CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 116, No. 4, Pt. 1 October 2004

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 116, NO. 4, OCTOBER 2004

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 116, NO. 4, OCTOBER 2004

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 116, NO. 4, OCTOBER 2004

CONTENTS—*Continued from preceding page*

ACOUSTICAL NEWS—USA

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

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

75th Anniversary Meeting of the Acoustical Society of America

The 75th Anniversary Meeting of the Acoustical Society of America was held 24–28 May 2004 at the Sheraton New York Hotel and Towers in New York, New York. The meeting included the usual meeting events plus special activities. The entire day on Wednesday was devoted to the celebration of the Society's 75th anniversary, with the afternoon left open for meeting attendees to participate in tours and other activities.

The meeting drew a total of 1719 registrants from 38 countries including 226 students (see Table I). A total of 1280 papers organized into 101 sessions covered the areas of interest of all 13 Technical Committees and the Committees on Education in Acoustics and Standards.

Special features of this meeting included a banquet, a celebration of the Society's 75th Anniversary, a variety of tours, and an organ concert.

TABLE I. Meeting attendees by country.

Argentina	1	Greece	1	Poland	4
Austria	1	Hong Kong	2	Portugal	1
Australia	9	India	4	Russia	5
Belgium	5	Ireland	1	Slovakia	1
Brazil	3	Israel	2	Singapore	1
Canada	57	Italy	9	South Africa	1
Chile	4	Japan	41	Spain	6
China	3	Malaysia	1	Sweden	5
Denmark	13	Mexico	2	Taiwan	11
Finland	2	Netherlands	10	Turkey	2
France	26	New Zealand	1	Ukraine	1
Germany	21	Norway	6	United Kingdom	30

On Monday evening, Leo Beranek and David Griesinger presented a tutorial on ''Listening to the Acoustics in Concert Halls'' to a standing-room only audience of over 200 people. Distinguished lectures were presented by Jens Blauert on ''Communication Acoustics'' and by Laymon Miller whose presentation was titled ''Noise: My 62 Years of It.'' Arrangements were made for the daily broadcast of the film recorded at the Society's 25th Anniversary meeting in 1954. Technical sessions covered a large range of acoustical topics including sessions focused on special topics. Special sessions were organized to honor Leo Beranek, D. Vance Holliday, Robert Apfel, Donald Griffin, and Charles Watson.

On Tuesday evening, the Society began the celebration of its 75th anniversary with a banquet attended by over 470 people. To capture the atmosphere of the 25th anniversary banquet, the reception hall was decorated with candelabras and strolling violinists provided entertainment during dinner. A ''plenary session'' followed dinner at which new fellows were announced, a Society prize was presented, and meeting organizers recognized.

ASA Vice President Anthony Atchley announced the election of the following new fellows who were presented with their Fellowship certificates by ASA President Ilene Busch-Vishniac: Abeer Alwan, David H. Chambers, Robin O. Cleveland, Li Deng, Alain de Cheveigne, Gary J. Heald, Jian Kang, Andrew J. Oxenham, Christopher J. Plack, Daniel Rouseff, Carrick Talmadge, D. Keith Wilson, Eric J. Woods (see Fig. 1). Stanley Dosso was announced as the recipient of the 2004 Medwin Prize in Acoustical Oceanograhy.

Meeting Cochairs were Russell Johnson and Damian Doria. Damain Doria (see Fig. 2) introduced the members of the local committee and thanked them for their work in organizing the meeting: Gary W. Elko, Technical Program Chair; Katharine Sawicki, Assistant to the Cochairs and Audio/Visual; Pam Brooks, Accompanying Persons Program; Konrad Kaczmarek, Signs; Dan Clayton, Technical Tours; Fredericka Bell-Berti, Education; Subha Maruvada, Photography and Jeffrey Ketterling, Poster Sessions.

The following members of the Technical Program Committee were also thanked and their work in arranging the technical portion of the meeting was recognized: Gary W. Elko, Chair; Ellen Livingston and Kyle Becker, Acoustical Oceanography; James Simmons and Robert Kull, Animal Bioacoustics; John Erdreich, Architectural Acoustics; Jeffrey Ketterling and Elisa Konofagou, Biomedical Ultrasound/Bioresponse to Vibration; Victor Sparrow, Education in Acoustics; Gary Elko and Ilene Busch-Vishniac, Engineering Acoustics; Victor Sparrow, Musical Acoustics; Bennett Brooks and Brigitte Schulte-Fortkamp, Noise; Ronald Roy and E. Carr Everbach, Physical Acoustics; Glenis Long, Psychological and Physiological Acoustics; Geoffrey Edelson, James Preisig, and Zoi-Heleni Michalopoulou, Signal Processing in Acoustics; Fredericka Bell-Berti and Laura Koenig, Speech Communication; Courtney Burroughs, Structural Acoustics and Vibration;

FIG. 1. New fellows of the Society with ASA President Ilene Busch-Vishniac (far left) and Vice President Anthony Atchley (far right).

FIG. 2. Damian Doria, New York Meeting Cochair.

James Preisig, and Natalia A. Sidorovskaia, Underwater Acoustics.

The members of the 75th Anniversary were announced and thanked for the organization of this extraordinary event: Leo L. Beranek and Patricia K. Kuhl, Cochairs; Anthony A. Atchley, Fredericka Bell-Berti, Richard H. Campbell, William J. Cavanaugh, Dan Clayton, Lawrence A. Crum, Christopher J. Jaffe, and Francis Kirschner. We also thank the following ASA volunteers who assisted with tours and photography arrangements during the meeting: William Hartmann, Christine Hartmann, Judy Dubno, Robert Frisina, and Clare Hurtgen.

The Vice President's gavel was presented to Anthony Atchley by Vice President-Elect Mark Hamilton (see Fig. 3). The ''plenary'' portion of the event was concluded with the presentation of the President's tuning fork to Ilene Busch-Vishniac by President-Elect William Kuperman (see Fig. 4).

The banquet was turned over to Larry Crum and Anthony Atchley who

FIG. 3. ASA Vice President-Elect Mark Hamilton (l) presents gavel to Vice President Anthony Atchley (r).

FIG. 4. ASA President-Elect William Kuperman presents tuning fork to President Ilene Busch-Vishniac.

served as Masters of Ceremony for the remainder of the evening. Thomas Rossing was introduced and lead everyone in singing happy birthday to the Society. He and three other ASA members then gathered as a ''Barbershop Quartet'' and entertained the audience with several songs. Other members of the quartet were Uwe Hansen, D. Murray Campbell, and Sten Ternström (see Fig. 5).

Next on the celebration ''agenda'' was the premier of a film celebrating the Society's ''Glorious Past'' which included segments of reminiscences by past presidents of the Society and footage from a film recorded during the banquet held at the 25th Anniversary meeting in 1954. This film,

FIG. 5. Barbershop quartet, (l–r) Uwe Hansen, Murray Campbell, Thomas Rossing, and Sten Ternström, entertains at banquet.

together with the 75th Anniversary Book, was given to each meeting participant as a memento (see Fig. 6).

The evening's celebration closed with brief presentations by three ASA

FIG. 6. Banquet participants view 75th Anniversary film.

fellows and Past Presidents who have received the President's National Medal of Science—Kenneth N. Stevens, James L. Flanagan, and Leo L. Beranek.

Wednesday morning was dedicated to the celebration of the 75th Anniversary and was held at the City Center in New York City. The event began with the presentation of four Society awards, by President Ilene Busch-Vishniac.

The 2003 Silver Medal in Musical Acoustics was presented to Johan E. F. Sundberg ''for contributions to understanding the acoustics of singing and musical performance and for leadership in musical acoustics research.'' The R. Bruce Lindsay Award was presented to Michael R. Bailey ''for contributions to the understanding of shock wave lithotripsy and nonlinear acoustics.'' The Helmholtz–Rayleigh Interdisciplinary Silver Medal in Architectural Acoustics and Noise was presented to David Lubman ''for work in noise and standards and for contributions to architectural and archeological acoustics.'' The Gold Medal was presented to Chester M. McKinney ''for pioneering research and leadership in underwater acoustics and high resolution sonar and for dedicated service to the Society (see Figs. 7–10).

The next segment of the celebration was titled ''Looking Forward.'' A group of nine ASA members were selected to present brief talks on the future of acoustics. The nine presenters included Kelly Benoit Bird, Dani Byrd, John Fahnline, Andrew J. Oxenham, Tyrone Porter, Purnima Ratilal, Lily M. Wang, Preston S. Wilson, and Ning Xiang. The presentations covered the range of acoustical topics covered by the Society's current activities (see Fig. 11).

The celebration ended with the demonstration of a virtual orchestra.

FIG. 7. ASA President Ilene Busch-Vishniac presents the Silver Medal in Musical Acoustics to Johan Sundberg.

The virtual orchestra included 24 loudspeakers with tonal quality to match the acoustical instrument that was being emulated. Two singers followed a professional conductor as part of the demonstration (see Fig. 12).

The celebration was adjourned and meeting participants departed to participate in a Fellows Reception at the City Center and in a variety of tours that had been scheduled for Wednesday afternoon.

There were 250 participants in the Fellows Reception. Photographs were taken of fellows who were grouped by year of election. Wednesday afternoon tours to the Steinway Piano Factory in Astoria, New York, the Rose Planetarium at the Museum of Natural History in New York, a tour of two church organs in Manhattan and a Circle Line cruise around Manhattan Island drew many enthusiastic participants. In the evening meeting participants were treated to a special organ concert held at St. Thomas Church near Fifth Avenue.

On Thursday evening a Social Hour was held at the Sheraton Hotel which provided a place for meeting attendees to socialize. This was followed by meetings of several of the Society's technical committees, some of which held joint meetings.

See Figures 13–25 for additional photos taken during the banquet.

In all the celebration was judged a great success by the participants. For members who were unable to attend, the ASA has made the 75th Anniversary Book and film available at cost. See the ad in this issue for details on how to obtain copies.

FIG. 8. ASA President Ilene Busch-Vishniac congratulates Michael R. Bailey, recipient of the R. Bruce Lindsay Award.

FIG. 9. ASA President Ilene Busch-Vishniac with David Lubman, recipient of the Helmholtz-Rayleigh Interdisciplinary Silver Medal.

I end this report with the closing words of the Introductory Chapter to the 75th Anniversary book: ''... celebrate the accomplishments of the people who make up the Acoustical Society of America and look ahead to all of the new reasons to celebrate that we will have at our 100th birthday!'' For those members who were able to attend, I hope you found your participation in the meeting productive and enjoyable. For members who were not able to join us, I hope this report has given you a sense of the wonderful celebration that was held in honoring the achievements of the past as well as the excitement for the future of the Society and of its members.''

ILENE J. BUSCH-VISHNIAC
President 2003–2004

FIG. 10. ASA President Ilene Busch-Vishniac with Chester McKinney, recipient of the Gold Medal.

FIG. 11. Patricia Kuhl, Cochair of the 75th Anniversary Celebration Committee (center) with the future of acoustics presenters.

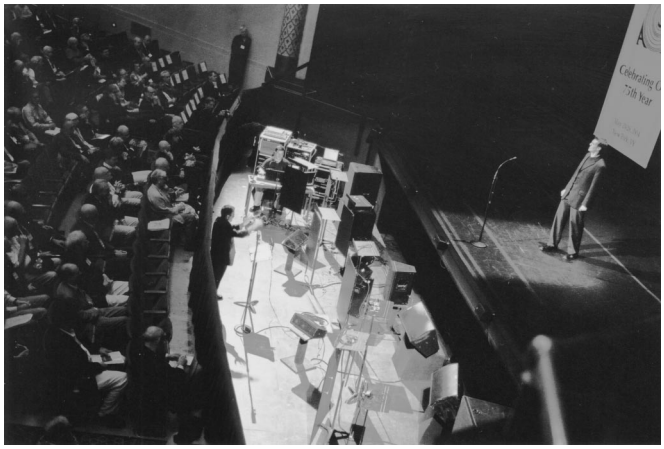


FIG. 12. Demonstration of virtual orchestra.

FIG. 13. Banquet Masters of Ceremony Anthony Atchley (l) and Lawrence Crum (r).

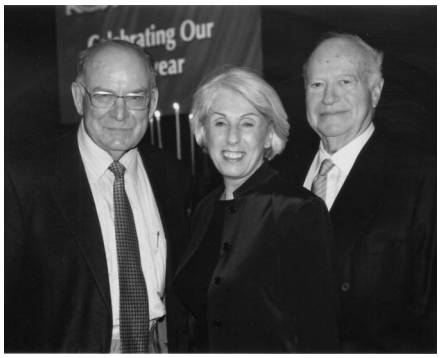


FIG. 14. Francis Kirschner (r), member of the 75th Anniversary Committee, with Mrs. Kirschner and Manfred Schroeder.

FIG. 15. Punita Singh and Thomas Rossing.

FIG. 16. Judy Dubno, Fredericka Bell-Berti, and Dan Clayton (l–r).

FIG. 17. The Rossings and the Crums (l–r): Jane Crum, Dolores Rossing, Thomas Rossing, and Lawrence Crum.

FIG. 18. Paul Schomer, Whitlow Au, and Mardi Hastings (l–r).

FIG. 19. Paul Ostergaard, Acoustical Society Foundation Chair (l) and Jim West, ASA Past President.

FIG. 20. Pat Kuhl, 75th Anniversary Meeting Cochair and James L. Flanagan, ASA Past President and recipient of the National Medal of Science.

FIG. 21. Jane and Larry Crum enjoy the strolling violinists at the banquet.

FIG. 22. David and Marjorie Blackstock.

FIG. 23. Robert Beyer (l) and Leo Beranek.

FIG. 24. Joel Sackett climbs ladder to take banquet photo.

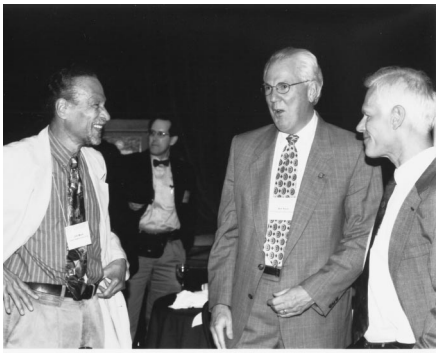


FIG. 25. Jim West, Bill Yost, and Bill Hartmann enjoying conversation at the banquet.

ASA North Texas Regional Chapter awards at Science Fair

More than 800 students from 86 North Texas schools competed in the 47th Dallas Morning News-Toyota Regional Science and Engineering Fair. This year two Senior Division contestants won Outstanding Acoustics Project Awards.

Mr. Cristian Oncescu, with the guidance of Ms. Karen Shepard, Plano Senior School, Plano ISD, investigated low frequency sounds accompanying physical phenomena (e.g., sea storms, tornadoes, volcanic eruptions, and auroras) (see Fig. 1). He developed JAVA pattern recognition and classification programs to distinguish and locate two types of atmospheric disturbances. For the second year, Mr. Oncescu won First Place in the Senior Earth and Space Sciences Division, and Third Place in Earth and Environmental Science at the Texas State Science and Engineering Fair.

Mr. Yi Fang, with the guidance of Mr. Richard Smith, Jasper HS, Plano ISD, investigated engineering considerations in microphone design. Mr. Fang received Honorable Mention in the Senior Physics Division.

Dr. Peter Assmann, Dr. Deb Rekart, Dr. Ben Seep, and Dr. Laurie Bornstein represented the Society.

FIG. 1. Christian Oncescu.

Report of the Auditor

Published herewith is a condensed version of our auditor's report for calendar year ended 31 December 2003.

Independent Auditors' Report

To the Executive Council
Acoustical Society of America

We have audited the accompanying statements of financial position of the Acoustical Society of America as of 31 December 2003 and 31 December 2002 and the related statements of activities and cash flows for the years then ended. These financial statements are the responsibility of the Society's management. Our responsibility is to express an opinion on the financial statements based on our audits.

We conducted our audits in accordance with auditing standards generally accepted in the United States of America. Those standards require that we plan and perform the audit to obtain reasonable assurance about whether the financial statements are free of material misstatement. An audit includes examining on a test basis, evidence supporting the amounts and disclosures in the financial statements. An audit also includes assessing the accounting principles used and significant estimates made by management, as well as evaluating the overall financial statement presentation. We believe that our audits provide a reasonable basis for our opinion.

In our opinion, the financial statements referred to above present fairly, in all material respects, the financial position of the Acoustical Society of America as of 31 December 2003 and 31 December 2002 and the changes in its net assets and its cash flows for the years then ended in conformity with accounting principles generally accepted in the United States of America.

CONROY, SMITH & CO.
Certified Public Accountants
April 7, 2004
New York, NY

ACOUSTICAL SOCIETY OF AMERICA
STATEMENTS OF FINANCIAL POSITION
AS OF 31 DECEMBER 2003 AND 2002

	December 31	
	2003	**2002**
Assets:		
Cash and cash equivalents	$2,349,616	$1,282,066
Accounts receivable	222,195	303,948
Marketable securities, at market	5,053,884	5,171,458
Furniture, fixtures, and equipment—net	55,089	79,532
Other assets	483,789	362,264
Total assets	$8,164,573	$7,199,268
Liabilities:		
Accounts payable and accrued expenses	$ 327,250	$ 313,297
Deferred revenue	1,196,196	1,430,153
Total liabilities	$1,523,446	$1,743,450
Net assets:		
Unrestricted	$5,640,080	$4,588,629
Temporarily restricted	634,148	502,170
Permanently restricted	366,899	365,019
Total net assets	$6,641,127	$5,455,818
Total liabilities and net assets	$8,164,573	$7,199,268

ACOUSTICAL SOCIETY OF AMERICA
STATEMENTS OF ACTIVITIES
FOR THE YEARS ENDED 31 DECEMBER 2003 AND 2002

	December 31	
	2003	**2002**
Changes in Unrestricted Net Assets		
Revenues:		
Dues	$ 664,508	$ 659,240
Publishing—JASA	2,132,642	2,074,151
Standards	304,778	322,925
Spring Meeting	186,353	182,808
Fall Meeting	196,830	239,367
Other member services revenue	17,488	16,175
Other	102,190	93,071
Net assets released from restrictions	68,348	72,637
	$3,673,137	$3,660,374
Expenses:		
Publishing	$1,566,927	$1,458,604
Standards	423,467	460,782
Spring Meeting	236,943	232,650
Fall Meeting	201,540	289,208
Member Services	225,964	186,211
Other	471,991	471,317
Administration	430,978	462,961
	$3,557,810	$3,561,733
Net income from operations	$ 115,327	$ 98,641
Nonoperating activities:		
Interest and dividends	$ 120,426	$ 130,915
Realized (loss) on investments	(569,494)	(89,461)
Unrealized gain (loss) on investments	1,385,192	(606,821)
	$ 936,124	($ 565,367)
Increase (decrease) in unrestricted net assets	$1,051,451	($ 466,726)
Changes in Temporarily Restricted Net Assets		
Contributions	$ 23,341	$ 10,858
Interest and dividends	22,768	27,137
Realized (loss)	(107,668)	(18,543)
Unrealized gain (loss)	(261,885)	(125,791)
Net assets released from restrictions	(68,348)	(72,637)
Increase (decrease) in temporarily restricted net assets	($ 131,978)	($ 178,976)
Changes in Permanently Restricted Net Assets		
Contributions	$ 1,880	–
Increase (decrease) in permanently restricted net assets	$ 1,880	–
Increase (decrease) in net assets	$1,185,309	($ 645,702)
Net assets at beginning of year	5,455,818	6,101,520
Net assets at end of year	$6,641,127	$5,455,818

USA Meetings Calendar

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

2004

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

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; Email: asa@aip.org; WWW:http://asa.aip.org].

16–19 May Society of Automotive Engineering Noise and Vibration Conference, Traverse City, MI [Patti Kreh, SAE International, 755 W. Big Beaver Rd., Ste. 1600, Troy, MI 48084, Tel.: 248-273-2474; Email: 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 October 150th Meeting joint with Noise-Con, Minneapolis, Minnesota, [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: 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

Aaron, Peters F., NSWCCD Code 713, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700

Andrade, Joseph D., Dept. of Bioengineering, Univ. of Utah, 50 South Central Campus Dr., Rm. 2480 MEB, Salt Lake City, UT 84112-9202.

Barber, Frank E., National Ctr. for Physical Acoustics, Univ. of Mississippi, One Coliseum Dr., University, MS 38677.

Barnhart, Timothy J., NSWCCD CODE 713, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700.

Barrett, Jillian G., 5520 Corte Sierra, Pleasanton, CA 94566.

Benfield, Mark C., Coastal Fisheries Inst., Louisiana State Univ., 2179 Energy, Coast & Environment Bldg., Baton Rouge, LA 70803.

Binns, Michael D., Acoustical Solutions, Inc., 2852 East Parham Rd., Richmond, VA 23228.

Bischoff, Lance B., Bernhard Shipps and Associates, 654 Chester Ave., Moorestown, NJ 08057.

Bitton, Gabriel, Glucon Medical, Pinsker 70, Petach Tikva, 49130 Israel.

Bocko, Mark F., Electrical and Computer Eng., Univ. of Rochester, 205 Hopeman Bldg., Rochester, NY 14627.

Brambilla, Giovanni, Via Rimini 14, Rome 00182 Italy.

Brock, Derek P., Naval Research Lab., Code 5513, 4555 Overlook Ave., SW, Washington, DC 20375.

Buchholz, Jorg M., Univ. of Western Sydney, MARCS Auditory Labs., Locked Bog 1797, Penrith South DC NSW 1797 Australia.

Buchner, Herbert W., Univ. of Erlangen-Nuremberg, Telecommunications Lab., Cauerstr. 7, Erlangen D-91058 Germany.

Caldwell, Jack, 314 Greenpark Dr., Houston, TX 77079-6414.

Chen, Angela C., The Greenbusch Group, Inc., 1900 West Nickerson St., Ste. 201, Seattle, WA 98119.

Cheng, Jason Y., Emergency Medicine, Univ. of Cincinnati, 231 Albert Sabin Way, Cincinnati, OH 45267-0769.

Clarkson, Marsha G., Georgia State Univ., Dept. of Psychology, MSC 2A1155 33 Gilmer St., SE Unit 2, Atlanta, GA 30303-3082.

Cushner, Joshua M., Shen, Milson and Wilke Inc., Acoustics, 417 Fifth Ave., New York, NY 10016.

D'Souza, Karl S., ABAQUS, Inc., Engineering Specialists, 1080 Main St., Pawtucket, RI 02860.

Davidson, Lisa, New York Univ., Linguistics, 719 Broadway, 4th Fl., New York, NY 10003.

Disner, Sandra F., 619 Tuallitan Rd., Los Angeles, CA 90049.

Duncan, Edward C., Resource Systems Group, Inc., 331 Olcott Dr., Ste. U1, White River Junction, VT 05001.

Duty, Jason R., Charles M. Salter Associates, Inc., 130 Sutter St., Ste. 500, San Francisco, CA 94104.

Eckenrode, Glenn M., NSWCCD CODE 713, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700.

Giurgiutiu, Victor, Mechanical Engineering, Univ. of South Carolina, 300 South Main St., Columbia, SC 29208.

Grigos, Maria I., Dept. of Speech Lang. Pathology Audio., New York Univ., 719 Broadway, Ste. 200, New York, NY 10003.

Groves, David L., 4076 Jami Ln., Snellville, GA 30039.

Hahn, Thomas R., AMP, Univ. of Miami, RSMAS, 4600 Rickenbacker Cswy., Miami, FL 33155.

Halberstam, Benjamin, 329 Aycrigg Ave., Passaic, NJ 07055.

Hedges, Richard B., SPL Control, Inc., 1400 Bishop St., Cambridge ON N1R 6W8 Canada.

Holland, Stephen D., 1125 Florida Ave., #510, Ames, IA 50014.

Howse, Geoffrey D. S., Industrial Acoustics Co Ltd., Engineering, IAC House, Moorside Rd., Winchester, Hants SO23 7US, United Kingdom.

Hsieh, Elaine Y., Charles M. Salter Associates Inc., Acoustical Engineering Consultant, 130 Sutter St., 5th Fl., San Francisco, CA 94104.

Johnstone, Tom, Univ. of Wisconsin-Madison, WM Keck Lab. for Functional Brain Imaging and Behavior, 1500 Highland Ave., Madison, WI 53705.

Kinch, Matthew S., Hoover & Keith, Inc., 11391 Meadowglen Ln., Ste. D, Houston, TX 77082.

Kochanski, Greg P., Phonetics Lab., Univ. of Oxford, 41 Wellington Square, Oxford OX1 2JF, UK.

Konofagou, Elisa E., Columbia Univ., Biomedical Engineering MC 8904, 1210 Amsterdam Ave., 351 Eng. Ter., New York, NY 10027.

Kroeker, John P., Eliza Corporation, 350G Cummings Ctr., Beverly, MA 01915.

Kumar, Anurag, Geo-Marine, Inc., 11846 Rock Landing Dr., Newport News, VA 23606-4206.

Kunisato, Seto, Mechanical Dept., Saga Univ., 1 Honjo-machi, Saga 840-8502 Japan.

Kurtz, Alexis D., 2277 Westlake Court, Oceanside, NY 11572.

Li, Pai-Chi, Electrical Engineering, National Taiwan Univ., No. 1, Sec. 4, Roosevelt Rd., Taipei, 106 Taiwan, ROC.

Liangwu, Cai, Kansas State Univ., Mechanical & Nuclear Engineering, 331 Rathbone Hall, Manhattan, KS 66506.

Martin, Rainer, Ruhr-Univ. Bochum, Inst. of Communication Acoustics IC1-131, Bochum 44780, Germany.

McClatchie, Sam, South Australian Aquatic Sciences Center, Sardi Aquatic Sciences, 2 Hamra Ave., West Beach, Adelaide SA 5024 Australia.

Miller, Denise M., Lewis Goodfriend and Associates, 760 Route 10 West, Whippany, NJ 07981.

Mitra, Partha P., Neuroscience, Cold Spring Harbor Lab., Freeman Bldg., 1 Bungtown Rd., Cold Spring Harbor, NY 11724.

Munhall, Kevin G., Psychology, Queen's Univ., 62 Arch St., Kingston ON K7L 3N6 Canada.

Nakashima, Ann M., Defense Res. and Develop. Canada Toronto, Human Factors Res. and Eng. Section, 1133 Sheppard Ave., West, Toronto, ON M3M 3B9 Canada.

Nocke, Christian, Alte Raad 20A, Oldenburg D-26127 Germany.

Nogaki, Geraldine, Auditory Implants and Perception, House Ear Inst., 2100 West Third St., Los Angeles, CA 90057.

Orlando, Aristizabal, Skirball Inst. NYU Medical Ctr., Structural Biology, 540 First Ave., 5th Fl., Lab 13, New York, NY 10016.

Park, Junhong, National Research Council, MS 462 NASA Langley Research Ctr., Yorktown, VA 23681-2199.

Parker, Judith A., English, Linguistics and Speech, Mary Washington College, 1301 College Ave., Fredericksburg, VA 22401.

Persson Waye, Kerstin B.-M., Environmental Medicine, Gotebord Univ., The Sahlgrenska Academy, Box 414, Medicinaregatan 16A, Goteborg 40530 Sweden.

Pinto, Mario A., NSWCCD Code 713, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700.

Pinyard, Scott, Shen Milsom and Wilke, Inc., 417 5th Ave., 5th Fl., New York, NY 10016.

Polonichko, Vadim, SonTek, 6837 Nancy Ridge Dr., Ste. A, San Diego, CA 92121.

Pride, Jeffrey B., Engineering Dynamics International, 8420 Delmar Blvd., St. Louis, MO 63132.

Qiong, Zhang, Dalian Scientific Test & Control Tech. Inst., Administer, P.O. Box 67, Zhongshan District, Binhai St. 14, Dalian, Liaoning 116013, People's Republic of China.

Quinlan, John A., Rutgers Univ., Inst. of Marine and Coastal Sciences, 71 Dudley Rd., New Brunswick, NJ 08901.

Quintos, Dario J., IAA Technologies, Unit 125 Regalia Park Tower, 150 P. Tuazon Blvd., Cubao, Quezon City, 1109 Philippines.

Restrepo, Juan M., Mathematics Dept., Univ. of Arizona, Mathematics Bldg., Tucson, AZ 85721.

Runia, John, Theatre Projects Consultants, 25 Elizabeth St., South Norwalk, CT 06854.

Rybakov, Leonard, Overseas Connection Company, 7510 Sunset Blvd., #1044, Los Angeles, CA 90046.

Schwob, Michael A., JBA Consulting Engineers, 5040 West Edna Ave., Las Vegas, NV 89146.

Shimamura, Tetsuya, Saitama Univ., Dept. of Inform. & Comput. Sci., 255 Shimo-Okubo, Saitama 338-8570 Japan.

Showen, Robert L., ShotSpotter, Inc., 809B Cuesta, #205, Mountain View, CA 94040.

Sifakis, Minas K., Lakonias 4 St., Athens 115 23, Greece.

Tutton, Robert L., NSWCCD Code 713, 9500 MacArthur Blvd., West Bethesda, MD 20817-5700.

Vanhille, Christian, ESCET, Univ. Rey Juan Carlos, Tulipan, s/n, Mostoles, Madrid, 28933 Spain.

Walsh, Timothy F., Computational Solid Mechanics and Structural Dynamics, Sandia National Labs., P.O. Box 5800, MS 0380, Albuquerque, NM 87185.

Warda, Joseph, 43-60 Douglaston Parkway, Apt. 420, Douglaston, NY 11363.

Wester, Eric C., Engineering Dynamics, Industrial-Research Limited, 24 Balfour Rd., Parnell, P.O. Box 2225, Auckland 0001 New Zealand.

Wiggins, Sean M., MPL, Scripps Inst. of Oceanography, 9500 Gilman Dr., MC 0205, La Jolla, CA 92093-0205.

Wilder, William D., 14200 Olive View Dr., Sylmar, CA 91342.

Williams, Duncan P., Underwater Systems, DSTL, Rm. N106, Bldg. A32, Winfrith Technology Center, Dorchester DT2 8WX, UK.

Wong, Willy, Dept. of Electrical and Computer Eng., Univ. of Toronto, 10 King's College Rd., Toronto ON M5S 3G4 Canada.

Wood, Gary, SoniSys, Inc., 735 North Water St., Ste. 830, Milwaukee, WI 53202.

Wort, Philip M., 61 Findlay Ave., Ottawa ON, Canada K1S 2V1.

Yontz, Barbara F., Watkins College of Art, 2298 MetroCtr. Blvd., Nashville, TN 37228.

Yu, Alan C., Univ. of Chicago, Linguistics, 1010 East 59th St., Chicago, IL 60637.

Zhang, Weiguo, 37620 Scotsdale Circle, #203, Westland, MI 48185.

New Students

Abercrombie, Clemeth L., 105 East Green C14, Champaign, IL 61820.

Alexander, Jennifer A., Linguistics, Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208.

Allen, Matt S., 2402 Inverloch Circle, Duluth, GA 30096.

Angert, Phillip E., 2518 Dana St., Berkeley, CA 94704.

Baird, David, New Mexico Tech, 801 Leroy Place, Socorro, NM 87801.

Baker, Brad N., Mechanical Engineering, Univ. of Idaho, P.O. Box 440902, Moscow, ID 83844-0902.

Barger, Mari M., 13433 Roosevelt Way, North #201, Seattle, WA 98133.

Bauer, II, Paul F., P.O. Box 8742, Bloomington, IN 47407.

Bechwati, Fouad G., Flat 2305 City Heights, Victoria Bridge St., Salford, Lancashire M3 5AS, UK.

Bodony, Daniel J., Dept. of Aeronautics and Astronautics, Stanford Univ., 496 Lomita Mall, Rm. 362, Stanford, CA 94305-4035.

Botto, Lorenzo, Mechanical Engineering, Johns Hopkins Univ., 3400 North Charles, Baltimore, MD 21218.

Brown, Meredith A., 320 Memorial Dr., Cambridge, MA 02139.

Bubnash, Brian, 775 Changing Seasons Rd., Westminster, MD 21157.

Budhlakoti, Suvrat, Artificial Intelligence Group, Beckman Inst. of Advanced Sci. and Tech., 405 North Mathews Ave., Urbana, IL 61801.

Ca, Ji, Industrial System Engineering, Univ. of Regina, 3737 Wascana Parkway, Regina SK S4S 0A2, Canada.

Calandruccio, Lauren, 361 Slocum Heights, Apt. 2, Syracuse, NY 13210.

Campos-Astorkiza, Rebeka, USC Linguistics, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693.

Cardillo, Dominic J., 30817 Hidden Pines Ln., Roseville, MI 48066.

Chan, Kwan, 70 Pacific St., Apt. 539B, Cambridge, MA 02139.

Chandrasekhar, B., 1, 2nd Main, 3rd Block, Goraguntepalya, Bangalore, Karnataka, 560022 India.

Chenausky, Karen V., 40A Cottage St., Watertown, MA 02472.

Colburn, Michael G., 2111 Kabold Dr., G-201, Lawrence, KS 66047.

Collis, Jon M., 51 Christie St., Troy, NY 12180.

Danilov, Pavel, 2215 Elder St., Apt. D, Durham, NC 27705.

Dohen, Marion S., Institut de la Communication Parlee, 46 Ave. Felix Viallet, Grenoble 38 031 France.

Elliot, William J., 9396 Chapman Rd., New Hartford, NY 13413.

Gilcrist, Laura E., 110 Lyrae Dr., Getzville, NY 14068.

Gilichinskaya, Yana D., 820 180th St., Apt. 54, New York, NY 10033.

Grebner, Dawn M., 805D West Aaron Dr., #11, State College, PA 16803.

Gunel, Banu, Apt. 4, Huntingale, 20 Annadale Ave., Belfast, Antrim, BT7 3JH, UK.

Haire, Jay, 525 West Deming, #210, Chicago, IL 60614.

Hallenbeck, Stephen A., Rush Univ., Communication Disorders and Sciences, 600 South Paulina St., Chicago, IL 60612.

Hanna, Emily J., Linguistics, Cornell Univ., 203 Morrill Hall, Ithaca, NY 14853.

Hervais-Adelman, Alexis G., Medical Research Council, Cognition and Brain Sciences Unit, 15 Chaucer Rd., Cambridge CB2 3AL, UK.

Hill, Randy J., College of Marine Science, Univ. of South Florida, 140 7th Ave., South, St. Petersburg, FL 33701.

Holdhusen, Mark H., 738 Myrtle St., NE, Apt. 6, Atlanta, GA 30308.

Khanna, Rajat, 600 West 26th St., Apt. C120, Austin, TX 78705.

Klasen, Thomas J., Electrical Engineering-SCD, Katholieke Univ. Leuven, Kasteelpark Arenberg 10, Leuven-Heverlee 3001, Belgium.

Kotas, Charlotte W., 1185 Collier Rd., Apt. 16D, Atlanta, GA 30318.

Krivokapic, Jelena, Linguistics, Univ. of Southern California, 3601 Watt Way, Grace Ford Salvatori, Rm. 301, Los Angeles, CA 90089.

Kroesen, Maarten, Technology Policy and Management, Technical Univ. of Delft, Jaffalaan 5, Delft 2600GA, The Netherlands.

Le Cocq, Cecile, Ecole de technologie Superieure, 1100 rue Notre-Dame Ouest, Montreal QC, Canada H3C 1K3.

Lee, Judy, 65 Buchanan Ave., North Balwyne, Melbourne VIC 3104, Australia.

Leider, Martha D. K., Speech and Hearing Science (Student), Hofstra Univ., Hempstead Turnpike, Hempstead, NY 11549.

Lilly, Christopher F., 1200 North Perkins Rd., Apt. Q-1, Stillwater, OK 74075.

Liu, Fang, 572 Central Ave., 5B, New Haven, CT 06515.

Lobdell, Bryce E., 202 East White St., 36, Champaign, IL 61820.

MacGillivray, Alexander O., 315-900 Tolmie Ave., Victoria BC, Canada V8X 3W6.

Madani, Vahid, Downing College, Regent St., Cambridge CB2 1DQ, England.

Mahmud, Shohel, Mechanical Engineering, Univ. of Waterloo, 200 Univ. Ave., West, Waterloo ON, Canada N2L 3G1.

Marentakis, Georgios N., Computing Science, Univ. of Glasgow, 17 Lilybank Gardens, Glasgow, Strathclyde, G12 8QQ, UK.

Mari, Jean-Martial, CREATIS, INSA de Lyon, Bat. Blaise Pascal, Villeurbanne 69621, France.

Martinson, Eric B., College of Computing, Georgia Inst. of Technology, Atlanta, GA 30332.

McCreery, Anthony B., Mechanical Engineering, Georgia Inst. of Technology, 801 Ferst Dr., NW, Atlanta, GA 30332.

Miller, Erik T., 185 Pine St., 820, Manchester, CT 06040.

Miller, Jonathan, P.O. Box 22, Ashville, PA 16613.

Mueller, Jacob L., Rensselaer Polytechnic Inst., Acoustics, 110 8th St., Troy, NY 12180.

Nielsen, Kuniko Y., Linguistics, UCLA, 3125 Campbell Hall, Los Angeles, CA 90095-1543.

Nordhoff, Charles M., 705 Mulberry Court, Algonquin, IL 60102.

Ntanos, Christos, Tiverton Rd. 163, Birmingham B29 6 BS, UK.

Ortmann, Amanda J., 5535 Covode St., Apt. 10, Pittsburgh, PA 15217.

Paik, Soonkwon, 2501 Lake Austin Blvd., F103, Austin, TX 78703.

Park, Hanyong, Linguistics Dept., Indiana Univ., 322 Memorial Hall, 1021 East 3rd St., Bloomington, IN 47405.

Parnum, Iain M., Ctr. for Marine Science and Technology, Applied Physics, GPO Box U1987, Perth WA 6845, Australia.

Pillai, Jagan A., Communication Sciences and Disorders, Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208.

Poore, Meredith A., Speech-Language Hearing, Univ. of Kansas, 3001 Dole Ctr., 1000 Sunnyside Ave., Lawrence, KS 66045.

Pueo, Basilio, Urb Las Mezquitas 23, San Juan 03550 Spain.

Que, Ying P., TITR, Univ. of Wollongong, Level 1, Bldg. 4, Northfield Ave., Wollongong NSW 2522, Australia.

Queenan, Alexander E., 9 Jessie Ter., Southampton SO14 3EF, UK.

Rodriguez, Benjamin, Speech, Language and Hearing Sci., Univ. of Colorado at Boulder, 2501 Kittredge Loop Rd., Boulder, CO 80309-0409.

Rolshofen, Wolfgang, Inst. for Mechanical Engineering, Technical Univ. of Clausthal, Robert-Koch Strasse 32, Clausthal-Zellerfeld, 38678, Germany.

Ryan, Robert H., 1048 West Huron, 3W, Chicago, IL 60622.

Schmidt, Benjamin A., 112 Wellington Rd., 7, Elora ON, Canada N0B 1S0.

Shah, Jashmin K., Electrical and Computer Engineering, Temple Univ., 1947 North 12th St., College of Engineering, Philadelphia, PA 19122.

Shank, Kevin R., KRAS Muzik, LLC, 15813 Southeast Rivershore Dr., Vancouver, WA 98683.

Shi, Lufeng, Syracuse Univ., Communication Sciences and Disorders, 805 South Crouse Ave., Rm. 200, Syracuse, NY 13244.

Slagley, Jeremy M., West Virginia Univ., Industrial and Management Syst. Eng., 353 C Mineral Resources Bldg., Morgantown, West Virginia 26505.

Somerville, Andrew L., 935 North Honore St., Apt. 1F, Chicago, IL 60622.

Sperbeck, Mieko N., 703 Pelham Rd., Unit 505, New Rochelle, NY 10805.

Thompson, Mya E., Cornell Univ., Neurobiology and Behavior, Mudd Hall, Ithaca, NY 14850.

Thornhart, Tomas, NEMESIS Audio and Design, Acoustics, Av. Brig. Faria Lima 1826 cj 910, Sao Paulo, SP 01451-001, Brazil.

Tien Ping, Tan, 280 Lorong Beringin 5, Taman Beringin, Kulim, Kedah 09000, Malaysia.

Tollefsen, Cristina, Dept. of Physics and Physical Oceanography, Memorial Univ. of Newfoundland, St. John's NF, Canada A1B 3X7.

Utley, Daniel L., 1189 Beall Ave., C-2900, Wooster, OH 44691.

Velikic, Gordana, ECE, Univ. of Rochester, Hopeman 204, Rochester, NY 14627.

Vijayakumar, V., Multimedia Univ., FOSEE, Jalan Air Keroh Lama, Melaka 75450, Malaysia.

Villalba, Sara E., School of Communication Sciences, McGill Univ., 1266 av edes Pins, Montreal PQ, Canada H1E 4T5.

Walls, Chris P., 39 New Rd., Clanfield PO8 0NR, UK.

Walter, Michael J., 1301 West 24th St., Apt. P-33, Lawrence, KS 66046.

Wang, Taiwei, 550 Memorial Dr., Apt. 13B2, Cambridge, MA 02139.

Way, Evelyn R., 463 Fulton St., 3, Troy, NY 12180.

Xu, Minghua, Texas A&M Univ., Biomedical Engineering, 233 Zachry Eng. Ctr., TAMU 3120, College Station, TX 77843.

Yuen, Michelle, Hawaii Inst. of Marine Biology, Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734.

New Electronic Associates

Ackley, Robert S., Hearing Speech and Language Science, Gallaudet Univ., 8000 Florida Ave., NE, Washington, DC 20002.

Bemman, Ya-Juan, Eberspaecher North America, 29101 Haggerty Rd., Novi, MI 48377.

Brown, Richard O., The Exploratorium, 3601 Lyon St., San Francisco, CA 94123.

Bullmore, Andrew J., Hoare Lea Acoustics, Acoustics Group, 140 Aztec West Business Park, Almondsbury, SG BS324TX, United Kingdom.

Candiver, Amy, 142 Page Rd., Bedford, MA 01730.

Chu, William, McKay Conant Brook, 5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362.

Concilio, Antonio, CIRA, Sound and Vibration Lab., Via Maiorise, Capua, Casreta, 81043, Italy.

Davidson, Robert J., Convergent Digital Company, 608 West Hays St., Boise, ID 83702.

Davis, Ted E., Wichita State Univ., ECE, 422 Waverly, Wichita, KS 67218.

Gelb, James M., Applied Research Lab., Univ. of Texas at Austin, 10000 Burnet Rd., Austin, TX 78758.

Gwinn, William R., 504 Ivy Circle, Alexandria, VA 22302.

Inurrieta, M. Eugenia, Ikerlan, S. Coop., Documentation, Paseo J.M. Arizmendiarrieta 2, Mondragon Gipuzkos 20500, Spain.

Johnson, David A., 7945 Ruststone Court, Colorado Springs, CO 80919.

Kanada, Yasusi, Hitachi, Ltd., Central Research Lab., Kokubunji, Tokyo 185-8601 Japan.

Kerer, Markus, Raiffeisenstrasse 28, Dombirn Vorarlberg 6850, Austria.

Nishino, Takanori, Ctr. for Information Media Studies, Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8603, Japan.

Noel, Claire, Semantic, 39 Chemin de la Buge, Sanary Sur Mer 83110, France.

Ortiz-Berenguer, Luis I., Univ. Politecnica de Madrid, DIAC, Ctra. Valencia km 7, Madrid 28031, Spain.

Plantevin, Philippe, SOACSY s.a.r.l., Co/CEEI-Provence BP 88, Aix en Provence cedex 4, Bouches du Rhone 13545, France.

Rasmussen, Soren, COWI, Odensevej 95, Odense DK-5260, Denmark.

Robinson, Matthew D., Robinson-Howell, 615 South Grant St., Denver, CO 80209.

Sadaka, Alain, Boston Scientific, 47900 Bayside Parkway, Fremont, CA 94538.

Shimokura, Ryota, Univ. of Bologna, Dept. of Energy, Nucl. and Environ. Control Eng., Viale Risorgimento 2, Bologna, Emilia Romagna 40136, Italy.

Spaargaren, Alamo, Ensigma Technologies, Turin House, Station Rd., Chepstow, Monmouthshire, NP1 65PB, UK.

Spaulding, Rick, The Environmental Company, 6765 NE Day Rd., Bainbridge, WA 98110.

Thomas, Norton F., Kinetics West Inc., 5808 South Rapp St., 140, Littleton, CO 80120.

Yates, Douglas A., 84 Candlestick Rd., North Andover, MA 01845.

Zhang, Yu, Univ. of Wisconsin, 1300 Univ. Ave., Rm. 5745, Madison, WI 53706.

New Corresponding Electronic Associates

Fugiel, Bogustaw, Inst. of Physics, Silesian Univ., Uniwersytecka 4, Katowice 40-007, Poland.

Ilkorur, Onur I., Yildiz Technic Univ., Mechanical Engineering, Barbaros Bulv., Istanbul 80750, Turkey.

Ouni, Kais, 10 Rue Abou El Kacem Chabbi, Denden, Tunis, 2011, Tunisia.

Renato, Alejandro, Genius Instituto de Tecnologia, ASR, Av. Acai 875 Bloco E. Dist. Indust., Manaus AM 69075904, Brazil.

Shevchenko, Tatiana I., Mira Prospect, 124-1-32, Moscow 129164, Russia.

Singh, Nandini C., National Brain Research Center, Nainwal Mode, Manesar, Dist.-Gurgaon, Gurgaon Haryana 122050, India.

Members Elected Fellows

M. A. Ainslie, J. J. Finneran, K. Genuit, R. F. Gragg, K. A. Hoover, P. A. Keating, V. Mellert, R. E. Remez, P. Roux, Y. Suzuki, S. C. Thompson, and M. E. Zakharia.

Associate Elected Fellow

S. A. Shamma

Associates Elected Members

W. M. Beltman, M. Bertsch, E. Bossy, R. Bresin, J. W. Choi, H. Dong, M. A. Epstein, R. J. Fedewa, V. N. Fokin, M. S. Fokina, J. T. Foulkes, W. L. J. Fox, B. N. Gover, B. S. Henderson, B. A. Henry, C.-J. Hsu, M. A. Huaquin, L. C. Hufnagle, Jr., M. J. Jarzabkowski, P. M. Jordan, S. Lee, F. Lin, Z. Lu, K. I. Matveev, K. D. Mitchell, J. A. Paulauskis, O. Poroy, J. J. Ridings, J.-S. Seo, D. B. Sharp, C. M. Traweek, J. L. Van Doorn, J. Vos, D. L. Watry, and L. Wu.

Students to Associates

B. D. Dooley, M. O. Lammers, B. W. Libbey, R. D. McCormick, G. S. Meltzner, T. M. Porter, B. D. Simpson, W. M. Whitmer, and H. Yuan.

Electronic Associates to Associates

Y.-J. Bemman and S. M. Reynolds.

Associates to Students

K. V. Chenausky and A. M. Freeman.

Associates to Electronic Associates

W. J. N. De Lima, R. Pitre, C. M. Ting, and M. F. Yama.

Students to Electronic Associates

R. Lee, S. K. Lineaweaver, J.-P. Roy, and C.-G. Tsai.

Resigned

J. Reisse, J. van den Eijk—*Fellows*

S. Anton, I. R. Bcknell, F. de Ribaupierre, T. V. Frazier, D. Gainer, J. M. Guinter, H. R. Hall, N. Maropis, I. Thalmann, K. Yamamura—*Members*

R. Allnor, F. A. DeVito, L. Halldin, V. A. Miraglia—*Associates*

E. L. Marzano—*Corresponding Electronic Associate*

Deceased

F. W. Cuomo, M. E. Hawley, R. C. Jones, I. G. Mattingly—*Fellows*

A. Bell, R. A. Isberg, C. O. Kleesattel, C. S. Kaminsky, E. Leighton, M. K. Wynne—*Members*

G. H. Sutton—*Associate*

Dropped

H. G. Davies, D. L. Folds, W. Kendall, S. G. Revoile, R. D. Stoll, E. H. Trinh—*Fellows*

W. A. Ainsworth, G. Alker, D. J. Anderson, S. W. Anderson, W. H. Anderson, M. K. Ang. S. Anton, H. A. Arts, D. P. Ayyappan, C. J. Bajdek, R. J. Barton, S. M. Bates, D. J. Bierenbaum, F. A. Boettcher, P. G. Bremner, D. J. Burr, G. W. Caille, Y. Champoux, D. D. B. Chan, D. Cheeke, R, A. Cooper, J. C. Cox, R. G. Daniloff, C. W. Dietrich, L. B. Dozier, A. D. Drake, F. A. Dube, J. P. Ertel, A. V. Furduev, G. P. Gibbs, A. Gopinath, R. M. Gorman, B. Guerin, M. P. Hagelberg, M. E. M. Head, C. E. Hime, W. L. Ho, W. Hogan, W. J. Jobst, D. R. Jones, M. G. Kim, S. L. Kneessi, S. Kojima, S. A. Kostarev, R. A. Krakow, N. Kraus, P. V. Krishnan, N. J. Lass, G. E. Leighton, Y. T. Y. T. Leung, T. Lin, P. R. Lizana, L. Lyamshev, A. I. Marash, T. S. McCain, J. V. McNicholas, J. L. Melanson, J. B. Merrill, R. B. Monsen, R. Moulder, T. W. Murray, V. V. Muzychenko, R. A. Nobili, J. E. Orban, A. T. Parsons, H.-S. Pei, L. A. Pflug, E. I. Plotkin, W. Qiao, J. S. Robertson, M. A. Rodriguez, H. F. Routh, S. T. Roweis, R. Saliba, A. Sarabasa, J. Shipps, M. D. Slurzberg, B. A. Snyder, S. R. Speer, T. Z. Strybel, A. M. Sutin, P. L. Tan, Y. C. Tong, N. B. Traylor, III, C. Tsuchitani, Y. Tsukahara, C. H. Venet, C. Wang, G. S. Wasserman, M. L. Wiederhold, Q. Qu, R. K. Yarber, X. Zhao, Z. Zhu, L. M. Zurk—*Members*

T. J. Adam, J. A. Adeff, H. Ajisaka, S. Y. Akhtar, M.-L. Alexander, S. F. Ali, H. E. Amorosa, J.-A. Bachorowski, J. Banchet, J.-P. R. Becker I. Y. Belyaeva, T. K. Bhatt, G. Biaswi, B. J. Bloom, C. E. Bradley, L. R. Brownlee, N. J. Buggy, D. L. Burwen, A. Cassone, H. Cha, K. Chandrasekaran, P. P. Chang, A. Cheung, H. C. Choe, K. Chung, W. Ciaramitaro, M. A. Cohen, A. J. Coleman, R. J. Comparin, J. F. Covey, J. C. Crew, K. M. Crosswhite, J. R. Curran, Y. Dain, K. Davis, P. R. Davis, A. S. De Brito, J. L. Delany, L. Dozsa, L. A. Drake, C. E. Drost, P. G. Dutilleux, M. O. Elegbe, C. M.

Elliott, S. Fang, B. J. Feng, A. M. Fincham, K. V. Fisher, K. H. Franck, T. J. Freeberg-Renwick, E. Gaja, V. GaLn.nko, C. W. Garnham, A. E. Geers, M. Gentil, N. V. Gerashehenko, Ghazanfar, R. J. Golembeski, P. Gori, S. M. Grace, D. A. Gruenhagen, L. D. Hager, J. E. Hamrin, V. P. Harper, M. Harun, R. O. Hauptmann, M. A. Hayner, J. F. Heake, M. J. Hirayama, H. J. Hoffman, J. S. Hollingshead, I. G. House, D. M. Houston, D. C. Hoyle, C. Huang, W. O. Hughes, K. Iida, A. D. Jenkins, E. R. Jensen, R. R. Johnson, P. Jonte, P. R. Kellock, E. Kelly, S. R. Khosla, P. R. Killeen, W. Kim, J. S. Kindred, K. Kobayashi, M. S. Kushwaha, E. H. Langendijk, A. I. Lavrentyev, B. A. Lawson, T. J. Leamy, B.-G. Lee, L. Lefebvre, D. R. Leon, J. W. Lewis, J.-F. Li, D. Lin, M. Lin, S. T. Link, D. S. Linton, H. Lu, W.-L. D. Ma, A. M. Machleder, S. F. Magruder, J. A. Mann III, T. L. Mann, M. Mantakas, G. B. Marshall, M. A. McHenry, T. N. KcKaig, M. F. McKinney, W. S. Melchert, X. Meng, A. M. E. Merlen, N. Michiko, G. R. Minasyan, A. Minguez, F. S. Miyara, T. M. Mueller, S. Muhdramli, J. Murdoch, D. A. Najolia, A. A. Nash, A. J. Nucci, D. N. O'Brien, J. S. O'Neill, P. A. Oates, M. R. Olsen, M. L. Ostendorf, C. Ozturk, T. Pahuja, R. A. Pappas, C. E. Parra, G. Pavic, F. R. Pereira, S. P. Perry, T. W. Piazza, P. Poesio, C. W. Ponton, E. D. Pope, G. Prado, S. Prasad, R. D. Rackley, L. A. C. Reis, A. F. Reller, M. P. Robb, I. E. Rogers, E. J. Rosenberg, A. M. Salazar, A. J. Sanchez-Guzman, R. Santos, E. K. Scheer, T. C. Schultze, M. J. Schulz, C. D. Scott, J. E. Scott, M. C. Sheppard, P. E. Shinn, J. J. Sidtis, Y. S. Sininger, W. B. Siong, C. E. Smith, A. Sodal, A. V. Sousa, D. F. St. Mary, J. F. Stasiewicz, Jr., B. D. Storey, S. B. Suppappola, D. A. Swanson, M. J. Swiergosz, S. Taherzadeh, D. Terzi, J. A. Tetnowski, D. B. Thiessen, C. T. Petrou, G. G. Tisato, N. N. P. Todd, T. M. Tomilina, P.-C. Torng, H. Traunmuller, C.-H. Tseng, T. N. Tyson, E. Tzdukermann, B. L. Vaupel, J. J. Venditti, E. Verona, J. A. Viator, M. D. Wagstaff, Jr., M. R. Walsh, M. T. Wazenski, E. White, G. M. Whitelaw, N. A. Whitmal, S. Xiaorui, M. Xu, M. B. Younger, I. Zawadzki, S. Zeroug, Y. Zhang, E. J. Zimmerman—*Associates*

W. A. Aasland, M. Agarwal, K. Akana, H. I. Alabi, M. C. Anzalone, T. C. Apple, J. C. Armstrong, S. E. Atev, J. L. Auer, A. Aziz, K. R. Ball, B. A. Becker, G. J. L. Beckers, D. A. Bell, F. W. Bentrem, E. A. Bernhardt, J. C. Berry, E. C. Bevill, A. R. Bindell, R. B. Biziorek, M. J. Blenman, B. D. Blodgett, T. A. Bolstad, R. Bozinoski, L. A. Bravo, N. Buedenbender, S. M. Burcsak, N. L. Callaway, P. A. Callen, E. M. Caner, J. G. Carlson, M. B. Cartwright, H. M. Carty, A. H. Chan, K.-K. Chan, S. S. Chang, S. R. Chapin, A. D. Chiesa, E. A. Chilcote, M. R. Chipley, S. R. Choate, E. A. Collison, M. T. Cone, Z. A. Constan, O. Couture, C. Cretton, A. L. D'Agostino, D. P. Davis, P. D. Deshpande, E. B. Dick, S. D. Dobson, T. C. Dow, K. T. Dunckley, G. Eberle, J. A. Elias, R. W. Elkhater, A. M. Englund, P. Escudero, R. W. Fisher, B. E. Fouts, A. B. Gertner, K. D. Girardin, C. Gonzalez, S. Griffin, N. Gupta, S. Hansen, M. J. Hartman, D. Herstad, B. G. Hiner, J. C. Hoffman, C. Holden, B. A. Hollingsworth, D. C. Holt, O. Houix, J. Huang, R. J. Hubbard, J. A. Hunt, L. G. Ireland, M. C. Ivazian, H. Jang, D. A. Johnson, A. D. Jorge, W. A. Kaf, J.-H. Kim, A. L. Kinney, J. F. Knapwurst, J. F. W. Kock, K. J. Kolbeck, I. Kourakata, O. Kwern, A. D. Lancaster, J. Y. Lau, M. A. Lawson, B. Lee, E. A. Lerner, D. M. Leslie, J. A. Lewis, Q. Li, Y. Li, Z. Li, K. C. Liao, C. Lin, J. M. Lorang, M. S. Luk'yanchuk, B. A. Lussier, G. M. Maki, J. P. Marozeau, B. J. McDermott, I. P. McGregor, C. F. Micallef, N. S. Miller, B. Monroy, M. S. Moondra, G. A. Moore, B. S. Muller, A. D. Munro, W.-B. Na, R. E. Nadel, D. R. Nair, M. F. Neelon, G. F. Newton, T. T. A. Nguyen, Y. Nie, L. C. F. Oliveira, D. M. Orbelo, C. Ordonez, A. D. Owens, D. A. Page, B. A. Pardo, C.-M. Park, H. Park, C. Ch. Pastiadis, A. M. Pawluk, H. B. Pincon, M. B. Pinyard, B. I. Raju, S. Ramamoorthy, A. A. Razak, D. Reby, M. C. Robinson, M. F. Roedelbronn, P. N. Rosso, C. M. Roup, C. D. Salthouse, A. A. P. Sarda, B. M. Schaffer, W. E. Schneider, M. Senova, J. E. Sheline, W. Shi, J. A. Skarulis, J. L. Smith, M. A. Smith, C. Sodsri, A. E. Sones, S. P. Stanton, G. L. Strycker, R. D. Summers, X. Sun, J. B. Tabeling, A. J. Tallon, K. S. Tarver, T. G. Terry-Williams, G. Tserdanelis, B. E. Tucker, A. S. Valentine, S. S. Vasan, G. J. Vazquez-Perez, P. Veprek, J. L. Via, C. A. Vila, J. Vliegen, N. M. Walker, N. K. Walker, F. Wang, J. A. Watson, J. F. Hayand, R. Weber, P. Welby, B. R. White, P. S. Whitehead, T. A. Wohlgemuth, J. Won, H. Y. Wong, M. K. Woodward, T. Wu, J. R. Yeh, S. Y. Yoon, M. S. Ziolkowski—*Students*

L. M. Anderson, d. Arbel, C. J. Calleja, W. H. Cho, T. A. Cox, C. K. Deacon, A. Dilorenzo, J. P. Evans, A. P. Goncalves, L. A. Hicks-Postar, Z. Hou, N. V. Joshi, I. Kocak, A. Konopka, A. Kordik, D. Krishnamurthy, J. O. Labahn, C. D. Logins, S. M. Mallicoat, K. P. McAloon, C. Menezes, J. W. Monroe, G. C. Moore, C. R. Pearsall, II, H. M. Rodriguez-Davila, M. S. Rutkiewicz, L. S. Ryden, G. P. Succi, W. B. Swain, A. K. Walden, A. H. Wootters, J. Wouters—*Electronic Associates*

Fellows	904
Members	2388
Associates	2695
Students	962
Electronic Associates	325
	7280

ACOUSTICAL NEWS—INTERNATIONAL

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

International Meetings Calendar

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

October 2004

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

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

6–7 **Institute of Acoustics Autumn Conference**, Oxford, UK. (Web: www.ioa.org.uk)

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

27–29 ∗**25th Symposium on Ultrasonic Electronics**, Sapporo, Japan. (Web: use-jp.org/USEframepage_E.html)

November 2004

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

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

8–9 ∗**17th Biennial Conference of the New Zealand Acoustical Society**, Wellington, New Zealand. (Web: www.acoustics.org.nz)

15–18 **15th Meeting of the Russian Acoustical Society**, Nizhny Novgorod, Russia. (Fax: +7 95 126 0100; Web: www.akin.ru)

17–19 **7th National Congress of the Turkish Acoustical Society**, Nevsehir-Cappadocia, Turkey. (Web: www.tak-.der.org)

December 2004

8–10 ∗**Tenth Australian International Conference on Speech Science & Technology**, Sydney, Australia. (Web: www.assta.org/sst/2004)

March 2005

14–17 ∗**31st Annual Meeting of the German Acoustical Society (DAGA'05)**, Munich, Germany. (Web: daga2005.de)

April 2005

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

May 2005

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

June 2005

20–23 ∗**IEEE Oceans05 Europe**, Brest, France. (ENST Bretagne—Technopôle Brest Iroise, 29238 Brest Cedex, France; Fax: +33 229 00 1098; Web: www.oceans05europe.org)

23–24 ∗**2nd Congress of the Alps-Adria Acoustical Association (AAAA2005)**, Opatija, Croatia. (Web: had.zea.fer.hr)

28–1 **International Conference on Underwater Acoustic Measurements: Technologies and Results**, Heraklion, Crete, Greece. (Web: uameasurements2005.iacm.forth.gr)

July 2005

11–14 **12th International Congress on Sound and Vibration**, Lisbon, Portugal. (Web: www.iiav.org)

August 2005

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

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

September 2005

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

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

11–15 **6th World Congress on Ultrasonics**, Beijing, China. (Secretariat of WCU 2005, Institute of Acoustics, Chinese Academy of Sciences, P.O. Box 2712 Beijing, 100080 China; Fax: +86 10 62553898; Web: www.ioa.ac.cn/wcu2005)

14–16 ∗**Autumn Meeting of the Acoustical Society of Japan**, Sendai, Japan. (Acoustical Society of Japan, Nakaura 5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; Fax: +81 3 5256 1022; Web: www.asj.gr.jp/index-en.html)

October 2005

19–21 **36th Spanish Congress on Acoustics Joint with 2005 Iberian Meeting on Acoustics**, Terrassa (Barcelona), Spain. (Sociedad Española de Acústica, Serrano 114, 28006 Madrid, Spain; Fax: +34 914 117 651; Web: www.ia.csic.es/sea/index.html)

June 2006

26–28 ∗**9th Western Pacific Acoustics Conference (WESPAC 9)**, Seoul, Korea. (Web: www.wespac9.org)

July 2007

9–12 **14th International Congress on Sound and Vibration (ICSV14)**, Cairns, Australia. (e-mail: n.kessissoglou@unsw.edu.au)

September 2007

2–7 **19th International Congress on Acoustics (ICA2007)**, Madrid, Spain. (SEA, Serrano 144, 28006 Madrid, Spain; Web: www.ica2007madrid.org)

9–12 ∗**ICA Satellite Symposium on Musical Acoustics**

(ISMA2007), Barcelona, Spain. (SEA, Serano 144, 28006 Madrid, Spain; Web: www.ica2007madrid.org)

June 2008
23–27

*__Joint Meeting of European Acoustical Association (EAA), Acoustical Society of America (ASA), and Acoustical Society of France (SFA)__, Paris, France. (e-mail: phillipe.blanc-benon@ec-lyon.fr)

Statistics—18th International Congress on Acoustics in Kyoto 2004

Statistics of the 18th International Congress on Acoustics which was held in Kyoto 4–9 April 2004 have now been released by the organizers.

The last column of the compilation below shows the number of registered participants (including students) from the various countries represented at the Congress. The other columns show the number of participants at previous Congresses held in Toronto (1986), Belgrade (1989), Beijing (1992), Trondheim (1995), Seattle (1998), and Rome (2001). The table reflects the fact that there have been some political changes in the 18 years covered by these statistics.

TABLE I.

Country	Toronto	Belgrade	Beijing	Trondheim	Seattle	Rome	Kyoto
Algeria	0	0	0	1	0	1	0
Argentina	5	0	0	0	0	5	0
Armenia			0	0	0	1	0
Australia	19	4	12	19	36	27	16
Austria	0	2	0	6	11	10	5
Belarus			0	0	0	2	0
Belgium	8	9	4	9	15	23	14
Brazil	4	1	3	3	12	18	4
Bulgaria	1	6	0	0	0	0	0
Canada	171	18	10	10	75	17	22
Chile	0	0	1	1	2	3	0
China	19	15	370	4	47	30	52
Colombia	0	0	0	0	1	0	0
Croatia			1	0	0	6	0
Czech Republic				4	3	11	3
Czechoslovakia	1	7	0				
Denmark	23	19	9	28	17	25	20
Egypt	1	1	0	0	1	1	1
Estonia			0	4	2	3	1
Finland	2	4	2	10	7	14	10
France	98	67	35	63	95	99	39
Germany (FRG)	71	55	32	41	49	64	58
Germany (GDR)	1	3					
Greece	0	3	0	0	1	4	0
Hong Kong	0	0	2	0	4	0	3
Hungary	3	22	0	11	1	6	1
Iceland	0	0	0	0	0	1	1
India	7	8	3	3	9	12	6
Indonesia	0	0	1	0	0	2	1
Iran	0	1	2	0	0	0	1
Ireland	0	0	0	0	3	3	0
Israel	2	0	3	0	3	3	2
Italy	15	13	8	16	28	160	21
Japan	143	87	106	96	173	237	723
Korea	2	0	21	4	33	25	69
Kuwait	0	0	0	0	0	1	0
Lebanon	0	0	0	0	0	2	0
Lithuania			0	0	0	2	2
Malaysia	0	0	2	0	0	0	0
Mexico	0	0	0	0	8	2	1
Morocco	0	0	1	0	0	0	0
Netherlands	18	21	8	14	27	19	21
New Zealand	4	1	0	2	9	5	3
Nigeria	1	0	0	0	0	0	5
Norway	8	9	3	89	18	24	7
Peru	0	0	0	1	0	0	1
Poland	10	16	2	24	12	36	13
Portugal	1	1	1	3	0	9	5
Romania	0	1	0	1	0	6	0
Russia			8	13	27	32	11

TABLE I. (*Continued*.)

Country	Toronto	Belgrade	Beijing	Trondheim	Seattle	Rome	Kyoto
Serbia & Monten.							2
Singapore	2	0	3	1	3	2	2
Slovakia				1	1	3	2
Slovenia			0	0	0	3	0
South Africa	8	1	1	0	3	0	0
Spain	16	4	2	7	13	25	12
Swaziland							7
Sweden	21	16	10	40	25	35	18
Switzerland	3	4	0	7	7	21	2
Taiwan							10
Thailand	0	0	1	0	0	0	0
Tunisia	0	0	0	0	0	2	1
Turkey	1	0	0	1	1	7	1
UAE	0	0	0	0	0	0	1
Uganda	0	0	0	0	0	0	2
United Kingdom	46	27	20	35	77	55	41
Ukraine			0	0	2	1	0
Uruguay	0	0	0	0	0	2	0
USA	201	61	64	69	1066	176	89
USSR	12	31					
Uzbekistan			0	0	0	1	0
Venezuela	0	0	0	0	2	3	0
Yugoslavia	6	169	3	0	3	6	
Total	954	707	754	642	1932	1293	1332
Accompanying persons	133	63	95	98	192	N/A	91

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).]*

Fundamentals of Ocean Acoustics (3rd edition)

L. M. Brekhovskikh and Yu. P. Lysanov

Springer-Verlag, New York, 2002.
xiv+277 pp. Price $99.00 (hardcover), ISBN: 0387954678.

In approximately 10 year intervals this book has been published, with some revisions, this being the third edition. The authors are two of the most distinguished undersea acoustic scientists in the world and this book exhibits their breadth and depth and their deft ability to convey the essentials to the reader. The first two editions were reviewed by Claude Horton and Bob Beyer, respectively, and it is indeed an honor to follow them.

Chapter one introduces the reader to some of the most important features of the sea that affect ocean acoustics and to the fundamental acoustic effects that they cause, such as sound speed variability, scattering from the boundaries, bubbles, and scattering layers. The authors have added a little more on ambient noise (although this topic is not covered in the remainder of the book), and introduced a relative newcomer to the field, intrathermocline lenses.

From Chapter two onward the fundamentals of the theories of propagation and scattering are treated. The book is compact and the reader is expected to be mathematically prepared for what is coming. (Indeed the authors, in the third edition, state in the preface ''The book is intended for experts in acoustics and oceanology, engineers, postgraduates and students of universities and institutes of geophysical and hydrometeorological profiles.'') The reader is expected, often, to fill in the mathematical steps between equations. Idealized physical models of the sea are only introduced to give physical insight to more realistic problems. Three new topics are presented: an interesting section on weakly divergent bundles of rays (in deep water), a section on coupled modes in range dependent wave guides, and a brief and somewhat descriptive section on ocean acoustic tomography. Chapters nine and ten introduce the methods for treating the sea as a stochastic medium. The theories of propagation and scattering in random media are developed, showing the correlation of acoustically measured properties with those of the media, such as turbulence and internal waves. Three new topics are presented: the small slope scattering approximation, scattering by highly anisotropic inhomogeneities with a fractal spectrum, and attenuation of low frequency sound in an underwater sound channel. The last chapter treats the propagation and scattering from bubbles in the sea. The effects of bubble clouds on the speed, scattering, and absorption of sound are shown by deriving the complex compressibility based on the scattering of sound from single spherical bubbles that are small compared to the acoustic wavelength. Some new material is presented but this reviewer does not see that it adds much to the fundamental issues of propagation, scattering, and absorption in the sea. None of the ocean experimental results that have been published in the last 20 years are reviewed or cited.

It is interesting to note that there is no reference to computational methods. Even in the section on the parabolic equation there is only one reference (to the work of Fred Tappert). There is no reason given for this omission in the preface. Other fields that are not covered are phase conjugation, time reversal, and matched field processing, all very popular today, but logically omitted from a book that does not include object detection as one of its objectives.

The references for each chapter are listed at the end of the book. There are some additional 60 references in this edition. The reader should be aware that the reference list is only a representative sample of the literature and is by no means meant to be comprehensive.

The book is highly recommended as a suitable advanced text or as a reference for the practitioner of acoustics. There are a few minor mistakes, but their discovery will serve the user well. For those who already have an earlier edition it is a personal call to know whether or not it is worth $100.00 for the new material.

RALPH R. GOODMAN
Department of Marine Science
University of Southern Mississippi
Stennis Space Center, Mississippi 39529

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:

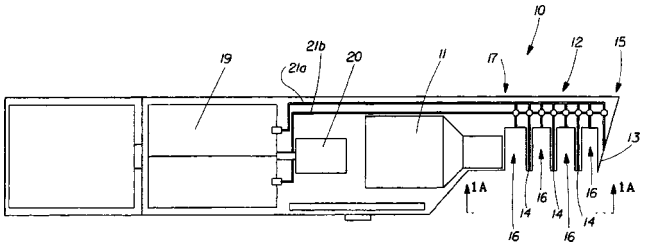
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID 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*
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*

6,732,744

43.20.Tb METHOD FOR THE ULTRASONIC TREATMENT OF HAIR AND OTHER KERATINOUS FIBERS

Michael Andrew Olshavsky and Ke Ming Quan, assignors to The Procter & Gamble Company
11 May 2004 (Class 132/200); filed 7 December 2001

The process for the oxidative treatment of keratinous fibers in human hair for the purpose of changing the color quality of the hair consists of placing a chelant (essentially a bleaching agent) and an ultrasonic treatment device close to the fiber. Energizing the ultrasonic device has the effect of



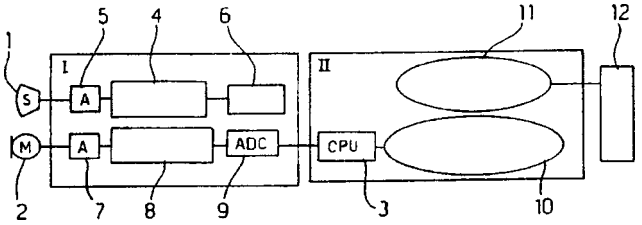
enhancing the deposition of the chelant onto the fiber. In another embodiment, a comb device, connected to two reservoirs for materials used to treat keratinous fibers, mixes and dispenses two materials to the hair under ultrasound activation.—DRR

6,728,661

43.20.Ye NONDESTRUCTIVE ACOUSTIC METHOD AND DEVICE, FOR THE DETERMINATION OF DETACHMENTS OF MURAL PAINTINGS

Giovanni Bosco Cannelli and Paola Calicchia, assignors to Consiglio Nazionale Delle Ricerche
27 April 2004 (Class 702/187); filed in Italy 25 June 1999

The detection of the amount of detachment of the separation zones in the layers in frescos and murals, or other stratified structures, without doing more damage to the piece of art or structure is described. The invention uses a nondestructive method that may be quite useful in the restoration process. The system uses an acoustic source **1**, acoustic receiver **2**, suitable transducer, and signal conditioning **I** and signal analysis **II**, particularly Cepstrum analysis, to perform the nondestructive analysis. This method is described as less destructive than a ''hammer'' technique and more precise than some other nondestructive techniques. The patent provides an introdution to the use of this technique and its proposed application.—NAS



6,732,515

43.35.Ud TRAVELING-WAVE THERMOACOUSTIC ENGINES WITH INTERNAL COMBUSTION

Nathan Thomas Weiland *et al.*, assignors to Georgia Tech Research Corporation
11 May 2004 (Class 60/520); filed 13 March 2003

The device described in this patent in essence is a thermoacoustic Stirling engine, which is comprised of a hot and a cold heat exchanger and a regenerator. The heat exchangers set up a thermal gradient in a regenerator that is located between them and that contains packing material that is fine enough so that the working fluid in the regenerator is essentially in equilibrium with the packing around it, but not so fine as to prevent the passage of acoustic waves. As an acoustic travelling wave passes through the regenerator's compressible fluid, it imparts pressure and velocity fluctuations to a small volume of the fluid and thus assists in the transfer of thermal energy. In devices according to the present patent the hot heat exchanger is replaced by a combustor, and means are provided for removal of the cold combustion products. A thermally insulated external feedback path directs a portion of the acoustic traveling wave from the hot side to the cold side of the regenerator.—EEU

6,725,704

43.35.Zc GAS ANALYZER

Jesper Lange and Jorgen Christensen, assignors to PAS Technology A/S
27 April 2004 (Class 73/23.2); filed in Denmark 14 January 2000

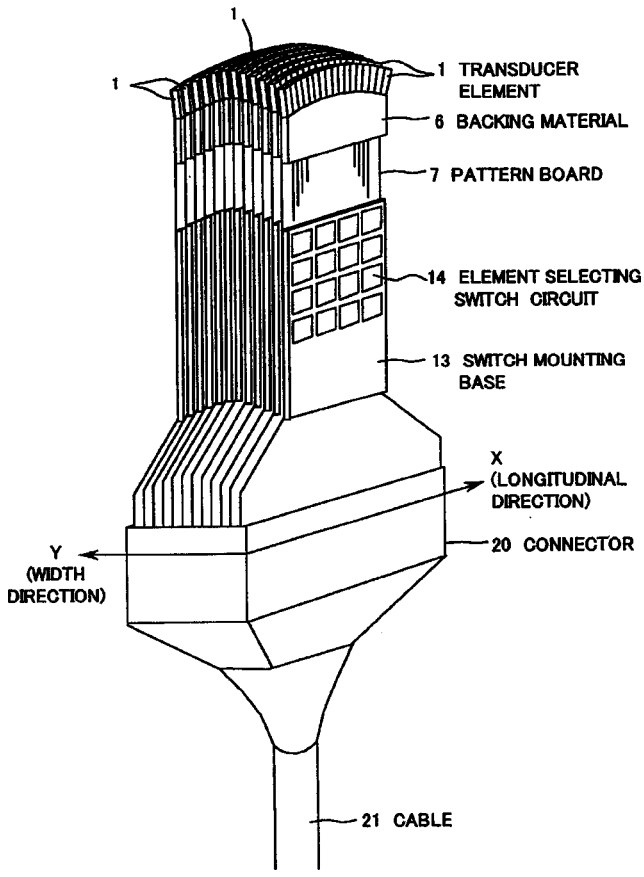
The gas analyzer described in this patent consists of a measuring chamber, gas inlet and outlet channels, and a means for providing a predetermined volume flow rate through the measuring chamber. As an applied pulsating magnetic field acts on the gas, it affects the oxygen in the gas, resulting in pressure changes that are sensed by a microphone. The application of pulsating electromagnetic radiation, such as infrared light, results in periodic heating and attendant pressure changes (depending on the wavelength of the radiation and the presence of certain components), which also are sensed by a microphone. The inlet and outlet channels consist of flow restrictions and cavities, designed so that external noise is kept out of the measuring chamber.—EEU

6,736,779

43.38.Fx ULTRASONIC PROBE AND ULTRASONIC DIAGNOSTIC DEVICE COMPRISING THE SAME

Shuzo Sano *et al.*, assignors to Hitachi Medical Corporation
18 May 2004 (Class 600/447); filed in Japan 17 September 1999

In this ultrasonic probe, two-dimensional array transducer elements are arrayed convexly in two perpendicular directions. This allows three-dimensional scanning of an object with an ultrasonic beam. By switching a selecting circuit, the shape and position of the probe's effective ultrasound emitting and receiving diameter are arbitrarily determined. Bundling the transducer elements in one ring decreases the number of connecting cables between the probe and the main body. The Fresnel ring is moved every ultrasonic reception cycle, thus scanning an object three-dimensionally with an ultrasonic beam.—DRR

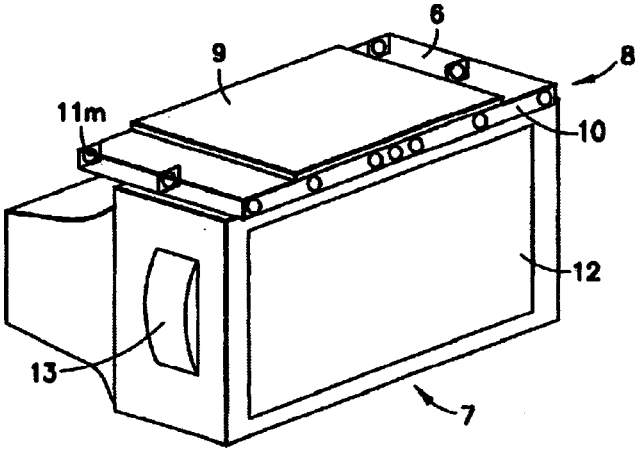


6,718,041

43.38.Hz ECHO ATTENUATING METHOD AND DEVICE

Bernard Debail, assignor to France Telecom
6 April 2004 (Class 381/66); filed in France 3 October 2000

An echo-cancelling teleconference terminal is made up of a conventional video monitor **7**, including loudspeaker **13**, plus an array of microphones **10**, and associated electronic circuitry. At the beginning of a conference session, a brief maximal-length sequence test signal is generated and the early-sound transfer function between the loudspeaker and each microphone is estimated in the form of an impulse response. At this point, antenna theory takes over, both in the signal processing itself and in the text of the



patent. Those who have no fear of triple integrals and spatial Fourier transforms will find the explanation enlightening. ''The channel-forming calculation is not performed solely on the basis of theoretical propagation models or of measurements performed in a quiet room, but also on the basis of estimated transfer functions obtained on site. Thus, the invention makes it possible to reduce direct and/or semi-direct coupling while imposing desired directivity and while controlling the maximum amplification of incoherent noise.''—GLA

6,724,899

43.38.Hz SOUND PICK-UP AND REPRODUCTION SYSTEM FOR REDUCING AN ECHO RESULTING FROM ACOUSTIC COUPLING BETWEEN A SOUND PICK-UP AND A SOUND REPRODUCTION DEVICE

Wolfgang Taeger and Gregoire Le Tourneur, assignors to France Telecom S.A.
20 April 2004 (Class 381/83); filed in France 28 October 1998

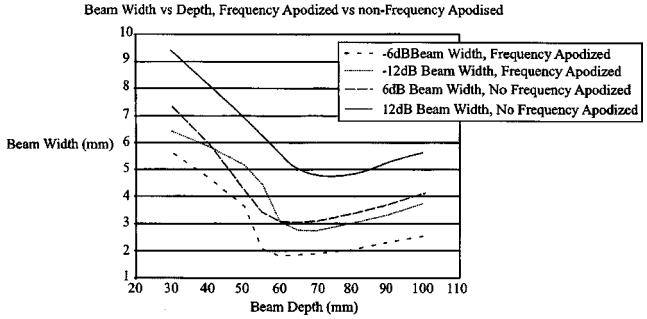
The performance goals and basic description of this invention are similar to those covered by Unites States Patent 6,718,041, reviewed above. In this instance, a different set of equations is used to control the digital signal processing circuitry. The system described includes ''at least two sound sensors situated at different distances from the reproduction device and a processing unit provided for using the amplitude and phase of the signals originating from each sound sensor in order to deliver a signal to the application unit whose echo signal has been reduced.''—GLA

6,726,631

43.38.Hz FREQUENCY AND AMPLITUDE APODIZATION OF TRANSDUCERS

Ram Hatangadi *et al.*, assignors to GE Parallel Designs, Incorporated
27 April 2004 (Class 600/459); filed 17 September 2001

A factor in interfering with high resolutions in medical ultrasonic scanning is the fact that the ultrasound signal undergoes attenuation and dispersion as it probes deeper into tissue. This degradation is governed by the Kramer–Kronig relationship [M. O'Donnel, E. T. Jaynes, and I. G. Miller, J. Acoust. Soc. Am. **69**, 696 (1981)]. One method for improving resolution is to frequency apodize the transducer aperture. The apparatus described in this



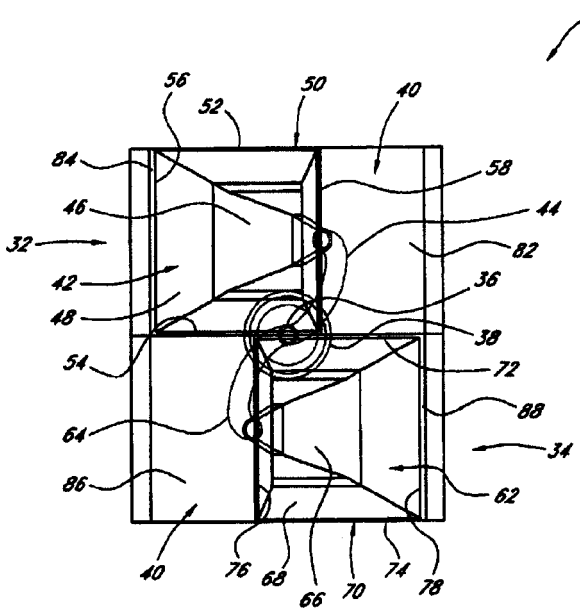
patent provides an ultrasonic transducer with frequency and amplitude apodization in order to improve signal quality and thereby improve ultrasonic imaging. Composite cuts into the piezoelectric material with varying widths and/or spacings according to a predetermined pattern result in a variation of the concentration of piezoelectric material across the surface of the transducer.—DRR

6,712,177

43.38.Ja CROSS-FIRED MULTIPLE HORN LOUDSPEAKER SYSTEM

Mark S. Ureda, Santa Ana, California
30 March 2004 (Class 181/152); filed 30 May 2001

Conventional high-frequency horns can be stacked and splayed to achieve wide horizontal coverage while reducing the vertical coverage angle. In most cases, this technique delivers better results than a fan-shaped array. Moreover, the general geometry can be adapted to provide a variety of coverage patterns. This patent describes a basic assembly consisting of a pair of horns joined to a common throat and a single compression driver. More than two-dozen figures and a total of 59 claims cover almost every conceivable variation.—GLA

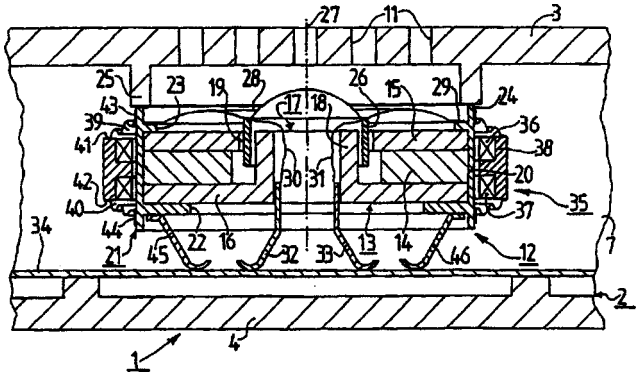


6,717,305

43.38.Ja APPARATUS HAVING AN ELECTROACOUSTIC TRANSDUCER FORMING A SOUND REPRODUCING MEANS AND A PART OF VIBRATION GENERATING MEANS

Ernst Ruberl *et al.*, assignors to Koninklijke Philips Electronics N.V.
6 April 2004 (Class 310/81); filed in the European Patent Office 17 February 2000

A cell phone must not only house a tiny, lightweight, efficient loudspeaker but a vibration generator as well. It seems logical to combine both functions in a single transducer. This patent describes an interesting ap-



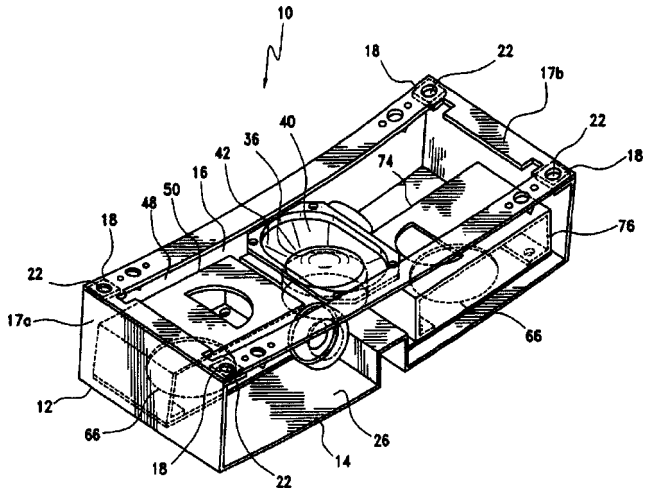
proach in which magnet **14**, voice coil **26**, and diaphragm **28** are elements of a conventional loudspeaker. Additional vibration-generating coils **36** and **37** simply make use of stray magnetic flux.—GLA

6,719,090

43.38.Ja SPEAKER ASSEMBLY

Dennis A. Tracy, Culver City, California
13 April 2004 (Class 181/145); filed 4 March 2002

The goal of this invention is a high-fidelity loudspeaker assembly design that meets the severe size and weight restrictions of private or commercial aircraft. So far, so good. What the inventor proposes is an open-ended



compartment housing loudspeakers that are partially covered ''to thereby alter the frequency response [and] also improve phase cancellation and resonant characteristics.'' Maybe so, but I would still like to hear it first.—GLA

6,720,708

43.38.Ja MECHANICAL-TO-ACOUSTICAL TRANSFORMER AND MULTI-MEDIA FLAT FILM SPEAKER

Lewis Athanas, assignor to Lewis Athanas
13 April 2004 (Class 310/324); filed 5 January 2001

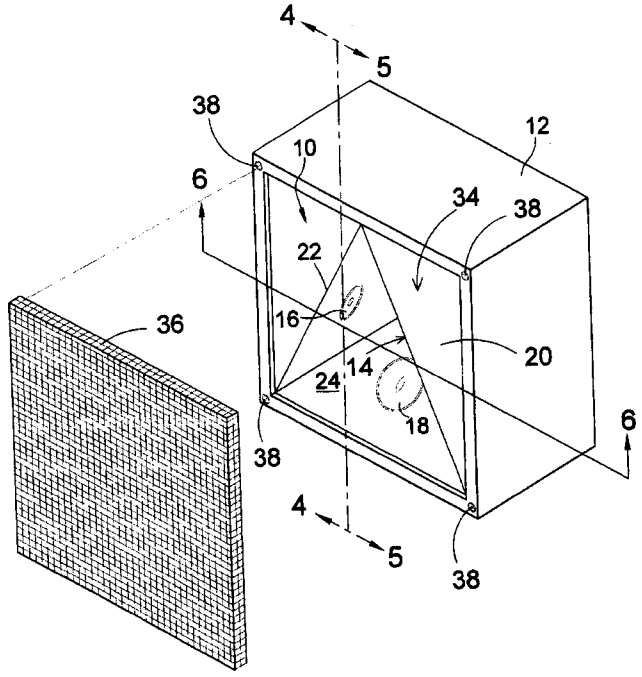
The invention is a kind of flextensional loudspeaker. Imagine a rectangular sheet of thin, flexible material (flexible, but not easily compressible). One edge is anchored to a rigid barrier. The other edge is driven uniformly along its length by one or more transducers. If the sheet is slightly bowed, and if the driving force is parallel to the general plane of the sheet, then a very small edge displacement will greatly increase or decrease the bowing of the sheet. The patent explains how this concept can be used to build a practical loudspeaker and includes frequency response graphs of an experimental unit. If the sheet is transparent, it can be mounted in front of a video screen.—GLA

6,721,431

43.38.Ja PRISMATIC LOUDSPEAKER/ MICROPHONE ARRAY

Dean E. Johnson, Stansbury Farmington, Michigan
13 April 2004 (Class 381/336); filed 24 January 2000

In the embodiment shown, a concave three-sided pyramid (the base is missing) is used to mount three loudspeakers or three microphones. This



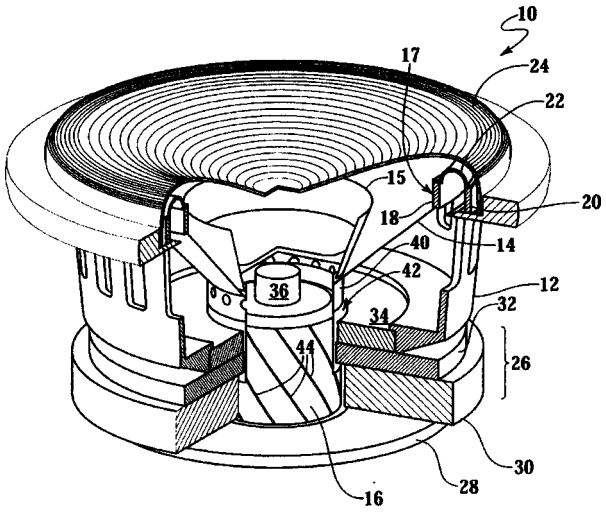
arrangement is said to result in ''a system which establishes positive phase control over the many and varied resonant characteristics encountered in the reproduction and presentation of audio energy.''—GLA

6,721,435

43.38.Ja ACOUSTIC LOUDSPEAKER WITH ENERGY ABSORBING BEARING AND VOICE COIL, AND SELECTIVE SOUND DAMPENING AND DISPERSION

Burton A. Babb and Alan J. Babb, assignors to Babb Laboratories
13 April 2004 (Class 381/407); filed 21 February 2001

The Babb sliding bearing loudspeaker is here improved through the addition of spiral ridges **44**. The voice coil former slides against the ridges,



minimizing friction and noise. Whizzer cone **15** and grill **24** are described in the patent text but not included in the claims.—GLA

6,721,436

43.38.Ja REMOTE EDGE-DRIVEN PANEL SPEAKER

Alejandro Bertagni and Eduardo Bertagni, assignors to Sound Advance Systems, Incorporated
13 April 2004 (Class 381/423); filed 29 March 2000

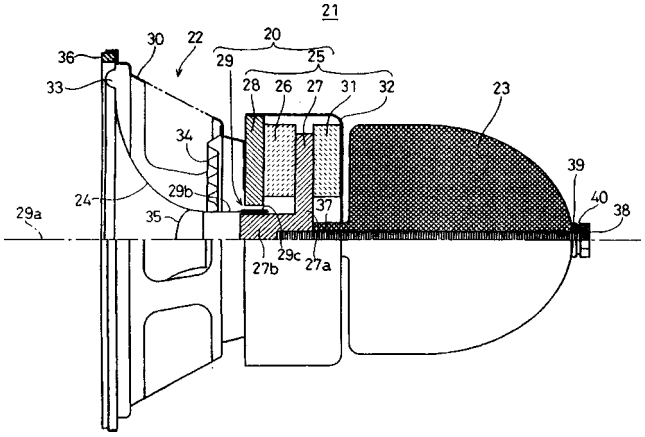
Like Unites States Patent 6,720,708 (reviewed above), this invention seems to be an edge-driven, bowed-panel loudspeaker. In this case, however, the preferred actuator is a moving coil motor, and the goal is to excite multiple bending waves. Such an arrangement lends itself to automotive applications. In fact, the invention defined in the patent claims is not a loudspeaker at all, but rather, ''A vehicle providing an audio environment for at least one person.''—GLA

6,724,909

43.38.Ja SPEAKER APPARATUS

Katsuhiko Tsumori *et al.*, assignors to Fujitsu Ten Limited; Timedomain Corporation
20 April 2004 (Class 381/412); filed in Japan 10 November 2000

What appears to be a transformer housing **23** is actually a solid weight



intended to act as an inertia sink. An active variant is also described at some length in the patent text but omitted from the two short claims.—GLA

6,726,336

43.38.Ja PROJECTOR HAVING A HINGED FRONT COVER WITH A LOUDSPEAKER

Naoya Matsuda *et al.*, assignors to Sony Corporation
27 April 2004 (Class 353/122); filed in Japan 12 January 2001

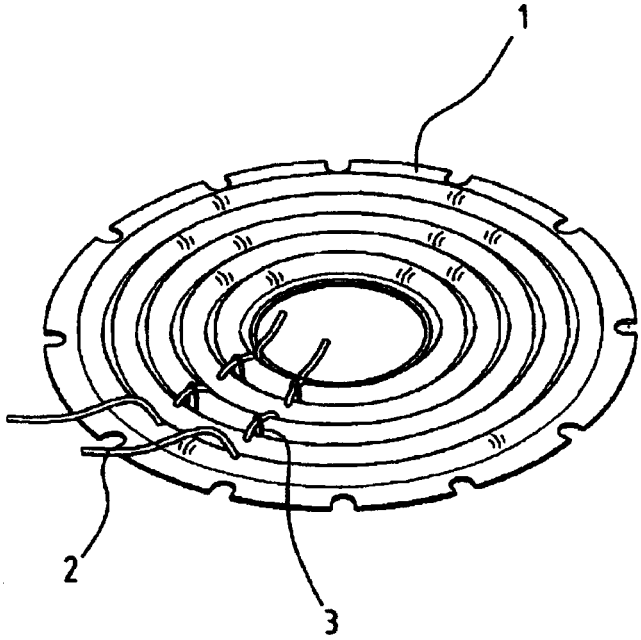
A portable LCD projector has the loudspeakers mounted in the cover to reduce the overall size of the projector enclosure. To further reduce the size of the enclosure, the controls to the projector are also mounted in the cover. The reduction in volume appears to be the sole purpose of the invention.—NAS

6,732,832

43.38.Ja SPEAKER DAMPER

Yen-Chen Chan, Taipei Hsien, Taiwan, Province of China
11 May 2004 (Class 181/166); filed 14 November 2001

The patent describes a ''complicated process and high cost'' method of weaving ''metallic whiskers'' **2** into a woven spider pad **1**, where the whiskers are fastened using a ''string'' **3**. ''Ruptures'' of the whisker can be



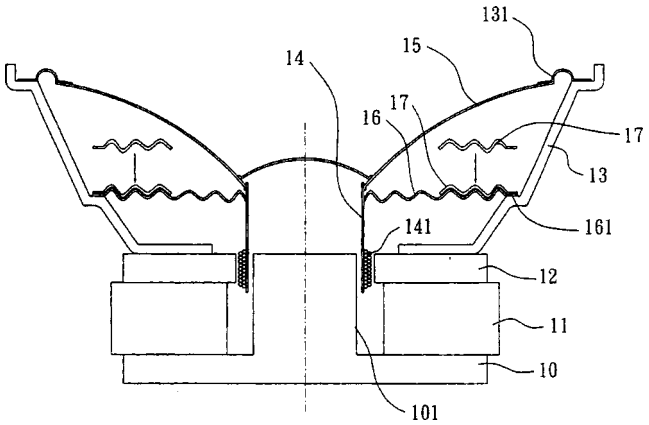
''significantly avoided,'' viewed and ''easily repaired'' and a phenomena described as ''jump-rope'' can be ''thoroughly eliminated.'' Labor costs being what they are, it may still be more cost effective to fix failures of ''whiskers'' in the current brute force fashion.—NAS

6,735,323

43.38.Ja SPEAKER

George Chang, assignor to Sun Technique Electric Company, Limited
11 May 2004 (Class 381/404); filed 30 January 2003

The patent describes a device where a second annular ring **17** is added to the spider **16**. The added annular ring can be either on the top or the bottom of the spider. The ''clear sound'' of the inner portion of the first annular corrugated plate **16** at low power levels and the ''clear sound'' of the outer portion of plate **16** with the help of the second annular corrugated plate **17** at high power levels is described, without any quantitative analysis. The



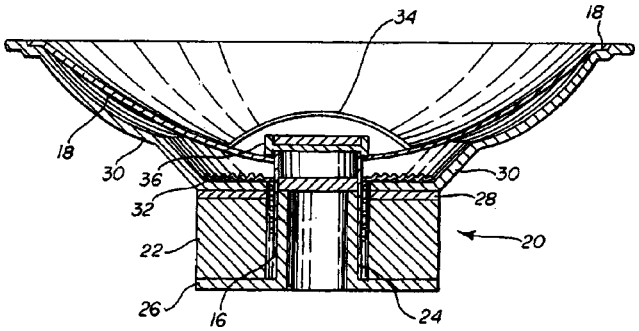
nonlinearities that may be introduced by this method of construction are not described.—NAS

6,738,490

43.38.Ja LOUDSPEAKER WITH INDEPENDENT MAGNETIC DAMPENING AND EXCURSION CONTROL

Eugene P. Brandt, West Paducah, Kentucky
18 May 2004 (Class 381/421); filed 11 January 2000

A magnetic motor is described with the top plate **28**, pole yoke **24**, and backplate **26** made from nonmagnetic materials. A small ring magnet (at the top of the pole piece and level with item **30**), that is mounted inside of the coil and rides up and down with the voice coil, interacts with a small coin magnet mounted either at the top plate side (inside **36**) or the backplate side (in another embodiment) of the yoke. These small magnets are mounted so that like poles face each other. Another embodiment is discussed where a



third damping magnet is introduced. All this is to prevent excessive excursion by the moving parts of the loudspeaker. Flux intensity for the damping magnets and other matters are discussed in prosaic terms as well as are 32 claims, but the motor strength of the primary magnet **22** using the novel nonmagnetic construction, the interaction of the magnetic fields from the primary magnets and damping magnets, and the effect on sensitivity, nonlinearity, etc., are not discussed.—NAS

6,725,110

43.38.Md DIGITAL AUDIO DECODER

Toshihiko Suzuki, assignor to Yamaha Corporation
20 April 2004 (Class 700/94); filed in Japan 26 May 2000

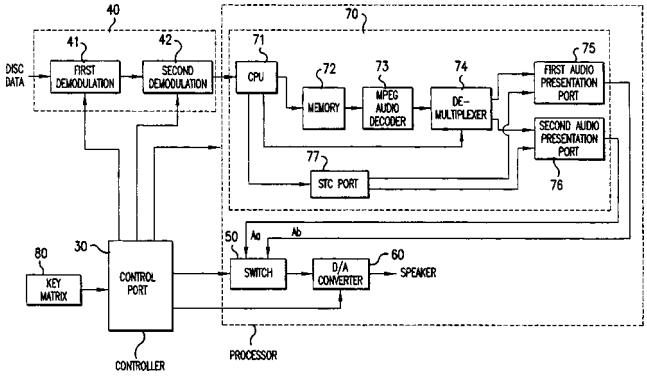
This patent quite simply shows how to mix down multiple MPEG-audio streams. Everything is obvious to those ''skilled in the art.'' The inventor missed Scott Levine's International Computer Music Conference paper in 1996 entitled ''Effects Processing on Subband Audio Data'' (three years before the patent was filed).—MK

6,714,494

43.38.Md APPARATUS AND METHOD FOR PROCESSING AUDIO SIGNALS RECORDED ON A MEDIUM

Han Jung and Ki Won Kim, assignors to LG Electronics Incorporated
30 March 2004 (Class 369/47.16); filed in the Republic of Korea 17 September 1993

This patent explains that the conventional karaoke audio-video format can encode two-channel stereo sound, or it can give the user a choice between mono with vocal and mono without vocal. A method is described for



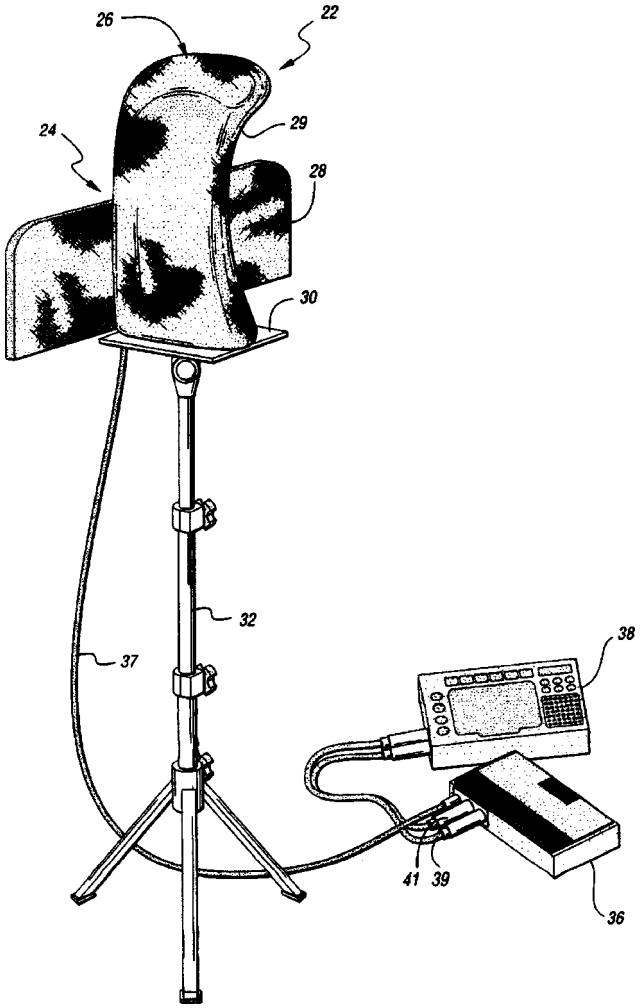
expanding the format to include stereo with vocal and stereo without vocal.—GLA

6,714,653

43.38.Md SOUND CAPTURING METHOD AND DEVICE

Douglas Peter Magyari, Royal Oak and David Keith Magyari, Madison Heights, both of Michigan
30 March 2004 (Class 381/26); filed 26 February 1998

One cannot fault this invention for lack of novelty. Its basic premise is that body vibration plays an important role in human hearing. The ''sound capturing device'' pictured is a kind of acoustic mannequin that includes a head portion **26** and a torso portion **28**. For stereo pickup, each of these



houses at least two transducer assemblies, and each assembly includes a condenser microphone to receive airborne sound plus a ''crystal microphone'' to pick up vibrations. Electrical signals from the various transducers are routed to a conventional audio mixer.—GLA

6,718,186

43.38.Md MELODY PLAYING SYSTEM

Eriko Aoki, assignor to NEC Corporation
6 April 2004 (Class 455/567); filed in Japan 28 January 2000

It is starting to dawn on the manufacturers of electronic gadgets that user-programmable features should be programmable by the average user. Suppose you purchase a cellular phone that allows you to create a melody to be used as a call signal. After going through the process, you find that the notes are correct but the tempo is too slow. To correct that seemingly minor flaw, you must start all over again. Designing a more friendly interface without exceeding the device's limited memory is not an easy task. This patent describes a melody-playing system that allows tempo to be changed independently of melody, yet requires only a modest increase in storage capacity.—GLA

6,727,421

43.38.Md REPRODUCING APPARATUS

Hidehito Izawa *et al.*, assignors to Kabushiki Kaisha Toshiba
27 April 2004 (Class 84/609); filed in Japan 21 January 2000

Essentially, this is a broad patent covering the use of a playlist in an MP3 player. The novelty of the scheme is difficult to discern since the exact format is not specified and, as a result, the description (while satisfying patent lawyers) is vague and useless.—MK

6,737,959

43.38.Md BAG WITH SOUND GENERATING MODULE

Chin-Lien Ho, Taichung City, Taiwan, Province of China
18 May 2004 (Class 340/384.1); filed 12 June 2001

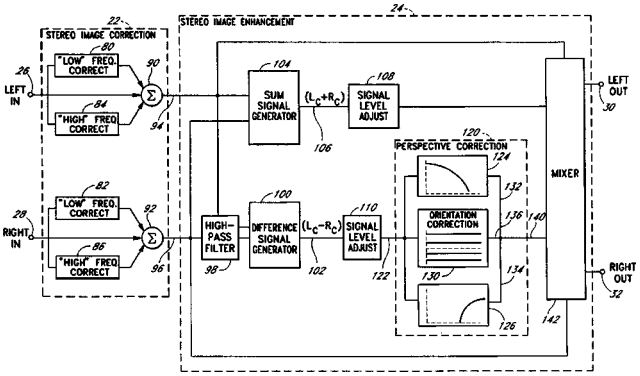
The ubiquitous sound chip finds yet another use—shopping bags. Coupled with wireless reception (not mentioned but fiendishly obvious), one can only imagine the possibilities: ''Buy more!,'' ''You're leaving the mall ALREADY?,'' and other exhortations.—MK

6,718,039

43.38.Vk ACOUSTIC CORRECTION APPARATUS

Arnold I. Klayman and Alan D. Kraemer, assignors to SRS Labs, Incorporated
6 April 2004 (Class 381/1); filed 9 October 1998

Although the invention is intended mainly for use in automotive stereo systems, other possible applications are described in the patent. In a ''stereo image correction'' section, high- and low-frequency bands are equalized separately. A ''stereo image enhancement'' section then makes use of the



inventor's earlier patents to generate an expanded stereo image. The patent is easy to follow, yet interested readers who scrutinize all 55 patent claims may find it difficult to decide exactly what has been patented here.—GLA

6,718,042

43.38.Vk DITHERED BINAURAL SYSTEM

David Stanley McGrath, assignor to Lake Technology Limited
6 April 2004 (Class 381/310); filed in Australia 23 October 1996

This patent can be filed under, ''Why didn't I think of that?'' Sophisticated virtual reality systems rely on head-tracking sensors to rotate the virtual sound field in response to movements of the user's head. The patent argues that simply assuming that the listener's head will turn slightly from time to time significantly improves the impression of spatialized sound sources. Moreover, in many cases, the head movements of typical listeners can be predicted in advance.—GLA

6,735,314

43.38.Vk EXPANDED STEREOPHONIC CIRCUIT WITH TONAL COMPENSATION

Alan Henderson Hoover, assignor to Thomson Licensing S.A.
11 May 2004 (Class 381/1); filed 13 May 2002

Modern commercial TV broadcasting normally includes two channels of audio that are often encoded with linear, program-directed cues to enhance spatial delineation by means of phasor reconstruction at the listener's ears. This field of development is a busy one, given that some transmission software standards are still in a state of flux. The patent specifically addresses the problems concerned with expanding the width of the stereo base angle by the addition of antiphase, frequency-shaped cross talk between the two channels. This alters the overall power content of the difference channel (L−R), which should logically be balanced by a conjugate adjustment in the sum channel (L+R).—JME

6,735,564

43.38.Vk PORTRAYAL OF TALK GROUP AT A LOCATION IN VIRTUAL AUDIO SPACE FOR IDENTIFICATION IN TELECOMMUNICATION SYSTEM MANAGEMENT

Pekka Puhakainen, assignor to Nokia Networks Oy
11 May 2004 (Class 704/246); filed in Finland 30 April 1999

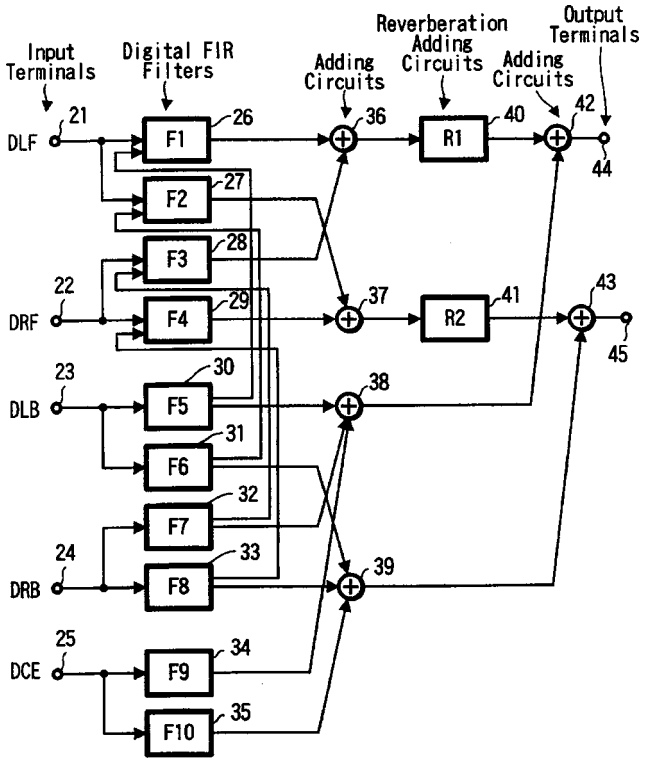
In certain mobile communications activities, a central dispatcher must keep track of a fairly large number of intercommunicating sources in the field. Traditionally, a visual monitor has been the dispatcher's chief aid in this activity. This patent proposes audible cues to reinforce the visual ones. Each source is given a specific location in stereo space, assigned arbitrarily by the dispatcher. Stereo (using two or more loudspeaker channels) or headphones can be used, depending on the complexity of the application.—JME

6,721,426

43.38.Vk SPEAKER DEVICE

Hirofumi Kurisu and Yuji Yamada, assignors to Sony Corporation; Keio University
13 April 2004 (Class 381/63); filed in Japan 25 October 1999

This ''speaker device'' is actually an audio processing circuit intended to make headphones sound like loudspeakers. Digital multiplication and the



addition of FIR filter outputs are used to efficiently generate long impulse responses.—GLA

6,731,327

43.40.Tm DYNAMIC STRUCTURAL COUPLING MECHANISM FOR REDUCING OPTICAL DEGRADATION IN VIBRATING ENVIRONMENTS

Thomas Joseph Kujawa *et al.*, assignors to Hypervision, Incorporated
4 May 2004 (Class 348/80); filed 24 November 1999

This patent relates generally to optics used in the process of testing semiconductors and integrated circuits. The optical system described here ''floats'' on springs relative to the item to be observed while it is in use, but when it is being adjusted it is clamped by means of an electromagnet to the stage holding the observed item.—EEU

6,732,033

43.40.Vn SEMI-ACTIVE SHOCK ABSORBER CONTROL SYSTEM

John A. LaPlante and William T. Larkins, assignors to Active Shock, Incorporated
4 May 2004 (Class 701/37); filed 13 January 2003

This shock absorber, which consists of a spring, a damper, and a control system that provides signals to adjust the damper valves, is considered to be semi-active because it includes no actuator that adds energy to the vibrating system. Sensors are used to determine the relative velocities and positions of the two ends of the spring, and the controller uses the sensors' signals to adjust the damping to near the critical value, to provide end stops without ''hard bottoming'' or ''hard topping,'' and to limit the transmitted force.—EEU

6,736,022

43.40.Tm VIBRATION REDUCTION GEAR STRUCTURE OF AN IMAGE INPUT AND OUTPUT DEVICE

Martin Chang and Thomas Sheng, assignors to Avision Incorporated
18 May 2004 (Class 74/439); filed in Taiwan, Province of China 10 August 2001

The smaller gear of a coaxial set is attached to the larger one via a resilient insert in the larger gear, so that torsional oscillations are attenuated.—EEU

6,725,968

43.55.Ti ACOUSTIC DOOR ASSEMBLY WITH CONTINUOUS CAM HINGE

Grant S. Quam *et al.*, assignors to Wenger Corporation
27 April 2004 (Class 181/287); filed 25 October 2002

An acoustical door assembly comprises a special insulated acoustic door, a frame, and a hinge to connect the door to the frame. In this case, the hinge is a cam-lift hinge that lifts the door when the door is opened, and sets the door down onto a perimeter gasket when the door is closed. The continuous hinge does not allow any discontinuity or sound leaks of the gasket along the jamb.—CJR

6,736,423

43.40.Tm APPARATUS AND METHOD FOR DAMPING VIBRATION OF A VEHICLE PART

Stepan S. Simonian *et al.*, assignors to TRW Vehicle Safety Systems Incorporated; TRW Incorporated
18 May 2004 (Class 280/731); filed 15 July 2002

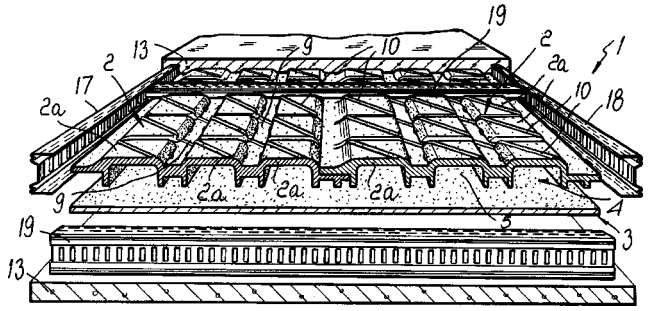
In this application of the well-known idea of damping by means of interacting particles, housings that are partially filled with particles are attached to components whose vibrations are to be damped. These particles dissipate energy as they collide with the inner walls of the housings and interact with each other.—EEU

6,729,093

43.55.Ti PREFABRICATED CONCRETE PANEL FOR INDUSTRIALIZED BUILDING WITH HIGH THERMAL AND/OR ACOUSTIC INSULATION

Sergio Zambelli and Benito Zambelli, both of Zanica, Italy
4 May 2004 (Class 52/405.1); filed in Italy 18 February 2000

In this invention, there are two layers of thermally insulating molded foamed polystyrene. One layer has cups and ridges; the other layer can be essentially flat. When the two layers are cupped together, they form interconnected air chambers. Concrete is then poured around these pockets, and



the result is a trapped internal air space that improves the noise reduction capability of the prefabricated panel, suitable for the exterior of an industrial building.—CJR

6,735,383

43.40.Vn VIBRATION CORRECTION APPARATUS, LENS APPARATUS, AND OPTICAL APPARATUS

Shinji Imada, assignor to Canon Kabushiki Kaisha
11 May 2004 (Class 396/55); filed in Japan 23 January 2001

A system for reducing image blurring due to vibration adjusts lens focus and aperture automatically on the basis of signals from a controller that acts on the basis of inputs from a vibration sensor.—EEU

6,725,150

43.58.Wc TURBO-CHARGED VEHICLE ENGINE SOUND SIMULATOR

Arthur Glandian, Glendale, California
20 April 2004 (Class 701/115); filed 7 February 2003

Once again (also see United States Patent 6,275,590 [reviewed in J. Acoust. Soc. Am. **111**, 2534 (2002)]), it is time to spoof the driving public with an imitation high-performance engine sound. In this case, it is a turbo-charger. Your Nash Rambler will never sound so good.—MK

6,727,419

43.58.Wc PULSATING METRONOME

Manuel Diaz, Columbus, Georgia
27 April 2004 (Class 84/484); filed 3 December 2002

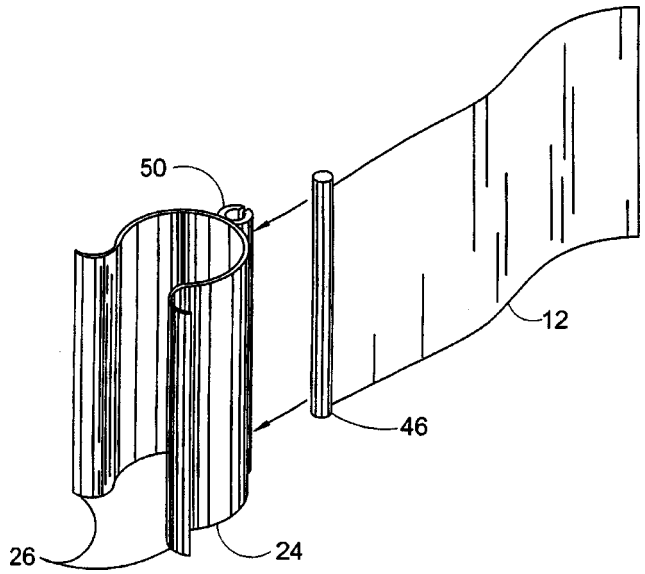
If Dick Tracy can have a two-way wrist radio, why not a wrist metronome? By adding a mechanical vibrator to signal the beat, the player can feel the pulse. What is more, with a strong enough vibration, the player can acheive rhythmic vibrato! The inventor neglected to think about how to synchronize multiple time pieces. More is the pity.—MK

6,733,399

43.58.Wc SPORTING EQUIPMENT AUDIBLE DEVICE

Lawrence J. Koncelik, Jr., East Hampton, LI, New York
11 May 2004 (Class 473/219); filed 24 April 2003

The inventor claims that the addition of a fluttering flag to a golf club or tennis racquet will encourage better form by only fluttering during impact. The aerodynamics of moving clubs and racquets is unfortunately not



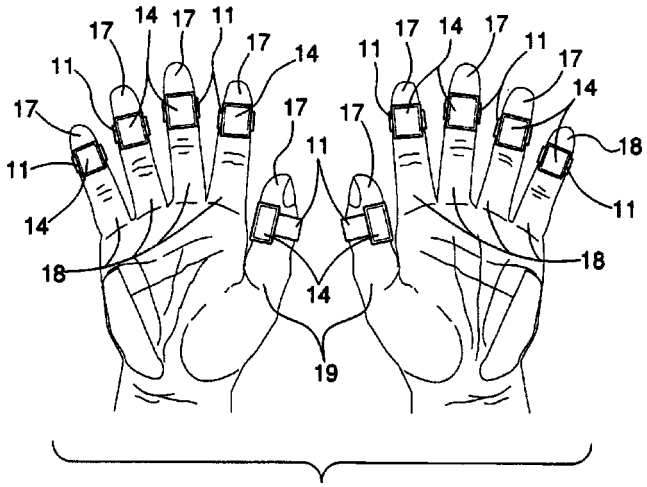
discussed, leaving the supposition in doubt. It is probably more annoying than useful.—MK

6,734,349

43.58.Wc FINGERTIP MUSICAL TAP ASSEMBLY

Conrad Adams, New York City, New York
11 May 2004 (Class 84/322); filed 21 March 2003

Do you like drumming your fingertips on the desktop? If so, then this invention is for you. Each finger gets a hard strap that can be adjusted for fat



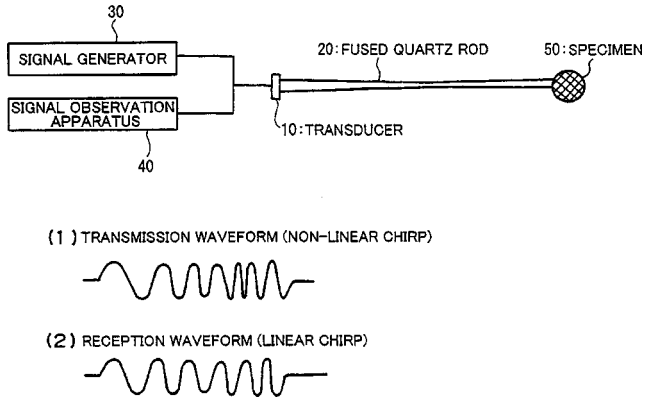
or thin fingers. The rest is up to you. However, fingernails are cheaper.—MK

6,730,029

43.60.Bf ULTRASONIC TRANSMITTER/RECEIVER BY PULSE COMPRESSION

Tadashi Moriya and Norio Tagawa, assignors to Japan Science and Technology Corporation
4 May 2004 (Class 600/437); filed in Japan 24 September 1999

An ultrasound transmission/reception apparatus, transmitting a signal with temporarily changing frequency and performing pulse compression on the received signal, is characterized by a single transducer and transmission line used for both transmitting and receiving the said ultrasound signal. The flexible waveguide transmission line is utilized as a transmission path and also serves as a delay medium to temporally separate the received signal and



a long-duration transmitted signal. A quartz rod with a narrowed center portion is an example of a usable transmission line. Side lobe suppression can be performed after pulse compression of a received signal by taking the correlation with an ideal compression waveform during further compression.—DRR

6,731,334

43.60.Mn AUTOMATIC VOICE TRACKING CAMERA SYSTEM AND METHOD OF OPERATION

Joonyoul Maeng and Errol R. Williams, assignors to Forgent Networks, Incorporated
4 May 2004 (Class 348/211.12); filed 31 July 1995

This complex patent deals with video conferencing environments in which camera assignments and audio cues must be coordinated. As the patent states, ''it is desirable in a video conferencing environment to provide automatic voice tracking of a speaker [talker] in order to control cameras such that there is natural camera movement in viewing a given speaker.'' The dominant principal here is beam forming via multiple microphones to identify the location of a speaker and to translate that information quickly into a corresponding reorientation of the most appropriate camera.—JME

6,726,628

43.60.Qv ANGLE-INDEPENDENT DOPPLER SYSTEM FOR SCREENING

David Vilkomerson, assignor to DVX, LLC
27 April 2004 (Class 600/454); filed 6 May 2003

The patent describes an apparatus and method for detecting the highest velocity of fluid flow within a volume that includes multiple flows with different velocities. The apparatus consists of multiple pairs of receiving transducers arranged about a region wherein each transducer in a pair is positioned on the same axis with respect to the region. A transmitting transducer is positioned to transmit into the fluid flow volume. This causes Doppler-shifted signals to be reflected, with the Doppler shift being a function of the fluid flow velocity. The receiving transducers receive the Doppler-shifted reflections. The transmitted and reflected signals are processed to yield a velocity signal for each pair of transducers. The maximum velocity yields the maximum fluid flow and the direction is determined by the axis of the receiving transducer pair.—DRR

6,735,317

43.66.Ts HEARING AID, AND A METHOD AND A SIGNAL PROCESSOR FOR PROCESSING A HEARING AID INPUT SIGNAL

Carsten Paludan-Mueller, assignor to Widex A/S
11 May 2004 (Class 381/317); filed 5 April 2002

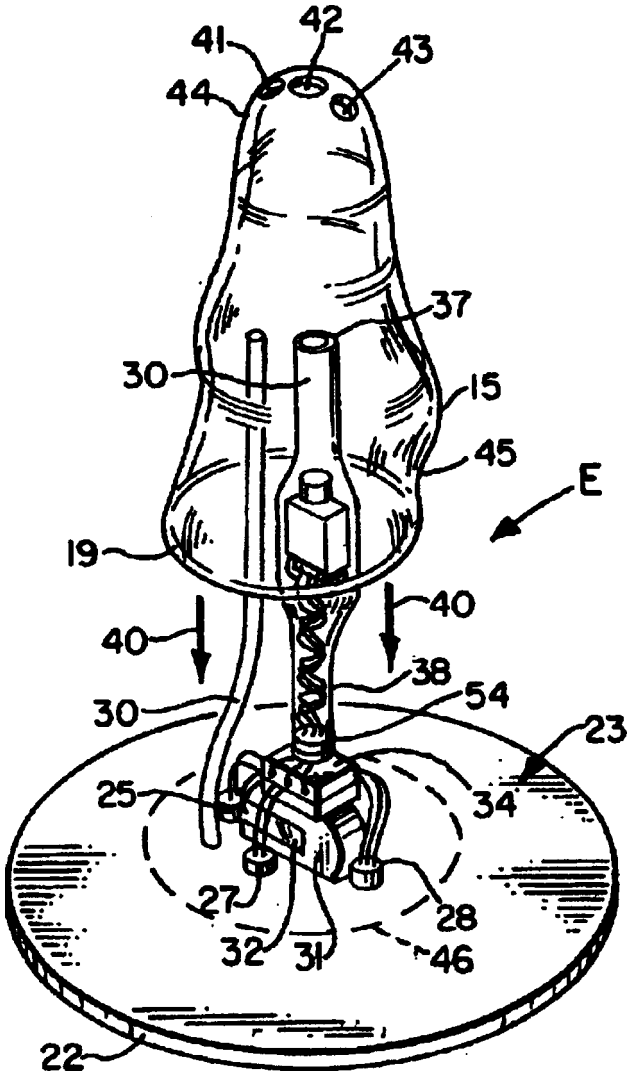
A noise attenuation algorithm estimates the level of background noise in a low-frequency band and at least one intermediate-frequency band. Based on the estimated noise level, gain is adjusted in these bands to squelch the noise. Similarly, if speech is thought to be present in at least one high-frequency band, the gain of at least one intermediate band is raised to enhance the speech components.—DAP

6,728,383

43.66.Ts METHOD OF COMPENSATING FOR HEARING LOSS

Roger P. Juneau *et al.*, assignors to Softear Technologies, L.L.C.
27 April 2004 (Class 381/322); filed 28 October 1998

A custom hearing aid consists of a rigid plate onto which is bonded a solid soft body encapsulating the electronic components. The shape of the



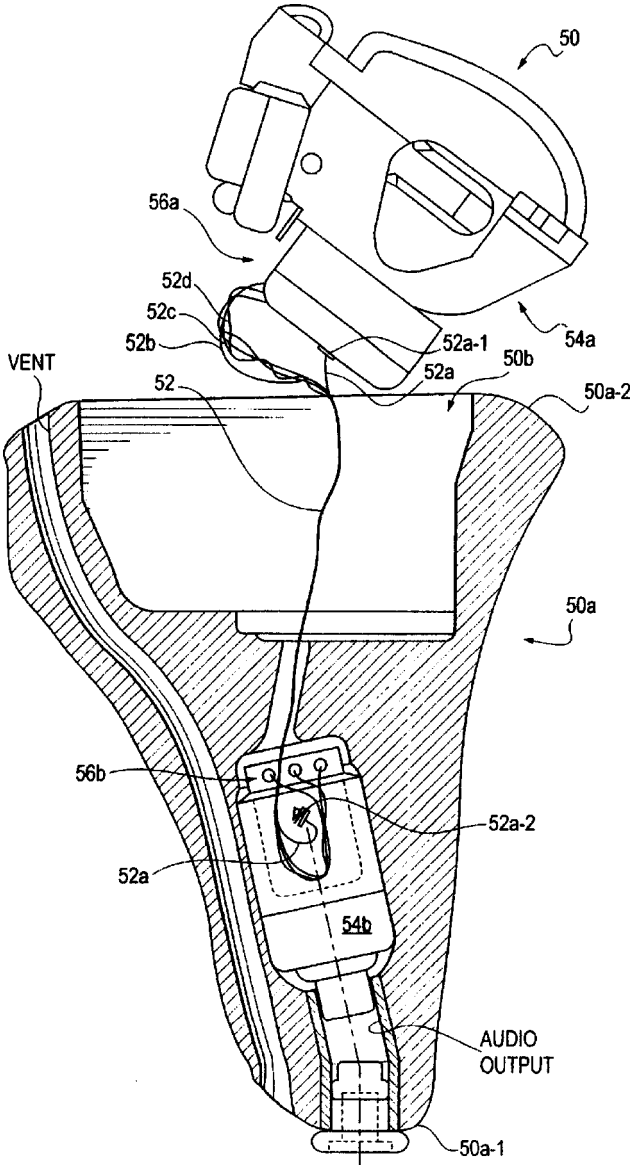
soft body conforms to changes in the ear canal of the wearer produced during jaw movement. Various stages of assembly methods are provided.—DAP

6,728,384

43.66.Ts HEARING AID CONNECTION SYSTEM

Paul R. Stonikas *et al.*, assignors to Beltone Electronics Corporation
27 April 2004 (Class 381/324); filed 27 June 2001

A nonvibration-transmitting wiring system is described for use in deformable or compressible hearing aids that change shape with changes in the ear canal geometry during, for example, talking and chewing. A nonstretch



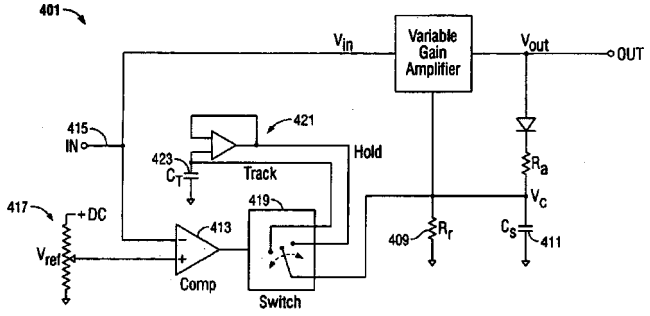
able thread or wire functions as a strain relief in combination with very flexible wires. Examples of fibers that may be used include high strength aramid-type and aromatic polyamide such as KEVLAR.—DAP

6,731,768

43.66.Ts HEARING AID HAVING SWITCHED RELEASE AUTOMATIC GAIN CONTROL

David J. Delage, assignor to Etymotic Research, Incorporated
4 May 2004 (Class 381/312); filed 26 July 2000

A switch controls the operation of a release circuit in an automatic gain control (AGC) function in a hearing aid. The release circuit is disabled and gain is maintained during low-level sounds at a compressed value resulting from the last high-level sound. When the input level exceeds a preset threshold, the release function is enabled and the AGC circuit behaves normally.—DAP

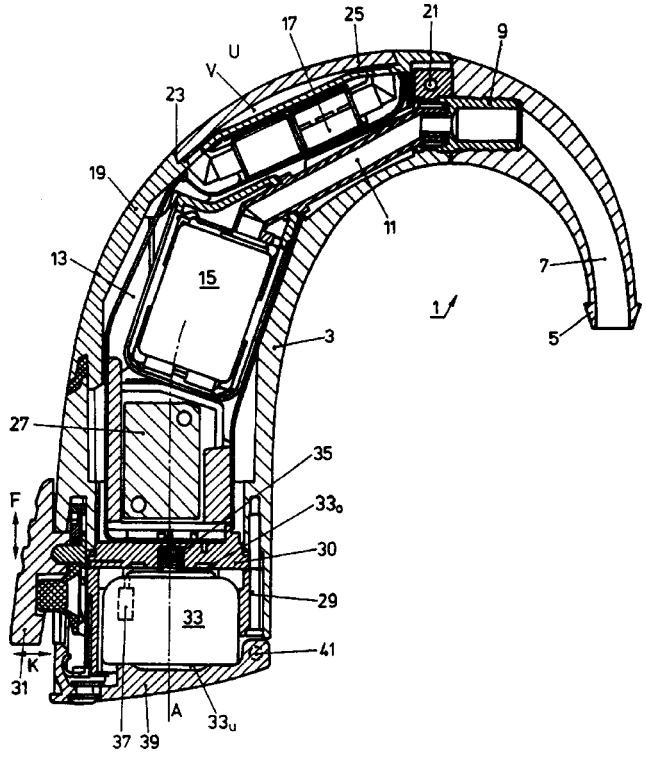


6,735,319

43.66.Ts BEHIND-THE-EAR HEARING AID

Andi Vonlanthen, assignor to Phonak AG
11 May 2004 (Class 381/330); filed in the World IPO 16 June 1999

Rather than the usual BTE hearing aid design with two case halves and the glue and screws and manual labor required to hold them together, a one-piece tubular housing is proposed. The tubular method is said to be particularly well suited for inserting and stacking components via automation.—DAP

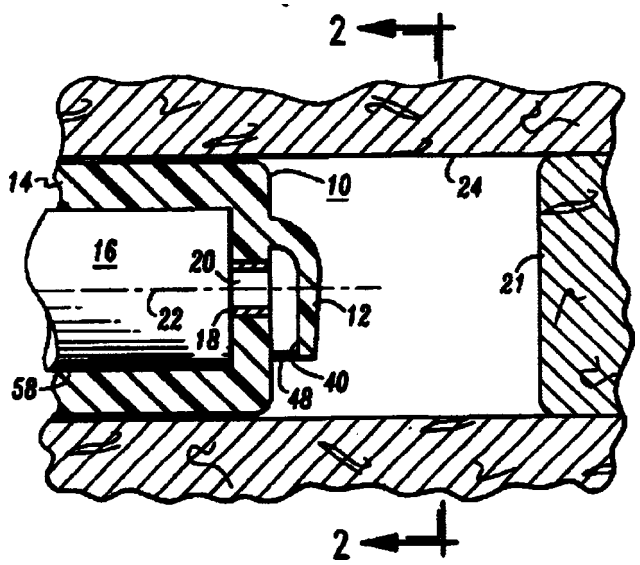


6,738,488

43.66.Ts PROTECTION AND SOLVENT WASHING OF IN-CANAL HEARING AIDS

Lawrence K. Baker, Gate City, Virginia
18 May 2004 (Class 381/325); filed 9 April 2002

Ear wax/fluid obstruction of hearing aid sound outlet tubes is a major problem. A hooded shield for the sound outlet of a hearing aid receiver is



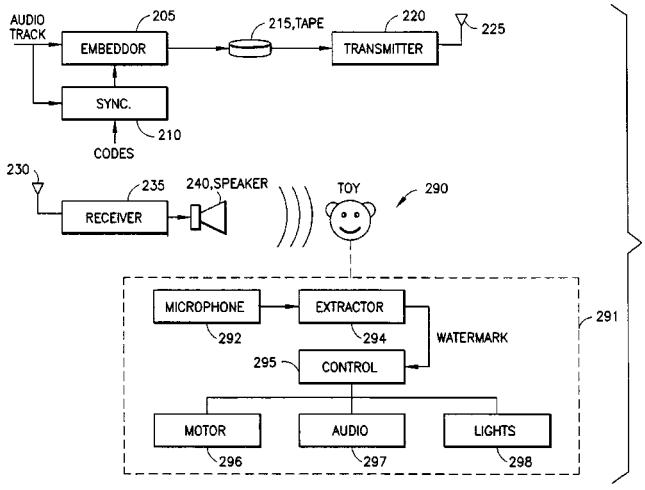
said to prevent fluids from entering the hearing aid from the ear canal. Solvent washing of the hearing aid is also recommended that may produce other problems.—DAP

6,737,957

43.72.Gy REMOTE CONTROL SIGNALING USING AUDIO WATERMARKS

Rade Petrovic and Joseph M. Winograd, assignors to Verance Corporation
18 May 2004 (Class 340/5.86); filed 16 February 2000

Hidden control information is embedded within an audio signal to remotely control a device. The detection of such a watermark may open a



time interval within which a user is allowed to respond, for example, by pressing a button. The detection time may be delayed if the bit error count of the watermark is too high.—DAP

6,681,208

43.72.Ja TEXT-TO-SPEECH NATIVE CODING IN A COMMUNICATION SYSTEM

Bin Wu and Fan He, assignors to Motorola, Incorporated
20 January 2004 (Class 704/260); filed 25 September 2001

The idea of this speech synthesis system is that, after a typical text-to-phoneme conversion, the synthesis can proceed directly from phonetic tables to vocoder codes, bypassing the additional steps of producing an audio stream and then processing that stream by a vocoder system. The central purpose, as expressed here, would be to use the existing vocoder tables, such as in a cell phone, eliminating any special synthesis tables. How well this would work is not clear.—DLR

6,684,187

43.72.Ja METHOD AND SYSTEM FOR PRESELECTION OF SUITABLE UNITS FOR CONCATENATIVE SPEECH

Alistair D. Conkie, assignor to AT&T Corporation
27 January 2004 (Class 704/260); filed 30 June 2000

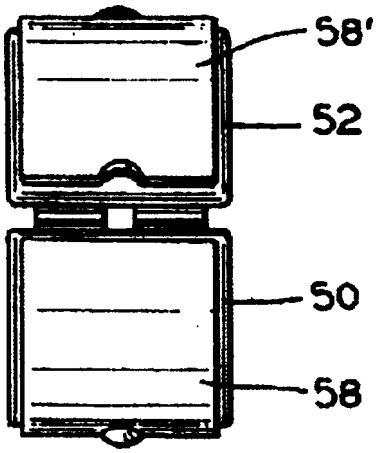
This seems to be a fairly traditional triphone synthesizer, except that somehow the process of locating the appropriate segment during the concatention operation has been sped up. Just how this is to be done is not entirely clear to this reader. A neighborhood of up to two phonemes on each side of the target is considered in evaluating the applicability of a particular segment.—DLR

6,687,675

43.72.Ja MESSAGE STORAGE DEVICE

Lurley Archambeau, Maumee, Ohio
3 February 2004 (Class 704/273); filed 31 May 2000

This is a high-tech version of the traditional keepsake locket, like the kind you pop open to reveal a tiny photo. Only here, you get not only an image, but a brief audio track, as well. The device may be in the form of a pendant, a bracelet, or some other form. Suggested uses include the usual



keepsake message, a medical reminder, or perhaps an important appointment reminder. The claims cover a device capable of holding a single message, but we can easily foresee the (obvious) next patent covering a device with a button to select one of several stored messages.—DLR

6,691,083

43.72.Ja WIDEBAND SPEECH SYNTHESIS FROM A NARROWBAND SPEECH SIGNAL

Andrew Paul Breen, assignor to British Telecommunications public limited company
10 February 2004 (Class 704/220); filed in the European Patent Office 25 March 1998

For good reasons, relatively little information is lost when a speech signal is bandwidth limited for transmission over a typical telephone line. Thus, it is not too surprising that most of the missing material can be fairly easily reinserted. Here, peaks are located for the available formants and a vocal tract shape model is used to improve the formant model accuracy. A

codebook entry is then located on the basis of the available formant frequencies, which supplies appropriate values for additional lower and/or higher formants that were supressed by the bandwidth reduction. There is no mention of the more perceptually significant, but much harder, task of restoring altered noise spectra.—DLR

6,694,123

43.72.Ja TALKING BOOK METHOD AND SYSTEM

Geoffrey Martindale and Suman Narayan, assignors to Texas Instruments Incorporated
17 February 2004 (Class 434/317); filed 7 December 2000

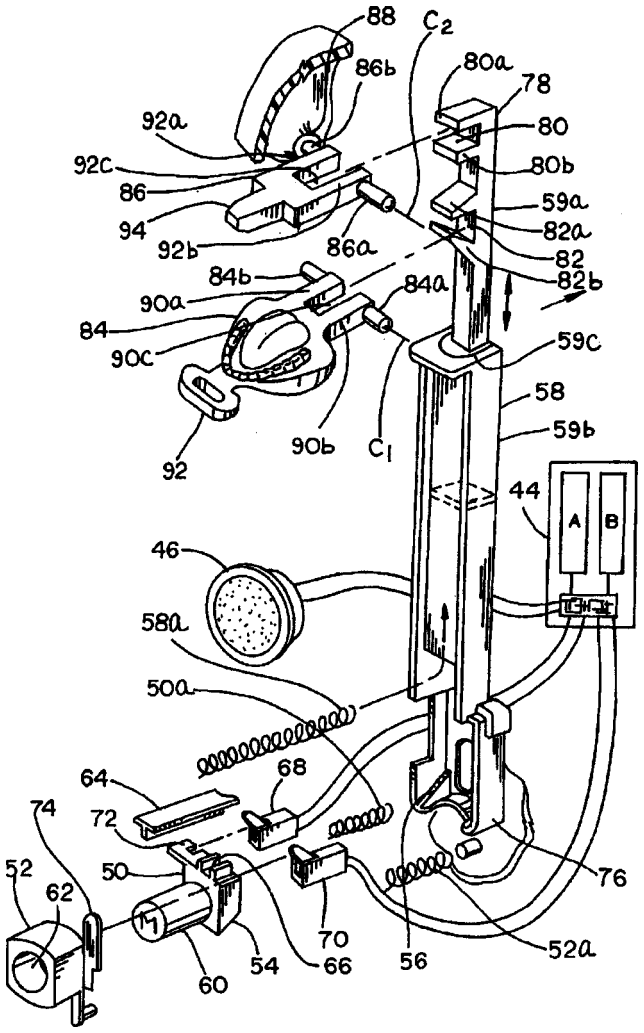
This patent is essentially a vehicle for the assignee to market a speech synthesis system, including a particular integrated circuit chip and an authoring software package geared toward organizing the text of a book for synthesis. The text is compressed using a method by which phrases are stored in a table and then represented by the table entry code. The software organizes and indexes chapters as sequences of phrase codes. Another software component encodes the phrases into a form suitable for the synthesis chip. Parts of the brief patent read like a software user's manual.—DLR

6,733,359

43.72.Ja TALKING ACTION FIGURE HAVING FACIAL EXPRESSIONS

Warren Leigh Jacobs, assignor to Hasbro, Incorporated
11 May 2004 (Class 446/300); filed 7 May 2003

An ''action figure'' is a doll by another name. Here, Hasbro illustrates how to add simple vocal gestures by moving the jaw up and down. As the computer animation industry knows, effective gestures are not that simple.—MK



6,681,206

43.72.Ne METHOD FOR GENERATING MORPHEMES

Allen Louis Gorin *et al.*, assignors to AT&T Corporation
20 January 2004 (Class 704/243); filed 18 October 2000

A method is described for collecting and analyzing speech utterances to build up a database of morphemes for recognition and understanding purposes. In fact, by significantly broadening the term ''morphemes,'' the system would include various nonverbal items, such as gestures, tablet strokes, body movements, mouse clicks, keypad entries, and DTMF codes, among many others, as well as the more commonly treated phones, diphones, words, grammars, etc. All of this information is collected by the recognizer system during normal use and applied to a semantic analysis of a phone dialer situation.—DLR

6,684,183

43.72.Ne GENERIC NATURAL LANGUAGE SERVICE CREATION ENVIRONMENT

Toby Korall *et al.*, assignors to Comverse Limited
27 January 2004 (Class 704/9); filed 6 December 1999

The purpose of this system is to simplify the process of creating recognition grammars for a variety of recognizer systems. In pursuit of this goal, the described system would collect exemplary speech phrases, coding these into a general Backus-Naur (BNF) format. Once a suitable grammar has been thus collected, tables would be consulted containing the specific coding details to convert the BNF grammar into the coding form required for a specific recognizer product.—DLR

6,691,088

43.72.Ne METHOD OF DETERMINING PARAMETERS OF A STATISTICAL LANGUAGE MODEL

Reinhard Blasig, assignor to Koninklijke Philips Electronics N.V.
10 February 2004 (Class 704/240); filed in Germany 21 October 1998

Sparse training data is a problem that typically occurs when training a large vocabulary speech recognizer using *n*-gram word sequence probablility models, because many of the possible *n*-gram patterns do not occur in the training data. The solution proposed here is that word classes would be formed based on semantic or syntactic features obtained from a lexicon. A representative *n*-gram is then formed when any word from a class occurs in an *n*-gram pattern. The brief patent goes into some detail covering the statistical effects of such a system.—DLR

6,727,418

43.75.St MUSICAL SCORE DISPLAY APPARATUS AND METHOD

Shuichi Matsumoto, assignor to Yamaha Corporation
27 April 2004 (Class 84/477 R); filed in Japan 3 July 2001

Instrumental players know to memorize when a begin repeat sign is seen so that they can frantically find the sign when they see the end repeat. How can an electronic/computer assisted score reader help? To begin with,

by displaying both the begin and ending measures. The patent discloses this and other mechanisms for resolving back references in score reading during performances. The patent writing is suitably turgid to give patents a bad name.—MK

6,727,417

43.75.Tv COMPUTERIZED MUSIC TEACHING INSTRUMENT

Dorly Oren-Chazon, Tel-Aviv, Israel
27 April 2004 (Class 84/470 R); filed 28 February 2002

In a patent reviewed earlier (United States Patent 6,337,434 [J. Acoust. Soc. Am. **112**, 23 (2002)]), the inventor proposed a set of mechanically controlled puppets for teaching solfège. Now, it has been updated to include a computer to enhance the instruction of preschool children—basically by expanding the already overloaded palette of musical metaphors.—MK

6,725,108

43.75.Wx SYSTEM AND METHOD FOR INTERPRETATION AND VISUALIZATION OF ACOUSTIC SPECTRA, PARTICULARLY TO DISCOVER THE PITCH AND TIMBRE OF MUSICAL SOUNDS

Shawn Anthony Hall, assignor to International Business Machines Corporation
20 April 2004 (Class 700/94); filed 28 January 1999

This remarkably readable and cogent patent addresses the problems of (a) logarithmic frequencies in Western scales, (b) timbre identification, and (c) missing fundamentals. The inventor proposes solving these conundra by (a) realizing a recursive design for the log FFT, (b) computing the normalized vector of overtone amplitudes, and (c) sorting the overtones. Admirably, the inventor is aware of the drawbacks, in particular timbre identification, but manages to put his best foot forward.—MK

6,728,664

43.75.Wx SYNTHESIS OF SONIC ENVIRONMENTS

Hesham Fouad, Arlington, Virginia
27 April 2004 (Class 703/2); filed 22 December 1999

Essentially, the inventor wants to apply neural-network-like techniques to sound environment generation. The idea of the probabilistic generation of sonic environments dates back at least to the 1950s with John Cage. Xenakis (and many others) extended this fundamental idea. It is a pity the inventor does not appear to be aware of this (much) earlier art.—MK

6,737,572

43.75.Wx VOICE CONTROLLED ELECTRONIC MUSICAL INSTRUMENT

John W. Jameson and Mark B. Ring, assignors to Alto Research, LLC
18 May 2004 (Class 84/741); filed 20 November 2001

The heart of the patent is an instrument that a naive player can play. The inventors propose a computer algorithm for pitch detection (with a kazoo input) followed by instrumental synthesis. A C++ code is sketched as well.—MK

6,730,349

43.80.Cs MECHANICAL AND ACOUSTICAL SUSPENSION COATING OF MEDICAL IMPLANTS

Marlene C. Schwarz and Stanley Tocker, assignors to SciMed Life Systems, Incorporated
4 May 2004 (Class 427/2.1); filed 1 March 2002

Coating of medical devices, such as catheters, stents, and the like, may be useful for facilitating local drug delivery, enhancing fluoroscopic visualization through the use of radiopaque materials, promoting biocompatibility of certain devices, or for improving surface properties such as slipperiness. In several embodiments of this apparatus, single or multiple coating materials are applied to medical devices by suspending the device and then coating at least a portion of the device. An example of such an apparatus includes a coating chamber, a vibrating structure within the coating chamber, means of suspending the device to be coated, and a coating source. In another embodiment, a medical device is moved into a coating area with a structure vibrating below the medical device. The vibration of the structure forces the device away from the vibrating surface, which results in coating the device.—DRR

6,726,904

43.80.Gx METHOD FOR TRANSMITTING ACOUSTIC WAVES

Werner Krauss *et al.*, assignors to Richard Wolf GmbH; TachnoGel GmbH & Company KG
27 April 2004 (Class 424/78.17); filed in Germany 9 October 2000

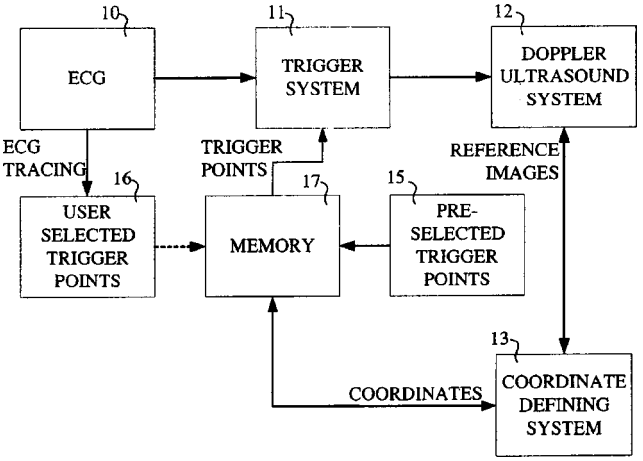
This patent refers to the use of a gel mass as a coupling medium for transmitting acoustic waves from a sound source into the body of a patient in order to minimize sound energy loss and to ensure high sound absorption of the ultrasound from the transducer irradiation surface. The gel mass is based on reaction products of the polyol components consisting of one or more polyols with hydroxyl numbers below 112. The isocyanate number of the reaction mixture lies in the range of 15 to 59.81 and the product of the isocyanate functionality and the functionality of the polyol components is at least 6.15, thus making it suitable as a coupling medium.—DRR

6,730,030

43.80.Qf METHOD AND APPARATUS FOR DETECTING ARTERIAL STENOSIS

Yoram Palti, Haifa, Israel
4 May 2004 (Class 600/441); filed 28 February 2002

The patent describes a high-speed apparatus for screening for arterial stenosis in which a Doppler ultrasound system automatically detects param-



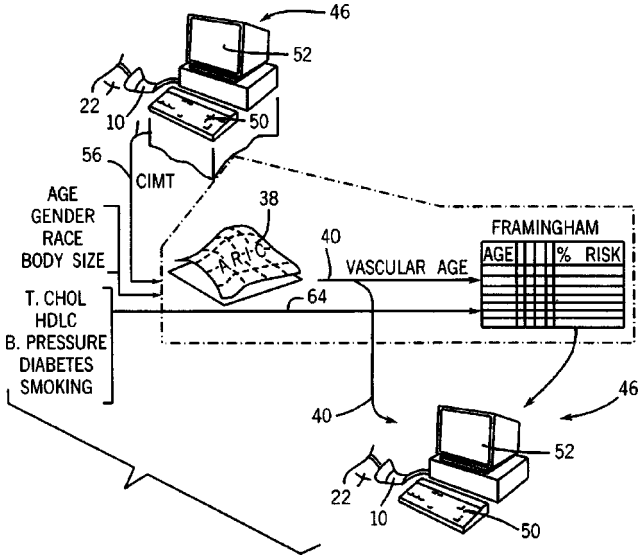
eters indicative of stenosis in the velocity profile of blood flow through the length of an artery without the need for analyzing visual imagery. The system automatically and repeatedly scans an area of interest to construct at least one velocity profile through segments of arteries within the area of interest. Each scan entails taking measurements of blood flow in the area of interest substantially simultaneously from different angles. The velocity and delta velocity profiles are analyzed automatically to determine a number of parameters that indicate potentially stenotic areas within the artery. These parameters and their changing values along the length of the artery are compared with predefined threshold values to establish potential stenotic areas. Scanning by the Doppler ultrasound system is triggered at a predefined point in an ECG cycle of the patient.—DRR

6,730,035

43.80.Qf ULTRASONIC APPARATUS AND METHOD FOR PROVIDING QUANTITATIVE INDICATION OF RISK OF CORONARY HEART DISEASE

James H. Stein, assignor to Wisconsin Alumni Research Foundation
4 May 2004 (Class 600/449); filed 7 November 2002

In this method, use is made of the fact that a quantitatively rigorous relationship between carotid intimadial thickness (CIMT) and risk of coronary heart disease (CHD) may be derived from existing Framingham Assessment data by converting CIMT into an equivalent ''vascular age'' that



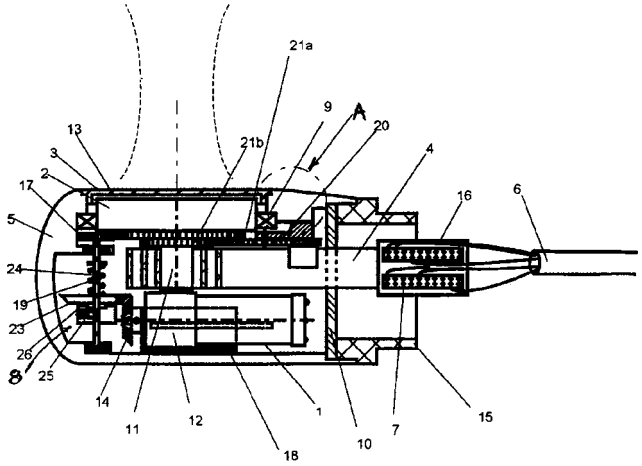
may be used in place of chronological age. The apparatus consists of a diagnostic machine having an ultrasonic transducer to be positioned near the carotid artery to obtain echo signals from that artery. The echo signal is processed to reveal the CIMT and to apply it to stored data relating CIMT of a population to a quantified risk of CHD. This quantified risk of heart disease constitutes the output.—DRR

6,733,457

43.80.Qf MOTORIZED MULTIPLANE TRANSDUCER TIP APPARATUS WITH TRANSDUCER LOCKING

Aimé Flesch and An Nguyen-Dinh, assignors to Vermon
11 May 2004 (Class 600/459); filed 11 June 2002

This ultrasonic endoscopic probe provides signals to an imaging system for transesophageal diagnosis. A removable housing is attached to the distal end of the endoscope. A ball-bearing-mounted phased array transducer within the housing, driven by an immersed micromotorized drive, rotates through multiple turns about its acoustic propagation axis. A torque-limiting gear box couples the drive to the transducer. A position encoder relays



transducer position information over 360° to the imaging system. An automatic transducer position locking/unlocking system is applied to lock the transducer in selected rotational positions.—DRR

6,730,048

43.80.Sh APPARATUS AND METHOD FOR ULTRASONIC MEDICAL DEVICE WITH IMPROVED VISIBILITY IN IMAGING PROCEDURES

Bradley A. Hare and Janniah S. Prasad, assignors to OmniSonics Medical Technologies, Incorporated
4 May 2004 (Class 601/2); filed 23 December 2002

A material of high radio opacity is used at one end of a catheter that vibrates ultrasonically in a standing wave pattern to remove tissue. The material allows the catheter to be visualized when the catheter is in the body. This improves the visability of the catheter tip during the procedure.—RCW

6,733,450

43.80.Sh THERAPEUTIC METHODS AND APPARATUS FOR USE OF SONICATION TO ENHANCE PERFUSION OF TISSUE

Andrei V. Alexandrov *et al.*, assignors to Texas Systems, Board of Regents
11 May 2004 (Class 600/439); filed 27 July 2000

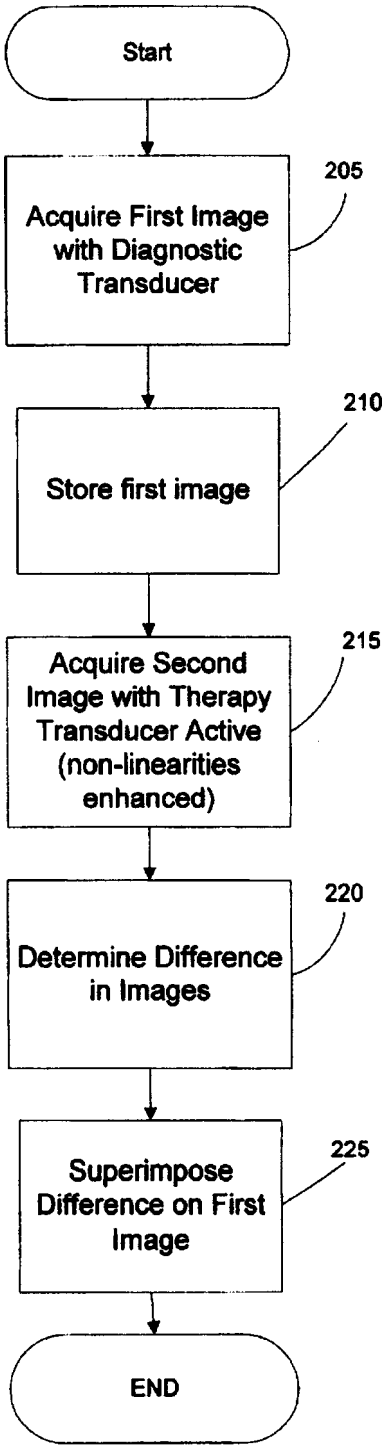
The method and its associated apparatus are designed to reduce tissue damage from ischemia by means of insonation through intact skin and/or bone to promote thrombolysis, reduce edema, and facilitate microcirculation, recanalization, increased collaterated interstitial flow, and delivery of lytic agents to clots located in supplying arteries, as well as nutrients or therapeutic agents to the ischemic tissue. An organ affected by ischemia or a portion of such an organ is exposed to low-frequency, low-power ultrasound generated by a multiple set of transducers of variable power limits which at least partially surround the organ, e.g., preferably at least 90% of a skull surface. Clinical studies have indicated that this approach maintains biological tissue function or viability in the induction of reduced perfusion by exposing the tissue or organ to ultrasound energy transmission.—DRR

6,726,627

43.80.Sh SYSTEM AND METHOD FOR ULTRASONIC HARMONIC IMAGING FOR THERAPY GUIDANCE AND MONITORING

Frederic Louis Lizzi and Cheri Xiaoyu Deng, assignors to Riverside Research Institute
27 April 2004 (Class 600/439); filed 24 January 2003

A method is provided to direct the application of an ultrasound system incorporating both therapeutic and diagnostic ultrasound transducers. The



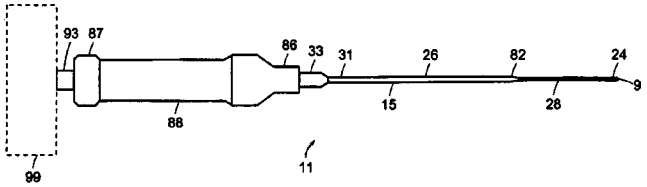
latter transducer is operated to acquire an ultrasound image and then both transducers are operated simultaneously over a time interval to yield a second image. The images are subtracted to obtain the difference that results from enhanced nonlinearities and propagation distortions induced by the high-intensity ultrasound therapy. A method is also provided to monitor the progress of the high intensity ultrasound indicative of changes due to *in situ* heating as well as permanent changes that result from cell microstructure alteration.—DRR

6,733,451

43.80.Sh APPARATUS AND METHOD FOR AN ULTRASONIC PROBE USED WITH A PHARMACOLOGICAL AGENT

Robert A. Rabiner *et al.*, assignors to OmniSonics Medical Technologies, Incorporated
11 May 2004 (Class 600/439); filed 25 March 2003

This device is essentially an ultrasonic probe that injects a pharmacological agent that enhances the ability of the probe to remove an occlusion. The pharmacological agent is released through a catheter to treat the occlusion and enhance the effect of a transverse ultrasonic vibration of the probe.



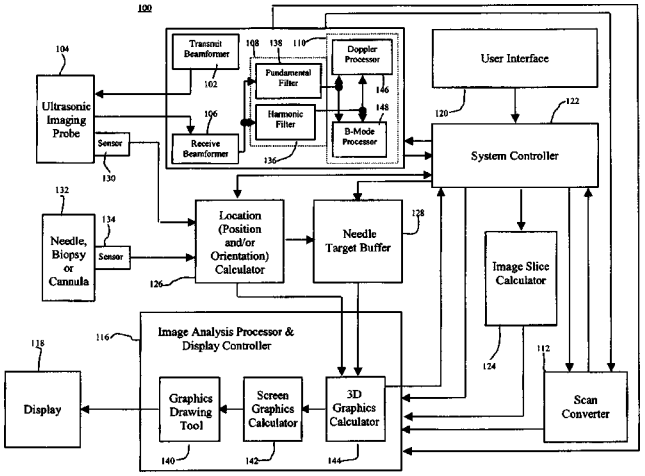
The pharmacological agent continues to travel downstream of the occlusion site and to work in conjunction with the probe to reduce the occlusion to a size that it can be easily removed from the body naturally in order to prevent reformation of the occlusion and other health risks.—DRR

6,733,458

43.80.Sh DIAGNOSTIC MEDICAL ULTRASOUND SYSTEMS AND METHODS USING IMAGE BASED FREEHAND NEEDLE GUIDANCE

Robert W. Steins *et al.*, assignors to Acuson Corporation
11 May 2004 (Class 600/461); filed 25 September 2001

A diagnostic medical ultrasound system incorporating an integrated invasive medical device guidance system is disclosed here. The guidance system obtains image slice geometry and other imaging parameters from the



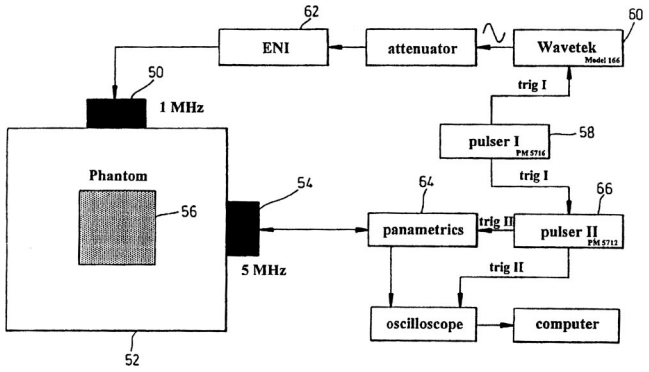
ultrasound system to optimize the guidance computations, the visual representation of the device, and the imaged portion of the subject. The ultrasound system also obtains guidance data that indicates the relative location, viz., the position and/or orientation of the medical device relative to the transducer and imaging plane in order to optimize the imaging plane and ultrasound beam characteristics.—DRR

6,726,629

43.80.Vj ULTRASOUND CONTRAST IMAGING

Peter J. A. Frinking *et al.*, assignors to Acuson Corporation
27 April 2004 (Class 600/458); filed in the United Kingdom 16 January 1998

An enhanced image is obtained using three pulse bursts of ultrasound. The first pulse burst has a low amplitude. The second pulse burst is at a different frequency and has a relatively higher amplitude that induces a



change in the contrast agent. The third pulse burst is at a third frequency and has a lower amplitude. Images from the first pulse burst and the third pulse burst are used to produce the enhanced image.—RCW

6,726,630

43.80.Vj ULTRASOUND DIAGNOSIS APPARATUS FOR IMAGING WITH A CONTRAST AGENT

Tetsuya Kawagishi, assignor to Kabushiki Kaisha Toshiba
27 April 2004 (Class 600/458); filed in Japan 8 November 2001

Two transmissions at a fundamental frequency are employed. The phase of the second transmission is opposite that of the first and the rates of the transmissions are different. A filter supresses a summed signal centered around a frequency that is twice the fundamental. After the suppression, an image is produced.—RCW

6,730,031

43.80.Vj EDITING AND TRIMMING APPARATUS AND METHOD FOR EXTENDED FIELD OF VIEW ULTRASONIC IMAGING SYSTEMS

Lingnan Liu and Lars Jonas Olsson, assignors to Koninklijke Philips Electronics, N.V.
4 May 2004 (Class 600/443); filed 21 September 2001

A sequence of partially overlapping images is acquired and stored. Spatial positions of the images are determined by correlation and the spatial positions are stored along with the images. A panoramic image is formed from the stored data. Images may be excluded from the panoramic image and the boundaries of the panoramic image may be trimmed.—RCW

6,730,032

43.80.Vj ULTRASONIC DIAGNOSTIC DEVICE AND IMAGE PROCESSING DEVICE

Masaki Yamauchi, assignor to Matsushita Electric Industrial Company
4 May 2004 (Class 600/443); filed in Japan 5 March 2001

Time-stamped images are stored. A contour within each image is found and associated with the image time stamp. From these contours, contours are interpolated at other times. A volume is determined using the contours.—RCW

6,730,037

43.80.Vj VASCULAR CATHETER HAVING LOW-PROFILE DISTAL END

Yue-Teh Jang, assignor to Cardiovascular Imaging Systems, Incorporated
4 May 2004 (Class 600/462); filed 7 March 2001

The distal region of this catheter has a reduced cross section compared to that of the proximal region. The proximal region contains two lumens, one that accomodates a guide wire and another that accomodates a rotatable element for imaging. The lumen in the distal region of the catheter is contiguous with both the lumens in the proximal region. The guide wire can be retracted from the distal region so that the lumen in the distal region is available for the rotating element.—RCW

6,733,452

43.80.Vj ULTRASOUND IMAGING METHOD AND APPARATUS FOR FORMING COLOR IMAGES BY USING A VARIABLE POWER THRESHOLD LEVEL

Moo Ho Bae and Ji Hoon Bang, assignors to Medison Company, Limited
11 May 2004 (Class 600/443); filed in the Republic of Korea 13 April 2001

Color Doppler images are formed using power threshold levels that depend on position and are determined using a noise signal, transmitted ultrasound signals, and received echoes.—RCW

6,733,453

43.80.Vj ELEVATION COMPOUNDING FOR ULTRASOUND IMAGING

Paul D. Freiburger and Bhaskar S. Ramamurthy, assignors to Siemens Medical Solutions USA, Incorporated
11 May 2004 (Class 600/447); filed 3 September 2002

Transducer arrays with multiple rows in elevation are employed to obtain image frames from essentially the same scan plane but with different spatial-frequency content in elevation. The frames are compounded to form an image with reduced speckle.—RCW

6,733,454

43.80.Vj AUTOMATIC OPTIMIZATION METHODS AND SYSTEMS FOR DOPPLER ULTRASOUND IMAGING

Müge M. Bakircioğlu and Bhaskar Ramamurthy, assignors to Siemens Medical Solutions USA, Incorporated
11 May 2004 (Class 600/453); filed 26 February 2003

Velocity scale, gain, baseline, and dynamic range are optimized based on Doppler data acquired at a standard or predetermined setting of the parameters and also based on the identification of an artifact in the Doppler data. The influence of the artifact on the determination of the parameter settings is then discarded or minimized.—RCW

6,733,455

43.80.Vj SYSTEM AND METHOD FOR ADAPTIVE CLUTTER FILTERING IN ULTRASOUND COLOR FLOW IMAGING

Larry Y. L. Mo *et al.*, assignors to Zonare Medical Systems, Incorporated
11 May 2004 (Class 600/454); filed 11 June 2002

An iterative algorithm is used to select a clutter filter for individual packets of color flow data. The cutoff frequency is chosen according to the frequency of the clutter. If significant clutter motion is present, a high-pass filter cutoff frequency is automatically set to suppress the clutter and associated flash artifacts. Low clutter frequencies result in low cutoff frequencies. The use of low cutoff frequencies for clutter allows the detection of low flows.—RCW

6,735,461

43.80.Vj FOCUSED ULTRASOUND SYSTEM WITH MRI SYNCHRONIZATION

Shuki Vitek *et al.*, assignors to Insightec-Txsonics Limited
11 May 2004 (Class 600/411); filed 19 June 2001

Magnetic resonance imaging (MRI) is used to execute a focused ultrasound procedure. A MRI system uses a timing sequence for transmitting radio frequency (rf) signals and detecting magnetic resonance response signals from a patient's body in response to the rf pulse sequences. A piezoelectric transducer is driven in such a manner that the transducer emits acoustic energy toward a target tissue region within the body. The parameters of the drive signals are changed during the timing sequence so as to minimize interference with the MRI system detecting magnetic response signals, e.g., during transmission of rf signals by the MRI system.—DRR

6,736,781

43.80.Vj ULTRASOUND IMAGING OF BREAST TISSUE USING ULTRASOUND CONTRAST AGENT

Roberta Lee, assignor to Manoa Medical, Incorporated
18 May 2004 (Class 600/458); filed 11 June 2002

A duct is identified using ultrasonic imaging. An ultrasound contrast agent is injected into the duct through an orifice in the nipple or through the duct wall. Ultrasonic images are obtained after the contrast is injected.—RCW

6,734,672

43.80.Vj METHOD FOR THE AUTOMATIC MEASUREMENT OF ACOUSTIC RESONANCE OF A MAGNETIC RESONANCE TOMOGRAPHY APPARATUS

Thorsten Feiweier, assignor to Siemens Aktiengesellschaft
11 May 2004 (Class 324/309); filed in Germany 2 May 2002

This is a method for determining acoustic resonance in a magnetic resonance tomography (MRT) system. A resonance measurement is implemented by applying a number of alternating gradient pulses that feature a fixed time interval relative to one another, applying an excitation pulse, and obtaining one or more magnetic resonance (MR) signals. The MR signal of the resonance measurement is evaluated with respect to at least one parameter characterizing the acoustic resonance of the MRT system. These steps are repeated with variation of the time spacing of the gradient pulses. A resonance curve is formed on the basis of the value of a characteristic parameter of the resonance measurement as a function of the varied time spacing. The resonance or resonances of the MRT system are established from the resonance curve.—DRR

6,736,780

43.80.Vj SYNTHETIC APERTURE FOCUSING METHOD FOR ULTRASOUND IMAGING BASED ON PLANAR WAVES

Tai Kyong Song and Jin Ho Chang, assignors to Medison Company, Limited
18 May 2004 (Class 600/447); filed in the Republic of Korea 31 December 2001

Unfocused waves are transmitted with a propagation angle that corresponds to the center position of a receive aperture. Echo signals are received in a subaperture, stored in a memory, and dynamically focused. The dynamically focused signals are used to form a beam for imaging.—RCW

Time-domain equations for sound propagation in rigid-frame porous media (L)

D. Keith Wilson[a)]
U.S. Army Engineer Research and Development Center, 72 Lyme Road, Hanover, New Hampshire 03755-1290

Vladimir E. Ostashev
NOAA Environmental Technology Laboratory, Boulder, Colorado 80305 and Physics Department, New Mexico State University, Las Cruces, New Mexico 88003

Sandra L. Collier
U.S. Army Research Laboratory, AMSRD-ARL-CI-EE, 2800 Powder Mill Road, Adelphi, Maryland 20783-1197

(Received 17 February 2004; revised 30 June 2004; accepted 6 July 2004)

A general set of time-domain equations describing linear sound propagation in a rigid-frame, gas-saturated porous medium is derived. The equations, which are valid for all frequencies, are based on a relaxational model for the viscous and thermal diffusion processes occuring in the pores. The dissipative terms in the equations involve convolutions of the acoustic fields with the impulse response of the medium. It is shown that the equations reduce to previously known results in the limits of low and high frequencies. Alternative time-domain equations are also derived based on a Padé approximation. [DOI: 10.1121/1.1785691]

PACS numbers: 43.20.Bi, 43.55.Ev [ANN] Pages: 1889–1892

I. INTRODUCTION

Most treatments of wave propagation and dissipation in porous media are formulated in the frequency domain. But interest in time-domain models has grown recently due to potential applications such as finite-difference, time-domain (FDTD) simulation of sound propagation in the presence of an absorbing surface[1,2] and determination of material properties with pulses.[3] Previous authors have developed time-domain formulations that are applicable either to low frequencies (e.g., Morse and Ingard[4]) or high frequencies (Fellah and Depollier[5]). In this letter, we derive a set of coupled integro-differential equations for time-domain propagation of linear sound waves that are valid for all frequencies. The derivation is based on the conceptual framework of viscous and thermal relaxations in rigid-frame porous media.[6,7] The resulting equations are shown to reduce to previously derived ones in the limiting cases of low and high frequencies.

II. FREQUENCY-DOMAIN RELAXATION EQUATIONS

The relaxation model for sound propagation in porous media is based on the concept that temperature perturbations introduced into the gas saturating the porous medium relax over time to the temperature of the frame material, as the frame generally has a much higher heat capacity than the saturating gas. Similarly, a change to the pressure gradient across the porous medium will induce a variation in the flow velocity that relaxes to a steady state determined by the permeability of the frame. Based on this conceptual framework, the following equations analogous to Euler's and mass conservation were proposed:[6,7]

$$-i\omega\mathbf{W}=-\tilde{V}\nabla P, \tag{1}$$

and

$$-i\omega\tilde{\beta}P=-\nabla\cdot\mathbf{W}, \tag{2}$$

where $\mathbf{W}$ is the acoustic particle velocity (averaged over a cross section normal to the nominal propagation direction and much larger than the size of all pores), P is the acoustic pressure, ω is the angular frequency, $\tilde{V}$ is the complex specific volume operator, and $\tilde{\beta}$ is the complex compressibility operator. The $\exp(-i\omega t)$ time convention is used here and uppercase letters indicate the temporal Fourier transforms (frequency-domain) velocity and pressure fields. The complex operators are given by[7]

$$\tilde{V}=V_\infty[1-S(\omega\tau_{\mathrm{vor}})], \tag{3}$$

$$\tilde{\beta}=\beta_\infty[1+(\gamma-1)S(\omega\tau_{\mathrm{ent}})], \tag{4}$$

where $V_\infty=\Omega/\rho q^2$, $\beta_\infty=\Omega/P\gamma$, Ω is the porosity, q is the tortuosity ($=\sqrt{\alpha_\infty}$ in the notation of some papers such as Refs. 8 and 9), ρ is the ambient air density, P is the ambient air pressure, γ is the ratio of specific heats, S is the relaxation function, and the τ's are characteristic relaxation times. The subscript "vor" indicates the vorticity (viscous) mode wave field and "ent" the entropy (heat conduction) mode.[10] The following simple phenomenological equation for the relaxation function provides a physically realistic approximation at all frequencies for the actual relaxation response of porous media:

a)Electronic mail: d.keith.wilson@erdc.usace.army.mil

$$S(\omega\tau)=\frac{1}{\sqrt{1-i\omega\tau}}. \tag{5}$$

The relaxation times may be "tuned" to fit the behavior of the porous medium well in a particular frequency range of interest, although similar values are usually obtained regardless of the frequency range of the tuning.[6] The values $\tau_{\mathrm{vor}}=2\rho q^2/\sigma\Omega$ and $\tau_{\mathrm{ent}}=N_{\mathrm{Pr}}s_{\mathrm{B}}^2\tau_{\mathrm{vor}}$ (where σ is the static flow resistivity, N_{Pr} is the Prandtl number, and s_{B} a pore shape factor) were shown in Ref. 7 to provide a relaxation model that is nearly indistinguishable from the pore microstructural model developed by Allard *et al.*[11] More complicated relaxation functions than Eq. (5), with additional free parameters, can be used if desired to produce a more accurate model for a given porous material.[7]

Equations (1)–(5) are similar to empirical models suggested by Johnson *et al.*[8] and Allard and Champoux.[9] Like the relaxation model, the others use radicals of the form $\sqrt{1-i\omega\tau}$ to interpolate in a simple way between the correct asymptotic behavior at low and high frequencies. However, the functional forms of the complex operators are different. Equations (1)–(5) were developed through comparison to known analytical solutions for simplified pore geometries and a heuristic solution of the diffusion equation for the vorticity and entropy modes in a porous medium.[6,7] The differences between the relaxation model and the Johnson *et al.* and Allard and Champoux equations appear to be important when undertaking the development of a time-domain model. As will be shown in the next section, the relaxation model readily transforms to the time domain without modification. We were unable to similarly transform the Johnson *et al.* and Allard and Champoux models.

III. CONVERSION OF THE EULER EQUATION TO THE TIME DOMAIN

Substituting Eqs. (3) and (5) into Eq. (1), one has

$$-i\omega\mathbf{W}=-V_\infty\left(1-\frac{1}{\sqrt{1-i\omega\tau_{\mathrm{vor}}}}\right)\nabla P. \tag{6}$$

When the inverse Fourier transform is applied to Eq. (6), the $-i\omega$ multiplying $\mathbf{W}$ on the left side of the equation becomes the time derivative. On the right side, the product of the relaxation function and pressure gradient becomes a convolution between the inverse transforms of these quantities. Thus, we have in the time domain,

$$\frac{\partial\mathbf{w}}{\partial t}=-V_\infty\nabla p+V_\infty\mathcal{F}^{-1}\left(\frac{1}{\sqrt{1-i\omega\tau_{\mathrm{vor}}}}\right)*\nabla p, \tag{7}$$

where the lowercase $\mathbf{w}$ and p indicate time-domain fields, $\mathcal{F}^{-1}$ is the inverse Fourier transform, and $*$ indicates convolution. The desired inverse transform is [Ref. 12, Eq. (3.382.6)]

$$s(t)=\mathcal{F}^{-1}\left(\frac{1}{\sqrt{1-i\omega\tau}}\right)=\frac{1}{\sqrt{\pi\tau t}}\exp\left(-\frac{t}{\tau}\right)H(t), \tag{8}$$

in which $H(t)$ is the Heaviside function (0 for $t<0$ and 1 for $t\geqslant 0$). The function $s(t)$ can be interpreted as the response of the medium to an impulsive change in the pressure gradient (or the pore temperature).[7] With the transformed response function, Eq. (7) becomes

$$\frac{\partial\mathbf{w}}{\partial t}=-V_\infty\nabla p+\frac{V_\infty}{\sqrt{\pi\tau_{\mathrm{vor}}}}\int_{-\infty}^{t}\frac{1}{\sqrt{t-t'}}\times\exp\left(-\frac{t-t'}{\tau_{\mathrm{vor}}}\right)\nabla p(t')\,dt'. \tag{9}$$

When τ_{vor} is very short in comparison to the time scale over which ∇p varies, the latter may be regarded as constant under the convolution integral. The integral then evaluates as $\nabla p(t)\sqrt{\pi\tau_{\mathrm{vor}}}$, and hence the second term on the right side of this equation becomes $+V_\infty\nabla p$, thereby cancelling the first term. This approximate cancellation of the two terms can be undesirable in calculations.

Alternatively, we could multiply both sides of Eq. (6) by the complex density, defined as $\tilde{\rho}=\tilde{V}^{-1}$. With some algebraic manipulation, one can show

$$\tilde{\rho}=V_\infty^{-1}\frac{\sqrt{1-i\omega\tau_{\mathrm{vor}}}}{\sqrt{1-i\omega\tau_{\mathrm{vor}}}-1}=V_\infty^{-1}\left(\frac{1-i\omega\tau_{\mathrm{vor}}}{-i\omega\tau_{\mathrm{vor}}}\right)\left(1+\frac{1}{\sqrt{1-i\omega\tau_{\mathrm{vor}}}}\right). \tag{10}$$

Therefore, instead of Eq. (6) we have

$$\left(1+\frac{1}{\sqrt{1-i\omega\tau_{\mathrm{vor}}}}\right)\left(\frac{\mathbf{W}}{\tau_{\mathrm{vor}}}-i\omega\mathbf{W}\right)=-V_\infty\nabla P. \tag{11}$$

Applying the inverse transform to this equation yields

$$\frac{\mathbf{w}}{\tau_{\mathrm{vor}}}+\frac{\partial\mathbf{w}}{\partial t}+\frac{1}{\sqrt{\pi\tau_{\mathrm{vor}}}}\int_{-\infty}^{t}\frac{\mathbf{w}(t')/\tau_{\mathrm{vor}}+\partial\mathbf{w}(t')/\partial t'}{\sqrt{t-t'}}\times\exp\left(-\frac{t-t'}{\tau_{\mathrm{vor}}}\right)dt'=-V_\infty\nabla p. \tag{12}$$

Equations (9) and (12) are equivalent, but one form or the other may be more convenient in a given application.

IV. CONVERSION OF THE MASS-CONSERVATION EQUATION TO THE TIME DOMAIN

A time-domain equation for mass conservation is developed by substituting Eqs. (4) and (5) into Eq. (2):

$$-i\omega\beta_\infty\left(1+\frac{\gamma-1}{\sqrt{1-i\omega\tau_{\mathrm{ent}}}}\right)P=-\nabla\cdot\mathbf{W}. \tag{13}$$

Transformation to the time domain then yields

$$\beta_\infty\frac{\partial p}{\partial t}+\beta_\infty(\gamma-1)\mathcal{F}^{-1}\left(\frac{1}{\sqrt{1-i\omega\tau_{\mathrm{ent}}}}\right)*\frac{\partial p}{\partial t}=-\nabla\cdot\mathbf{w}. \tag{14}$$

Applying Eq. (8), we find

$$\beta_\infty \frac{\partial p}{\partial t} + \frac{\beta_\infty(\gamma-1)}{\sqrt{\pi\tau_{\text{ent}}}} \int_{-\infty}^{t} \frac{\partial p(t')/\partial t'}{\sqrt{t-t'}} \exp\left(-\frac{t-t'}{\tau_{\text{ent}}}\right) dt' = -\nabla\cdot\mathbf{w}. \tag{15}$$

Unlike the Euler equation, the mass-conservation equation does not appear to possess two equivalent forms. Algebraically, the presence of $\gamma-1$ in Eq. (13) [instead of the -1 present in Eq. (6)] prevents a factorization analogous to Eq. (10). On the other hand, the approximate cancellation of terms discussed in connection with Eq. (9) does not occur in Eq. (15).

V. COMPARISONS WITH PREVIOUS RESULTS

Equations (9), (12), and (15) are the main new results of this letter. They provide time-domain equations for sound propagation in rigid porous media that apply in all frequency ranges of the acoustic disturbance. In this section, we compare the new results to previous ones for low or high frequencies.

Let us first consider Eq. (12) at a low frequency, specifically $\omega\tau_{\text{vor}} \ll 1$. At such a frequency, $\mathbf{w}(t)$ varies little over the relaxation time. The exponential factor under the convolution integral in Eq. (12) quickly decays over the acoustic period. Therefore we may consider solving the integral approximately by expanding $\mathbf{w}(t')$ and $\partial\mathbf{w}(t')/\partial t'$ in a Taylor series about t. Taking $\mathbf{w}$ to be on the order of unity $[O(1)]$, $\tau_{\text{vor}}\partial\mathbf{w}/\partial t$ would be $O(\omega\tau_{\text{vor}})$, etc. We have the approximations, to $O(\omega\tau_{\text{vor}})$,

$$\mathbf{w}(t') \simeq \mathbf{w}(t) + (t'-t)\frac{\partial\mathbf{w}(t)}{\partial t}, \qquad \tau_{\text{vor}}\frac{\partial\mathbf{w}(t')}{\partial t'} \simeq \tau_{\text{vor}}\frac{\partial\mathbf{w}(t)}{\partial t}. \tag{16}$$

When these expansions are substituted into the convolution integral in Eq. (12), it is readily evaluated as $\mathbf{w}(t)/\tau_{\text{vor}} + (\frac{1}{2})\partial\mathbf{w}(t)/\partial t$. Therefore Eq. (12) becomes

$$\frac{2}{\tau_{\text{vor}}}\mathbf{w} + \frac{3}{2}\frac{\partial\mathbf{w}}{\partial t} = -V_\infty\nabla p. \tag{17}$$

Setting $\tau_{\text{vor}} = 2\rho q^2/\sigma\Omega$ and $V_\infty = \Omega/\rho q^2$, as suggested in Sec. II, allows the preceding equation to be rewritten as

$$\sigma\mathbf{w} + \frac{3\rho q^2}{2\Omega}\frac{\partial\mathbf{w}}{\partial t} = -\nabla p. \tag{18}$$

Similar equations, relating the pressure gradient to the sum of a static flow term and a mass acceleration term, are common in the literature. However, the coefficient on the mass acceleration term is written in a variety of ways that are not necessarily equivalent. For example, in place of $3\rho q^2/2\Omega$, Morse and Ingard[4] use an "effective density" ρ_p. Fellah and Depollier[5] use $\rho\alpha_0$, where α_0 is the zero-frequency limit of the tortuosity. Attenborough[13] shows, for slanted circular cylindrical pores, that the coefficient is $4\rho q^2/3\Omega$. The main point here is that Eq. (18) is the phenomenologically correct low-frequency approximation, although the coefficient on the mass acceleration term is approximate with the version of the relaxation model used in this letter.

Considering next the low-frequency approximation of the mass-conservation equation, Eq. (15), we similarly make the approximation $\partial p(t')/\partial t' \simeq \partial p(t)/\partial t$ under the convolution integral. After performing the integration, we are left with

$$\beta_\infty\gamma\frac{\partial p}{\partial t} = -\nabla\cdot\mathbf{w}. \tag{19}$$

Since $\beta_\infty\gamma = \Omega/P$ is the isothermal compressibility, this result simply implies that at low frequencies ($\omega\tau_{\text{ent}} \ll 1$) the sound waves propagate isothermally in the pores, as is consistent with other models in the literature.[4,13]

Let us next consider high-frequency approximations. If the acoustic period is much shorter than the relaxation times τ_{vor} and τ_{ent}, the relaxation process always lags the sound wave and the exponential factor in the convolution integrals remains close to 1. (This is true so long as the high-frequency excitation has been stationary over an interval large compared to τ_{vor}.) Furthermore, when $\omega\tau_{\text{vor}} \gg 1$ in Eq. (12), we may neglect the terms involving $\mathbf{w}(t)/\tau_{\text{vor}}$ in comparison to $\partial\mathbf{w}(t)/\partial t$. Therefore Eqs. (12) and (15) approximate as

$$\frac{\partial\mathbf{w}}{\partial t} + \frac{1}{\sqrt{\pi\tau_{\text{vor}}}}\int_{-\infty}^{t}\frac{\partial\mathbf{w}(t')/\partial t'}{\sqrt{t-t'}}dt' = -V_\infty\nabla p \tag{20}$$

and

$$\beta_\infty\frac{\partial p}{\partial t} + \frac{\beta_\infty(\gamma-1)}{\sqrt{\pi\tau_{\text{ent}}}}\int_{-\infty}^{t}\frac{\partial p(t')/\partial t'}{\sqrt{t-t'}}dt' = -\nabla\cdot\mathbf{w}. \tag{21}$$

Alternatively, we could have derived these equations from Eqs. (11) and (14) by observing that the relaxation function at high frequencies is $S(\omega\tau) \simeq (-i\omega\tau)^{-1/2}$, and that $\mathcal{F}^{-1}[(-i\omega\tau_{\text{vor}})^{-1/2}] = (\pi\tau_{\text{vor}}t)^{-1/2}H(t)$. Equations (20) and (21) are the same as Fellah and Depollier's[5] if one sets $V_\infty = \Omega/\rho\alpha_\infty$, $\tau_{\text{vor}} = \Lambda^2\rho/4\eta$, and $\tau_{\text{ent}} = N_{\text{Pr}}(\Lambda'/\Lambda)^2\tau_{\text{vor}}$, where η is viscosity and Λ and Λ' are viscous and thermal characteristic lengths.

VI. OTHER TIME-DOMAIN EQUATIONS

Equations (18) and (19) describe sound propagation in porous media in the limit of low frequency. In this section, we develop extensions of these equations that are valid to a higher order in frequency but, unlike Eqs. (9), (12), and (15), do not involve convolution integrals.

Equation (11) can be written in the following form:

$$\sqrt{1-i\omega\tau_{\text{vor}}}\,\mathbf{W} = -(1-i\omega\tau_{\text{vor}})\mathbf{W} - \tau_{\text{vor}}V_\infty\nabla P. \tag{22}$$

On the left side of this equation, the square root can be written as $\sqrt{1-i\omega\tau_{\text{vor}}} = 1 - i\omega\tau_{\text{vor}}/2 + O(\omega^2)$, where $O(\omega^2)$ indicates terms of order ω^2. Using this formula, Eq. (22) becomes

$$i\omega\tau_{\text{vor}}\mathbf{W} = (\tfrac{4}{3})\mathbf{W} + (\tfrac{2}{3})\tau_{\text{vor}}V_\infty\nabla P + O(\omega^2). \tag{23}$$

In the time-domain, Eq. (23) coincides with Eq. (18). In order to generalize the latter equation to higher frequencies,

we apply a Padé 1,1 approximation to the square root in Eq. (22):

$$\sqrt{1-i\omega\tau_{\mathrm{vor}}}=\frac{1-i3\omega\tau_{\mathrm{vor}}/4}{1-i\omega\tau_{\mathrm{vor}}/4}. \tag{24}$$

Let us decompose both sides of Eq. (24) into a Taylor series. Then, the left and right sides will coincide to order ω^2 and differ by terms $O(\omega^3)$. Therefore, a Padé 1,1 approximation is formally equivalent to the second-order approximation of the square root $\sqrt{1-i\omega\tau_{\mathrm{vor}}}$. However, it is known that a Padé 1,1 approximation often works better than the second-order approximation of the square root since it is a polynomial approximation. Note that Padé approximants were also used by Chandler-White and Horoshenkov[14] to tie up the low- and high-frequency asymptotic limits of theoretical models of the acoustical properties of rigid frame porous media.

Substituting Eq. (24) into Eq. (22) and multiplying both sides of the resulting equation by $1-i\omega\tau_{\mathrm{vor}}/4$, we obtain

$$2(1-i\omega\tau_{\mathrm{vor}}-\omega^2\tau_{\mathrm{vor}}^2/8)\mathbf{W}=-\tau_{\mathrm{vor}}V_\infty(1-i\omega\tau_{\mathrm{vor}}/4)\nabla P. \tag{25}$$

In the time domain, this equation becomes

$$\sigma\left(1+\tau_{\mathrm{vor}}\frac{\partial}{\partial t}+\frac{\tau_{\mathrm{vor}}^2}{8}\frac{\partial^2}{\partial t^2}\right)\mathbf{w}=-\left(1+\frac{\tau_{\mathrm{vor}}}{4}\frac{\partial}{\partial t}\right)\nabla p. \tag{26}$$

This is a generalization of Eq. (18) to higher frequencies.

Let us next generalize the mass-conservation equation, (19). In Eq. (13), we use a Padé 1,1 approximation for the square root $\sqrt{1-i\omega\tau_{\mathrm{ent}}}$. The result is

$$[i\omega\gamma\beta_\infty-(2+\gamma)(i\omega)^2\tau_{\mathrm{ent}}\beta_\infty/4]P$$
$$=(1-3i\omega\tau_{\mathrm{ent}}/4)\nabla\cdot\mathbf{W}. \tag{27}$$

In the time domain, this equation reads

$$-\left[\gamma\beta_\infty\frac{\partial}{\partial t}+\frac{(2+\gamma)\tau_{\mathrm{ent}}\beta_\infty}{4}\frac{\partial^2}{\partial t^2}\right]p=\left(1+\frac{3\tau_{\mathrm{ent}}}{4}\frac{\partial}{\partial t}\right)\nabla\cdot\mathbf{w}. \tag{28}$$

Equations (26) and (28) are potentially useful for modeling sound interactions with porous media in FDTD simulations. Note that these equations contain second-order derivatives with respect to time. By expanding the square roots in Eqs. (22) and (13) into Taylor series and keeping terms of order ω^2 and ω, respectively, the following set of time domain equations can be derived:

$$\sigma\left(1+\frac{5\tau_{\mathrm{vor}}}{6}\frac{\partial}{\partial t}\right)\mathbf{w}=-\left(1+\frac{\tau_{\mathrm{vor}}}{12}\frac{\partial}{\partial t}\right)\nabla p \tag{29}$$

and

$$-\beta_\infty\gamma\frac{\partial p}{\partial t}=\left(1+\frac{(\gamma-1)\tau_{\mathrm{ent}}}{2\gamma}\frac{\partial}{\partial t}\right)\nabla\cdot\mathbf{w}. \tag{30}$$

[When deriving the first of these equations, in the term $(\omega\tau_{\mathrm{vor}})^2W$ appearing in the derivation, $\omega\tau_{\mathrm{vor}}W$ was replaced by the right side of Eq. (23) so that the resulting expression was proportional to the first power of ω.] Equations (29) and (30) are first-order differential equations with respect to time. Their range of applicability is narrower than that of Eqs. (27) and (28), however wider than that of Eqs. (18) and (19).

VII. CONCLUSION

A time-domain version of the relaxational model for sound propagation in rigid-frame porous media was developed in this paper. The new equations have physically realistic behavior for all frequencies and reduce to previously known results in the limits of low and high frequencies. The ease with which the time-domain model was developed appears to be due to the particular algebraic equations used in the relaxation formulation.

Both the Euler and mass-conservation equations in the time-domain relaxation model involve convolution integrals between the acoustic fields and an impulse response of the form $s(t)=(\pi\tau t)^{-1/2}\exp(-t/\tau)H(t)$, where τ is a characteristic relaxation frequency. It would be of interest to find other response functions, with additional free parameters, that lead to tractable time-domain models.

ACKNOWLEDGMENTS

This work was supported by the DoD High Performance Computing Modernization Office and the ERDC-CRREL AT24 direct program. We thank D. F. Aldridge, N. P. Symons, D. H. Marlin, and K. Attenborough for their helpful comments on this work.

[1] E. M. Salomons, R. Blumrich, and D. Heimann, "Eulerian time-domain model for sound propagation over a finite-impedance ground surface. Comparison with frequency-domain models," Acust. Acta Acust. **88**, 483–492 (2002).
[2] T. Van Renterghem and D. Botteldooren, "Numerical simulation of the effect of trees on down-wind noise barrier performance," Acust. Acta Acust. **89**, 764–778 (2003).
[3] Z. E. A. Fellah, M. Fellah, W. Lauriks, and C. Depollier, "Direct and inverse scattering of transient acoustic waves by a slab of rigid porous material," J. Acoust. Soc. Am. **113**, 61–72 (2003).
[4] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw–Hill, New York, 1968).
[5] Z. E. A. Fellah and C. Depollier, "Transient acoustic wave propagation in rigid porous media: A time-domain approach," J. Acoust. Soc. Am. **107**, 683–688 (2000).
[6] D. K. Wilson, "Relaxation-matched modeling of propagation through porous media, including fractal pore structure," J. Acoust. Soc. Am. **94**, 1136–1145 (1993).
[7] D. K. Wilson, "Simple, relaxational models for the acoustical properties of porous media," Appl. Acoust. **50**, 171–188 (1997).
[8] D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–401 (1987).
[9] J. F. Allard and Y. Champoux, "New empirical equations for sound propagation in rigid frame fibrous materials," J. Acoust. Soc. Am. **91**, 3346–3353 (1992).
[10] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY, 1994).
[11] J. F. Allard, C. Depollier, J. Nicolas, W. Lauriks, and A. Cops, "Propriétés acoustiques des matériaux poreux saturés d'air et théorie de Biot" ("Acoustical properties of air-saturated porous materials and the theory of Biot"), J. Acoust. **3**, 29–38 (1990).
[12] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, San Diego, 1994).
[13] K. Attenborough, "Acoustical characteristics of rigid fibrous absorbents and granular materials," J. Acoust. Soc. Am. **73**, 785–799 (1983).
[14] S. N. Chandler-White and K. V. Horoshenkov, "Padé approximants for the acoustical characteristics of rigid frame porous media," J. Acoust. Soc. Am. **98**, 1119–1129 (1995).

Submerged plane layered isotropic media: Properties of the scattering matrix eigenvalues with application to bilayers (L)

Hervé Franklin, Jean-Louis Izbicki, Teddy Marie-Françoise, and Pascal Rembert[a)]
Laboratoire d'Acoustique Ultrasonore et d'Electronique (LAUE) UMR CNRS 6068, Université du Havre, Place Robert Schuman, 76610 Le Havre, France

(Received 19 April 2004; revised 5 July 2004; accepted 6 July 2004)

The plot of the transmission coefficient moduli cannot be used to analyze all the resonances of layered structures because of the overlapping of the resonances. A method using the scattering matrix eigenvalues is presented. It allows, first, to annihilate the overlapping and to recover the resonance positions and half-widths and second, to divide in two independent sets most of the resonances involved in symmetric structures as well as in nonsymmetric ones. Comparisons between the results provided by the method and those coming from the poles of the transmission coefficient agree. In many cases, the transition terms constitute a generalization of the transmission coefficients. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1785615]

PACS numbers: 43.20.Ks, 43.20.Fn, 43.40.Dx [GCG] Pages: 1893–1896

I. INTRODUCTION

Nowadays, the theoretical methods to analyze acoustic wave properties in plane layered media, for instance in the goal of nondestructive evaluation (NDE) of materials, borrow mainly three different ways. At first, if modal properties are studied, a Newton–Raphson procedure is set to work to search for the roots of the secular equation and the dispersion curves are plotted.[1,2] Second, reflection (or transmission) coefficient modulus is plotted in the frequency or the angular variable. The positions and widths of the troughs (or the spikes) in the curves are compared with the experimental data.[3] In other cases, a resonant scattering formalism is built in order to obtain a best interpretation of the wave process.[4] The connection of this second way with the first is generally not easy because there is very little correspondence between the trough (peak) location in the curve of the reflection (transmission) coefficient modulus and the information provided by the mode's dispersion curves. The difficulty originates from the overlapping existing between modes.[5,6] That is why a study of the phase of the reflection coefficient is recommended at the same time. It allows one to determine the frequencies (or angles) where jumps occur in the phase. Unfortunately, the resonance width remains difficult to obtain. Third, the phase derivatives[7] of the reflection coefficient with respect to the variables (frequency or angle) or to the physical parameters are plotted. This way shows clearly the link existing between the first two ways, as it allows one to obtain both the position and the half-width of the resonances, but it necessitates the computation of derivatives and is therefore neither easy nor fast enough to set to work.

For the three ways we have just described, if the layered structure is a symmetrical stacking (that is, it has a physical plane of symmetry parallel to the interfaces), symmetric or antisymmetric vibration modes are easily distinguished. Conversely, if the stacking is an asymmetrical one, it becomes difficult to predict the symmetric or antisymmetric nature of a given mode (or resonance) and then to have a certainty of the family to which this mode belongs.

A simple method that is suitable for a fast evaluation of the resonant properties of layered media is proposed which connects the first two ways and avoids the derivative processes of the third method. It is based on the analysis of the S-matrix eigenvalues.[8] In Sec. II, we recall the main properties of this matrix for a layered structure immersed in a liquid. In Sec. III, in order to illustrate the method, we consider the case of an elastic and isotropic solid bilayer immersed in water (water-loaded aluminum/steel structure). Comparisons are made between the results provided by the eigenvalues of the S matrix and the results given by the roots of the dispersion equations and the transmission coefficient of the bilayer.

II. SCATTERING MATRIX PROPERTIES

Consider a plane bilayer $S1/S2$ with a welded contact between the elastic solids $S1$ and $S2$, the interfaces being in the Ox axis direction. Such a contact between solids is considered here simply for checking the method but other types of contact could be of concern as soon as dissipative processes are not taken into account. The surrounding liquid $F0$ has density ρ_0, sound wave speed c_0. The layers being indexed j with $j=1,2$, the densities are ρ_j, the L-wave (respectively, T wave) velocity is denoted c_{jL} (respectively, c_{jT}). A plane homogeneous monochromatic wave, with angular frequency ω, is incident with an angle θ_0 from the surrounding liquid onto the bilayer. The wave number in the interface's direction will be denoted $k_x = k_0 \sin\theta_0$, with $k_0 = \omega/c_0$. Numerical results will be, however, given with the help of the nondimensioned form of the wave number $\overline{k_x} = \sin\theta_0$.

Let R_j be the reflection coefficient of the whole bilayer plate and T_j the related transmission coefficient when the incident wave first encounters the $F0/Sj$ interface. One has

$$R_1 \neq R_2, |R_1| = |R_2|, T_1 = T_2.$$

These coefficients are frequency $f = \omega/2\pi$ and angle θ_0 dependent. From the energy conservation law across the struc-

a)Electronic mail: rembert@univ-lehavre.fr

ture, a 2×2 matrix called S matrix can be built[8] whose elements are the reflection coefficients on the diagonal and the transmission coefficients on the off-diagonal. Defined over the field of complex numbers, this matrix is unitary: $SS^{+}=S^{+}S=I$ (the + sign means Hermitian conjugate and I is the identity matrix). Both eigenvalues of this matrix, λ and μ [see Eq. (27) of Ref. 8], have the modulus equal to 1 and the associated eigenvectors define an orthonormal basis (e_λ, e_μ), that is, two independent subspaces.

A resonance of the bilayer will occur each time the (squared) modulus of the transition terms defined by $T_\lambda=(\lambda-1)/2i$ and $T_\mu=(\mu-1)/2i$ and plotted versus the frequency or the incidence angle θ_0 describes a peak. In Ref. 8, devoted to the problem of an immersed aluminum plate, it has been shown that the eigenvalue λ contains the symmetrical modes, whereas the eigenvalue μ contains the antisymmetrical modes. Here, because of the absence of material symmetry with respect to the median plane of the bilayer, it is no longer possible to speak of symmetrical or antisymmetrical modes. Nevertheless, the independence of the eigenvectors e_λ and e_μ suggests that the transition terms describe again two physical independent parts of the resonant spectrum, and this can be seen as the extension of the customary separation between symmetrical and antisymmetrical modes. As a consequence of this last remark, it can happen that a given resonance is drawn partially by each of the two eigenvalues. It becomes difficult then to specify the nature of the resonance and, by extension, the nature of a mode.

III. CASE OF A WATER LOADED ALUMINUM/STEEL BILAYER

For the computations, the following parameters are chosen: water ($\rho_0=7900\ \mathrm{kg\,m^{-3}}$, $c_0=1470\ \mathrm{m\,s^{-1}}$); aluminum ($\rho_1=2790\ \mathrm{kg\,m^{-3}}$, $c_{1L}=6380\ \mathrm{m\,s^{-1}}$, $c_{1T}=3100\ \mathrm{m\,s^{-1}}$); and steel ($\rho_2=7900\ \mathrm{kg\,m^{-3}}$, $c_{2L}=5790\ \mathrm{m\,s^{-1}}$, $c_{2T}=3100\ \mathrm{m\,s^{-1}}$). The layers each have a thickness $d=3$ mm. This choice of parameters corresponds to characteristic impedance ratios $\rho_0 c_0/\rho_j c_{jL,T}$ $(j=1,2)$ small when compared with 1, that is to say, we consider here light fluid loading. Figure 1 shows the curve of $|T_1|^2$ plotted versus the angular variable for the frequency 4.5 MHz. Most of the peaks never reach 1. Figure 2 shows the curves of $|T_\lambda|^2$ and $|T_\mu|^2$ for the same frequency. Contrary to the transmission coefficient, all the peaks have a magnitude equal to 1. For very narrow peaks, this is not obvious on the curve due to numerical resolution. The whole resonant spectrum is divided in two sets: resonances with the solid curve ($|T_\lambda|^2$) and resonances with the dot curve ($|T_\mu|^2$). Many results can be read on these curves: (1) Beyond the aluminum critical angle $\theta_{cT}\approx28.31°$, $|T_1|^2$ has only one (very narrow) peak, while each of the two transition terms exhibits one broad peak. (2) From the preceding point, we can deduce the existence of two kinds of resonance beyond the critical angle: first, the Rayleigh type located near angles 30.50° (water/aluminum) and 30.76° (water/steel) with corresponding angular half-widths 1.90° and 0.73°; second, the Stoneley type very narrow in comparison to the Rayleigh type and located near the angle 28.71°. Actually, because of the relatively great value of the frequency (i.e., short wavelength with regard to the thickness of the layers) and due to the angular domain (angles greater than θ_{cT}), the three interfaces are noncoupled ones. Then, we have to deal with three semi-infinite structures: $F0/S1$, $F0/S2$, and $S1/S2$. The Rayleigh-type resonances describe the generalized Rayleigh waves propagating along the $F0/S1$ and the $F0/S2$ interfaces. The Stoneley-type resonance describe the Stoneley wave propagating along the $S1/S2$ interface (the existence conditions[9,10] are filled here for the aluminum/steel combination). Converted into wave number units, one has for the generalized waves at the water/aluminum interface: $\overline{k_x}\approx0.507+i0.014$ and at the water/steel interface: $\overline{k_x}\approx0.511+i0.037$. For the Stoneley wave at the aluminum/steel interface, one has $\overline{k_x}\approx0.480+i0.000$. These values are identical to those obtained by solving the secular equation.

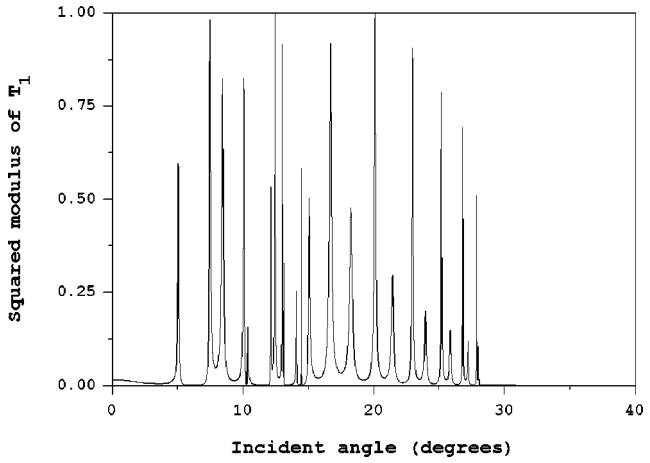


FIG. 1. Water-loaded aluminum/steel structure. The squared modulus of the transmission coefficient is plotted versus the incident angle at a frequency of 4.5 MHz ($fd=13.5$ MHz mm). Most of the amplitudes of the peaks never reach 1. Beyond the angle 29° no peak exists.

A zoom of the situation beyond the critical angle θ_{cT} is presented in Fig. 3, in the angular domain 26°–36°. The

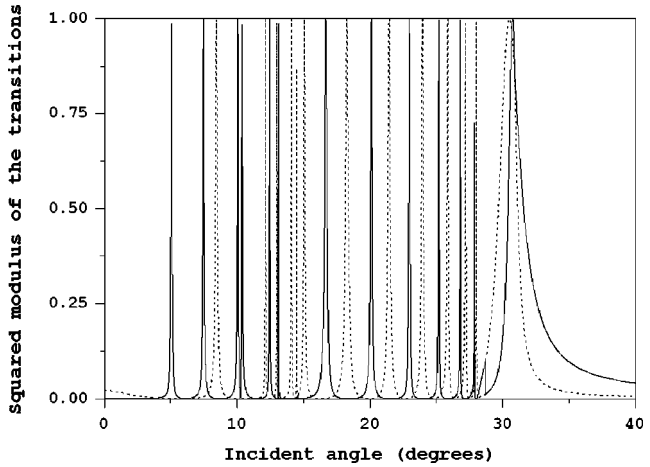


FIG. 2. Water-loaded aluminum/steel structure. The squared modulus of the transitions terms is plotted versus the incident angle at a frequency of 4.5 MHz. Three effects of the transition terms are (a) the distinction between two sets of resonances by the solid ($|T_\lambda|^2$) and dot ($|T_\mu|^2$) curves; (b) the obtaining of peaks with amplitudes equal to 1 according to the law of energy conservation; (c) the obtaining of resonance peaks beyond the angle 29° related to interface waves.

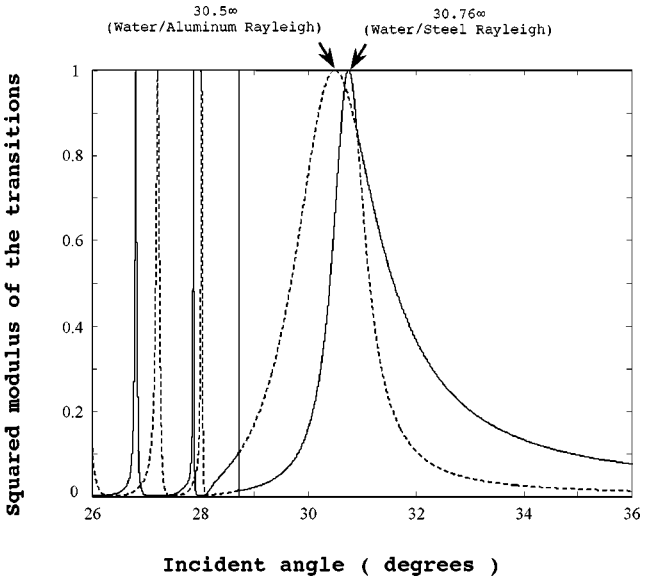


FIG. 3. A zoom of the preceding figure in the angular domain [26°, 36°] showing the very narrow Stoneley wave peak located at 28.71°, the generalized Rayleigh wave peaks located, respectively, at 30.50° and 30.76°. Due to the asymmetry of the bilayer, the Rayleigh resonances have no definite nature, being each partially described by the two eigenvalues. Both the fluid loading and the coupling between solids can influence the nature of the resonances.

Rayleigh resonances are not related to a given eigenvalue but arise from a contribution of both λ and μ. For instance, the curve of the water/steel Rayleigh resonance (amplitude of transition term equal to one at 30.76°) is described in the domain 29°–31° by the solid curve and in the domain 31°–36° by the dot curve. It is superposed on the curve of the water/aluminum Rayleigh resonance (amplitude of transition term equal to 1 at 30.50°) which is also described by the two eigenvalues and has a broader width. The fact that the Rayleigh resonances are each partially described by the eigenvalues expresses the absence of physical symmetry in the bilayer. It then results that the two eigenvalues contribute to each Rayleigh resonance contrary to the case of a single plate, where each Rayleigh resonance is entirely described by one eigenvalue.[8] The linear independence of the eigenvectors ensures for its part that the Rayleigh resonances of Fig. 3 are two different entities. However, in the spectrum of Fig. 2, one can also find resonances described by only one of the eigenvalues, that is, having a defined nature.

Another result can be obtained by examining the dispersion curves of the water-loaded aluminum/steel bilayer, Fig. 4, which furnish the real parts of the complex roots $\overline{k_x}$ of the dispersion equation versus the frequency–thickness product fd. The comparison with the curves of the transition terms at an arbitrary value of the frequency, says 0.66 MHz in Fig. 5, shows a good agreement between the data provided by the two methods. The peaks of the transition terms are located exactly as indicated on the dispersion curves. The half-widths correspond also to two times the imaginary part of the roots of the dispersion equation.

IV. CONCLUSION

A new method has been performed to analyze the acoustical properties (modes and resonances) of nondissipative

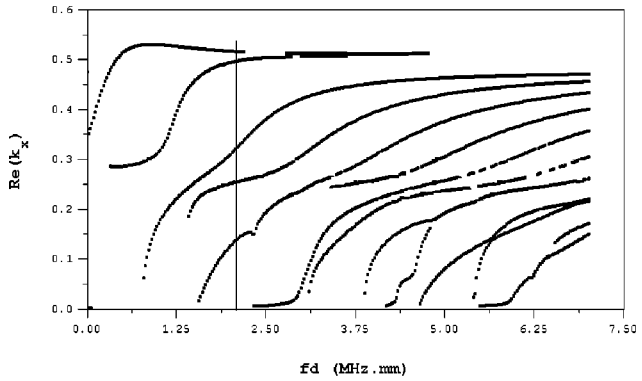


FIG. 4. Dispersion curves of the water-loaded aluminum/steel structure. Re($\overline{k_x}$) plotted versus the frequency–thickness product.

layered media. The reflection and transmission coefficients of the layered structures have not been studied individually as usual. We rather use the transition terms which are derived from the eigenvalues of the scattering matrix. The transition terms are a combination ot the reflection and transmission coefficients and contain more physical insight than the two coefficients. As a proof, all the resonant behaviors of the bilayers are sketched and most of the overlapping phenomena annihilated. In addition, a one-to-one connection is established between the points on the dispersion curves of the bilayer, which were usually found by computing the roots of the secular equation and our method. Technically, these transition terms are nothing but symmetrization and antisymmetrization operators in the case of symmetric bilayers. They can also be considered as a generalization of the transmission coefficient concept.

By considering sufficiently great values of the frequency and plotting the modulus of the transition terms versus the

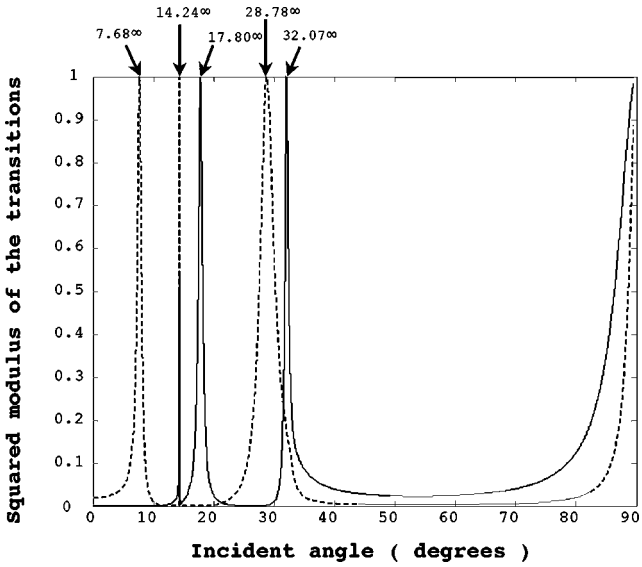


FIG. 5. Squared modulus of the transitions terms $|T_\lambda|^2$ (solid curves) and $|T_\mu|^2$ (dot curves) versus the incident angle, at 0.666 MHz ($fd \approx 2$ MHz mm). Comparison with the points of the dispersion curves at the same frequency (points of intersection with the vertical line positioned at $fd \approx 2$ MHz mm) shows, for the real parts of the wave numbers, a good agreement between the method using dispersion curves and the method of the scattering matrix eigenvalues (the abscissas of the peak maxima are indicated by arrows).

incident angle, surface wave characteristics of the layers are exhibited. In particular, for a water-immersed aluminum/steel bilayer, it is easy to understand that the surface wave resonances were not present in the reflection or the transmission coefficients because of the overlapping between modes that occur in those coefficients. The method is accurate and very easy to use for resonance studies.

This paper focused on immersed bilayers, but the formulation given here also can be applied to an arbitrary number of layers (liquid or solid) with, however, some limitations. For instance, the method cannot be applied in the present form in the entire angular domain. For more than two layers, band gaps make the modes concentrate at some angles. Even if it is possible to distinguish between two sets of modes, it becomes difficult to separate clearly the spikes and to obtain resonance curves having or approaching the Breit–Wigner shape.

Recently, this method has been used to recognize the symmetric or antisymmetric nature of the guided modes of a water-saturated poroelastic layer immersed in water[11] by plotting the squared modulus of each transition term in the $(\overline{k_x}, fd)$ plane. It remains now to experiment it on more complex cases such as elastic isotropic layered media for which the light fluid-loading assumption does not hold or anisotropic layered media.

[1] J. P. Jones, ''Wave propagation in a two-layered medium,'' J. Appl. Mech., Trans. ASME **31**, 213–222 (1964).

[2] F. Coulouvrat, M. Rousseau, O. Lenoir, and J.-L. Izbicki, ''Lamb-type waves in a symmetric solid–Fluid–Solid trilayer,'' Acust. Acta Acust. **84**, 12–20 (1998).

[3] D. F. Mc Cammon and S. T. McDaniel, ''The influence of the physical properties of ice on reflectivity,'' J. Acoust. Soc. Am. **77**(2), 499–507 (1985).

[4] P. D. Jackins and G. C. Gaunaurd, ''Resonance acoustic scattering from stacks of bonded elastic plates,'' J. Acoust. Soc. Am. **80**(6), 1762–1776 (1986).

[5] A. Freedman, ''On the overlapping resonances concept of acoustic transmission through an elastic plate. I. An examination of properties,'' J. Sound Vib. **82**, 181–195 (1982).

[6] A. Freedman, ''On the overlapping resonances concept of acoustic transmission through an elastic plate. II. Numerical examples and physical implications,'' J. Sound Vib. **82**, 197–213 (1982).

[7] O. Lenoir, J. Duclos, J.-M. Conoir, and J.-L. Izbicki, ''Study of Lamb waves based upon the frequential and angular derivatives of the phase of the reflection coefficient,'' J. Acoust. Soc. Am. **94**, 330–343 (1993).

[8] H. Franklin, E. Danila, and J.-M. Conoir, ''*S*-matrix theory applied to acoustic scattering by asymmetrically fluid-loaded elastic isotropic plates,'' J. Acoust. Soc. Am. **110**(1), 243–253 (2001).

[9] J. G. Scholte, ''The range of existence of Rayleigh and Stoneley waves,'' Mon. Not. R. Astron. Soc.: Geophys. Suppl. **5**, 120–126 (1947).

[10] K. Sesawa and K. Kanai, ''The range of possible existence of Stoneley waves and some related problems,'' Bull. Earthquake Res. Inst., Univ. Tokyo **17**, 1–8 (1939).

[11] G. Belloncle, H. Franklin, F. Luppe, and J.-M. Conoir, ''Normal modes of a poroelastic plate and their relation to the reflection and transmission coefficients,'' Ultrasonics **41**, 207–216 (2003).

The effects of single and double hearing protection on the localization and segregation of spatially-separated speech signals (L)

Douglas S. Brungart
Air Force Research Laboratory, WPAFB, Ohio

Alexander J. Kordik
Sytronics, Dayton, Ohio

Brian D. Simpson
Air Force Research Laboratory, WPAFB, Ohio

(Received 25 December 2003; revised 21 May 2004; accepted 9 July 2004)

Recent results have shown that auditory localization in the horizontal plane is dramatically worse for listeners wearing double hearing protection (earplugs and earmuffs) than it is for listeners wearing single hearing protection (earplugs or earmuffs alone). This suggests that double hearing protection might also impair the spatial unmasking that normally occurs when two simultaneous talkers are spatially separated in azimuth (the so-called ''cocktail party'' effect). In this experiment, normal hearing listeners wearing no hearing protection, single hearing protection (earmuffs), or double hearing protection were asked to perform a speech intelligibility task that required them to segregate two simultaneous talkers who were either presented from the same loudspeaker or spatially separated by 90° in azimuth. The listeners were also asked to determine the location of the target talker in each trial. The results show that the listeners were unable to reliably determine the location of the target talker when they wore double hearing protection, but that they were still able to benefit from the spatial separation of the competing talkers. This suggests that the use of double hearing protection causes spatially separated sound sources to be heard at locations that are inaccurate but still distinct enough to enhance the segregation of speech. [DOI: 10.1121/1.1786812]

PACS numbers: 43.50.Hg, 43.66.Vt, 43.66.Qp [AR] Pages: 1897–1900

I. INTRODUCTION

Hearing protection devices play an essential role in preventing long-term occupational hearing loss, but they can also impair the localization cues that listeners use to process and analyze surrounding sound sources. This is especially true in environments with noise levels that exceed 100 dBA, where double hearing protection (a combination of earplugs and earmuffs) should be worn to avoid the possibility of permanent hearing loss (NIOSH, 1998). Recent results from our laboratory have shown that listeners wearing double hearing protection are much worse at sound localization than those wearing earplugs alone or earmuffs alone. Earplugs and earmuffs have long been known to impair the spectral cues used for the localization of brief sounds in the horizontal plane (Abel and Hay, 1996; Abel and Armstrong, 1993; Vause and Grantham, 1999; Bolia, D'Angelo, Mishler, and Morris, 2001), but most studies have shown that listeners wearing single hearing protection can adequately localize sound sources in azimuth when they are on long enough to facilitate the use of exploratory head movements (Noble, Murray, and Waugh, 1990). However, two recent studies in our laboratory have shown that listeners wearing double hearing protection are reduced to near chance localization performance in the horizontal plane even with continous sound sources that allow them to make unrestricted head movements (Brungart, Kordik, Simpson, and McKinley, 2003; Simpson, Bolia, McKinley, and Brungart, 2002).

One possible explanation for the dramatic decrease in localization performance that occurs when double hearing protection is worn is that the interaural localization cues that are usually present in audio signals that enter the ears through the ear canals are being contaminated by bone- and tissue-conducted sounds that enter the listener's ears directly through the surface of the head. This kind of bone conduction is already known to place a limit on the amount of attenuation that can be achieved with conventional earplug and earmuff hearing protection devices (Zwislocki, 1957; Berger, Kieper, and Gauger, 2003). Once these devices attenuate the direct sound entering the ear canal below the level of the sound reaching the cochlea through bone or tissue conduction, additional attenuation no longer has any effect on the perceived level of the sound or on the long-term damage it can cause to the listener's hearing. Researchers who have estimated the bone-conduction limit by measuring sound detection thresholds with increasingly effective earplug and earmuff combinations have found that the bone-conduction limit on conventional hearing protection ranges from roughly 39 dB of attenuation at 2 kHz to roughly 55 dB of attenuation at 500 Hz (Berger, 1983; Berger *et al.*, 2003).

Although bone-conduction pathways have traditionally been examined in the context of the limitations they impose on the effective *attenuation* of hearing protection devices, it is likely that these flanking pathways also influence how well listeners are able to *localize* sound sources while wearing hearing protection. Sound localization in the horizontal plane

depends primarily on direction-dependent changes in interaural time and intensity differences caused by the diffraction of air-conducted soundwaves around the listener's head and torso. These interaural difference cues would be badly corrupted in a bone-conducted sound, which would propagate directly through the head at a much higher rate than a normal air conducted signal. Consequently, it may not be surprising that a listener wearing double hearing protection that reaches the bone-conduction attenuation limit at frequencies at or above 1 kHz would perform much worse in a sound localization task than a listener wearing single hearing protection that does not approach the bone-conduction limit at any frequency. Indeed, similar reductions in localization ability have been reported for listeners with severe conductive hearing losses who experience the same kind of increase in the level of bone-conducted sound relative to the level of air-conducted sound as listeners wearing double hearing protection (Zurek, 1986; Noble, Byrne, and LePage, 1994).

Although a reduction in sound localization ability is clearly one important consequence of the use of double hearing protection, it may not be the only consequence. The degradation in sound localization that occurs when double hearing protection is worn may also impact the processing of complex auditory scenes with multiple simultaneous sound sources. In normal listening environments, the ability to segregate and process multiple simultaneous sound sources is greatly enhanced when those sound sources are spatially separated. The classic example of this is the dramatic increase in speech intelligibility that occurs when a target speech signal is spatially separated from a masking speech signal, often referred to as the ''cocktail party'' effect (Cherry, 1953). To the extent that double hearing protection disrupts the interaural localization cues that listeners use to segregate spatially separated sounds, one might expect the benefits of spatial separation inherent in the cocktail party effect to be greatly reduced in listeners wearing double hearing protection. Indeed, Zurek (1986) has speculated that a similar reduction in spatial unmasking should occur in listeners with severe conductive hearing losses. If this kind of reduction in complex sound processing ability does occur in listeners wearing double hearing protection, then steps should be taken to alert the users of double hearing protection to this problem and to eliminate situations where a lack of spatial segregation ability could pose a safety hazard in high noise environments. The purpose of this study was to examine the effect that double hearing protection has on the segregation of multiple simultaneous speech sources.

II. METHODS

A. Subjects

Six volunteer subjects, five males and one female, were paid to participate in the experiment. All had normal hearing ($<$15 dB HL from 500 Hz to 8 kHz), and their ages ranged from 21 to 24 years.

B. Hearing protection devices

Three hearing protection (HP) conditions were tested in the experiment: No HP, where no hearing protection device was used; Single HP, where the subjects wore earmuffs (Tasco Soundshield); and Double HP, where the subjects wore fully inserted foam earplugs (E·A·R Classic ®) under earmuffs (TASCO Soundshield). In conditions where HP devices were used, they were fitted under the supervision of an experimenter. Prior to this study, the real ear attenuation at threshold method was used to determine the attenuation characteristics of the Single HP and Double HP hearing protection devices for each of the six participants in this experiment. These attenuation results, which are reported in detail elsewhere (Brungart *et al.*, 2003), indicate that the combination of earplugs and earmuffs used in the Double HP condition of this experiment reach the bone-conduction limit of attenuation for frequencies at or above 1 kHz, while the earmuffs used in the Single HP condition produce less attenuation than the bone conduction limit at all frequencies.

C. Apparatus

The experiment was conducted in the Air Force Research Laboratory's Auditory Localization Facility (ALF), a large anechoic chamber housing an aluminum-frame geodesic sphere with small loudspeakers mounted at each of its 272 vertices (Brungart *et al.*, 2003). These 272 loudspeakers were not capable of producing sufficiently intense auditory stimuli for the Double HP condition of this experiment. Consequently, the ALF was modified by mounting two large powered loudspeakers with twin 8-in. woofers (Barbetta Sona) at $\pm 45°$ azimuth in the horizontal plane 1.4 m away from the subject, who was positioned on a platform in the center of the sphere. These powered loudspeakers were used to generate all the stimuli used in the experiment. In order to allow the subjects to use a computer mouse to make their responses, a 14 in. color cathode-ray tube was also mounted outside the sphere between the two loudspeaker locations directly in front of the subject.

D. Threshold measurement

At the start of each block of trials, the subjects were fitted with the appropriate level of hearing protection and asked to stand in the middle of the sphere facing the midpoint between the two loudspeakers. Then they participated in a two-down, one-up adaptive threshold procedure (Levitt, 1971) to estimate their hearing threshold level with that hearing protection device. In each trial of this two-interval forced-choice procedure, light emitting diodes located over the left and right buttons of a two-button response box were used to visually cue two consecutive 250-ms observation intervals, separated by a pause of 100 ms. A 250-ms burst of pink noise[1] was presented in one of these two observation intervals, and the subject's task was to determine which interval contained the noise burst. Two consecutive correct responses resulted in a 3 dB decrease in the noise level, and each incorrect response resulted in a 3 dB increase in the noise level. This procedure was repeated until ten reversals occurred, and the last five reversals were averaged to estimate the detection threshold for that particular subject with that hearing protection device. The mean detection thresholds measured with this procedure were roughly -3 dB

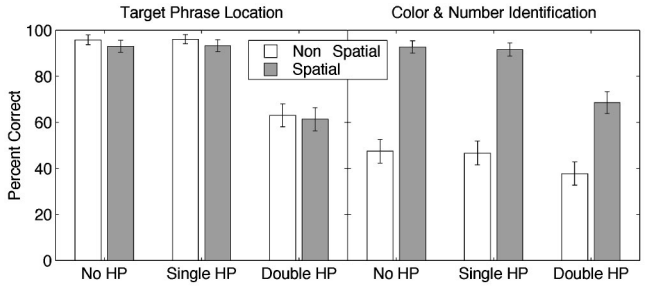


FIG. 1. The left panel shows performance in the left-right discrimination task, where the subjects had to identify which speaker the target phrase originated from. The right panel shows the percentage of correct color and number identifications in each condition of the experiment. The error bars represent the 95% confidence intervals calculated from all the raw data at each data point.

sound pressure level (SPL) in the No HP condition, 22 dB SPL in the Single HP condition, and 37 dB SPL in the Double HP condition.[2]

E. Procedure

Once the threshold measurement was complete, data collection started in the speech intelligibility test. This test was based on the coordinate response measure (CRM), a call-sign-based intelligibility test consisting of target and masking phrases of the form ''Ready (call sign) go to (color) (number) now.'' The phrases themselves were drawn from the publicly-available CRM corpus (Bolia, Nelson, Ericson, and Simpson, 2000), which contains CRM phrases for all combinations of eight call signs (''Arrow,'' ''Baron,'' ''Charlie,'' ''Eagle,'' ''Hopper,'' ''Laker,'' ''Ringo,'' ''Tiger''), four colors (''blue,'' ''green,'' ''red,'' ''white''), and eight numbers (1–8) spoken by four male and four female talkers. In this experiment, the stimulus always consisted of two simultaneous CRM phrases, a target phrase randomly selected from all the phrases containing the call sign ''Baron,'' and a masking phrase randomly selected from all the phrases spoken by a different same-sex talker with a different call sign, color, and number than the target phrase. These phrases were adjusted to set their rms power levels 30 dB higher than the hearing threshold measured at the start of the block, and presented either from the same randomly selected loudspeaker (''nonspatial trials'') or from different randomly selected loudspeakers (''spatial trials''). In each case, the subject's task was to first identify the color and number combination contained in the phrase containing ''Baron,'' and then to determine whether the target phrase came from the left or right loudspeaker. The subjects were permitted to move their heads during this procedure, but they were not specifically encouraged to do so. Each block consisted of 15 spatial and 15 nonspatial trials, and each subject participated in four blocks of trials with each of the three hearing protection levels, with the order of the blocks randomized across the different subjects.

III. RESULTS

The results of the experiment are shown in Fig. 1. The left panel of the figure shows performance in the left-right discrimination task of the experiment. These results show that left-right discrimination was near 100% correct in both the No HP and Single HP conditions, but that it fell to roughly 60% correct in the Double HP condition.[3] A two-factor within-subjects analysis of variance (ANOVA) on the arcsine-transformed data from each individual subject confirmed that this drop in performance was significant ($F_{(2,10)} = 27.95$, $p < 0.0001$). This finding is consistent with our earlier experiment that showed that listeners wearing double hearing protection are unable to reliably localize sound sources in the horizontal plane (Brungart *et al.*, 2003). Note that spatial separation of the talkers had no effect on left-right discrimination in any of the hearing protection conditions. Thus, it appears that the listeners were unable to reliably determine the location of the target talker in the Double HP condition even when both speech signals originated from the same loudspeaker and there was no need to use the target call sign to determine the location of the target phrase.

The right panel of Fig. 1 shows performance in the color-number identification task of the experiment. In this task, spatial separation significantly improved performance in every hearing protection condition tested (two-factor within subjects ANOVA on the individual arcsine-transformed scores) ($F_{(1,5)} = 211.63$, $p < 0.0001$). As in the left-right discrimination task, there was a significant decrease in performance in the Double HP condition ($F_{(2,10)} = 23.19$, $p < 0.0002$). There was also a significant (15 percentage point) decrease in the magnitude of the spatial advantage in the Double HP condition ($F_{(2,10)} = 5.60$, $p < 0.05$). Nevertheless, despite the poor location discrimination data in the Double HP condition, the intelligibility advantages of spatial separation in that condition were substantial (69% percent correct identifications spatial versus 38% non-spatial).

On the surface, it seems somewhat odd that listeners could obtain such a large intelligibility advantage from spatial separation in a condition where they could not reliably distinguish between two spatial locations that were separated by 90° in azimuth. Indeed, the 60% correct discrimination scores achieved in the left-right discrimination task, while technically better than chance, were lower than what is generally considered to be the discrimination threshold for the minimum audible angular change in the location of a sound source, or MAA (Mills, 1958). The most likely explanation for the spatial unmasking in the Double HP condition is that, although listeners in that condition could not accurately determine the actual locations of the talkers, they could tell that the talkers were originating from different locations, and this helped them to segregate the two speech signals into different streams and improve their score on the speech intelligibility test.

IV. SUMMARY AND CONCLUSIONS

This paper has presented the results of an experiment examining whether the decreases in localization ability that occur when listeners wear double hearing protection also impair their ability to segregate spatially separated speech signals. As in previous studies, the results of this experiment have shown that listeners who can reliably localize sound sources while wearing single hearing protection cannot do so

when they are wearing double hearing protection. However, despite this decrease in localization ability, listeners wearing double hearing protection still seem to gain substantial, albeit somewhat reduced, intelligibility benefit from the spatial separation of the competing talkers. Thus it seems that the degraded localization cues that prevent listeners from hearing sound sources in the correct locations do not necessarily make all sound sources appear to originate from the same location. This suggests that listeners wearing double hearing protection may not be as impaired in their ability to process complex auditory scenes as their poor localization scores would suggest. It also offers some hope that listeners who have been thoroughly trained while wearing double hearing protection (or those who are constantly exposed to bone-conducted sound due to conductive hearing loss) may eventually be able to learn to improve their localization ability. Previous research has shown that listeners who receive training can learn to adapt to a variety of different disruptions in the cues they would normally use to localize sounds. For example, Hofman and his colleagues (1998) conducted an experiment where listeners' normal localization cues were disrupted by inserting plastic molds into their left and right ears. The results of this experiment showed that listeners could, over a several day period, learn to localize relatively well with these disrupted cues. Furthermore, once they adapted to these cues, they were then able to localize with or without the earmolds with no addional periods of adaptation.

Shinn-Cunningham, Durlach, and Held (1998) reported similar results for localization cues that were systematically remapped in the horizontal plane. Thus, to the extent that double hearing protection produces disrupted location cues that change systematically with source location, one might expect listeners to eventually be able to learn to use these disrupted cues to localize sounds. Further research is now needed to determine the extent, if any, to which listeners are also able to adapt to the disrupted localization cues that occur with double HP.

ACKNOWLEDGMENT

This work was sponsored in part by AFOSR Grant No. 01-HE-01-COR.

[1]This noise was generated by using a DSP processor (Tucker-Davis RP2) to rectangularly gate the output of an analog pink noise generator (GenRad). Note that, although the noise was both pink and broadband prior to transmission to the loudspeakers, no attempts were made to correct for the loudspeaker and amplifier responses to ensure that this signal was pink at the location of the center of the head.

[2]The standard deviations of these threshold measurements were approximately 5 dB in the No HP condition, 7 dB in the Single HP condition, and 13 dB in the Double HP condition.

[3]Across the individual subjects in the Double HP condition, one listener performed relatively well at the left-right identification task (≈80% correct), one performed relatively poorly (≈30% correct), and the others fell between 50 and 70% correct responses. These differences are consistent with the relatively large individual differences found in our earlier study on localization with double hearing protection (Brungart *et al.*, 2003).

Abel, S., and Armstrong, N. (**1993**). ''Sound localization with hearing protectors,'' J. Otolaryngol. **22**, 357–363.

Abel, S., and Hay, V. (**1996**). ''Sound localization: The interaction of aging, hearing loss, and hearing protection,'' Scand. Audiol. **25**, 3–12.

Berger, E. (**1983**). ''Laboratory attenuation of earmuffs and earplugs, both singly and in combination,'' Am. Ind. Hyg. Assoc. J. **44**, 321–329.

Berger, E., Kieper, W., and Gauger, D. (**2003**). ''Hearing protection: Surpassing the limits to attenuation imposed by the bone-conduction pathways,'' J. Acoust. Soc. Am. **114**, 1955–1967.

Bolia, R., D'Angelo, W., Mishler, P., and Morris, L. (**2001**). ''The effects of hearing protectors on auditory localization in azimuth and elevation,'' Hum. Factors **43**, 122–128.

Bolia, R., Nelson, W., Ericson, M., and Simpson, B. (**2000**). ''A speech corpus for multitalker communications research,'' J. Acoust. Soc. Am. **107**, 1065–1066.

Brungart, D., Kordik, A., Simpson, B., and McKinley, R. (**2003**). ''Auditory localization in the horizontal plane with single and double hearing protection,'' Aviat., Space, Environ. Med. **65**, A31–38.

Cherry, E. (**1953**). ''Some experiments on the recognition of speech, with one and two ears,'' J. Acoust. Soc. Am. **25**, 975–979.

Hofman, P. M., Van Riswick, J. G. A., and Van Opstal, A. (**1998**). ''Relearning sound localization with new ears,'' Nat. Neurosci. **1**, 417–421.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Mills, A. (**1958**). ''On the minimum audible angle,'' J. Acoust. Soc. Am. **30**, 237–246.

NIOSH (1998). ''Criteria for a recommended standard-occupational noise exposure,'' Pub. No. 98–126, U.S. Department of Health and Human Services, Washington, DC.

Noble, W., Byrne, D., and LePage, B. (**1994**). ''Effects on sound localization of configuration and type of hearing impairment,'' J. Acoust. Soc. Am. **95**, 992–1005.

Noble, W., Murray, N., and Waugh, R. (**1990**). ''The effect of various hearing protectors on sound localization in the horizontal and vertical planes,'' Am. Ind. Hyg. Assoc. J. **51**, 370–377.

Shinn-Cunningham, B., Durlach, N., and Held, R. (**1998**). ''Adapting to supernormal auditory localization cues I: Bias and resolution,'' J. Acoust. Soc. Am. **103**, 3656–3666.

Simpson, B. D., Bolia, R., McKinley, R., and Brungart, D. (2002). ''Sound localization with hearing protectors: Performance and head motion analysis in a visual search task,'' *Proceedings of the Human Factors Society 46th Annual Meeting, Baltimore, MD, Sept 30–Oct 4, 2002*, 100–104.

Vause, N., and Grantham, D. (**1999**). ''Effects of earplugs and protective headgear on auditory localization ability in the horizontal plane,'' Hum. Factors **41**, 282–294.

Zurek, P. (**1986**). ''Consequences of conductive auditory impairment for binaural hearing,'' J. Acoust. Soc. Am. **80**, 466–472.

Zwislocki, J. (**1957**). ''In search of the bone conduction threshold in a free field,'' J. Acoust. Soc. Am. **29**, 795–804.

Acoustic scattering from finite cylindrical shells: Influence of end-caps (L)

L. Haumesser
LUSSI/GIP-ULTRASONS, FRE CNRS 2448, rue de la Chocolaterie BP3410, 41034 Blois cedex, France

N. Touraine
DCN Cherbourg, Place Bruat, BP 440, 50104 Cherbourg-Octeville Cedex, France

D. Décultot[a] and G. Maze
LAUE UMR CNRS 6068, Université du Havre, place Robert Schuman, 76610 Le Havre, France

(Received 19 September 2003; revised 9 July 2004; accepted 12 July 2004)

By using an experimental approach, the acoustic backscattering from air-filled cylindrical shells, immersed in water, is investigated. The targets considered in this work are ended either by hemispherical caps or by flat circular lids. Given the 1% shell thickness and the frequency window of this study, only echoes resulting from the propagation of S_0 and T_0 waves (respectively, quasi-compressional and quasi-shear waves of lowest order) are observable. On both types of objects, measurements are carried out at various aspect angles of incidence over the broadside of the target. The influence of the type of end-caps on the propagation of helical waves (i.e. waves initially generated on the cylindrical part of the objects) is studied using incidence-angle/time representations. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1788731]

PACS numbers: 43.40.Fz, 43.20.Fn [ANN] Pages: 1901–1904

I. INTRODUCTION

Many authors examine the acoustic radiation from fluid-loaded finite cylindrical shells excited by a plane wave. Analytical methods are generally used on such a problem to carry out investigations in various ranges of incidence angles on the broadside. In most cases, for incidence angles which correspond to the generation of waves on the cylindrical part, the finite shell is considered to be simply supported at its extremities.[1–4] Theoretical results are validated by a comparison with experimental ones. The comparison is widely accepted, even when the shells used in the experiments are ended by hemispherical caps.[1,3] In Ref. 3, the author calculates a backscattering pattern in the aspect-angle/frequency domain from a finite cylindrical shell considered to be simply supported at its extremities and compares it with a measured backscattering pattern obtained from a cylindrical shell ended by hemispherical caps. The backscattered pressure modulus presented in frequency domain does not highlight the influence of the extremity shape of studied shells. So various presentations of results are used: (a) aspect-angle/frequency domain,[3–5] (b) aspect-angle/time domain,[6,7] and (c) the time/frequency domain for given aspect-angles.[7] In addition, in the frequency domain, an identification of resonance vibration modes of a finite flat-ended cylindrical shell has been achieved by an experimental bistatic setup in Ref. 8.

Other results from computation are obtained by using a coupled finite element method/boundary element method.[9–12] In these studies, the axial incidence on cylindrical shells ended by hemispherical caps, where waves are generated on the hemispherical part, is investigated. In addition, some of these studies analyze influences of internal reinforcements on wave propagation on the composite shell.[11,12]

Some authors use the ray theory to describe helical waves which propagate along finite cylindrical shells.[6,13] In addition, authors of the Ref. 13 study influences of a transversal ring, a transversal bulkhead, joints between hemispherical endcaps and the pipe and caustics of hemispherical endcaps.

In this paper, experimental results of acoustic scattering from thin finite cylindrical shells ended by, on the one hand, hemispherical caps and, on the other hand, circular lids are compared. Helical waves are obliquely generated on the broadside of targets and on the basis of previously published analyses,[14,15] time domain experimental data are examined. Many of the existing differences due to the type of wave and to the shape of objects, are highlighted when the objects are insonified on their cylindrical part.

II. TARGET DESCRIPTION AND EXPERIMENTAL CONFIGURATION

The two objects studied in this work are made by assembling together a number of separate pieces (Fig. 1). In both targets, the main part is composed of two identical cylindrical shells made of stainless steel 347 ($C_T = 3100$ m/s; $C_L = 5790$ m/s; $\rho_s = 7900$ kg m^{-3}), with extremities fitted onto each other and soft soldered together. The extra solder is filed out to obtain a smooth junction, without altering of the thickness. Dimensional characteristics of the cylindrical part are the outer radius $a = 29.3$ mm, the radius ratio $b/a = 0.99$ (b: inner radius) and the length $L = 400$ mm (see Fig. 1). Object A is ended by flat disks of radius a, made of PVC and glued onto each extremity. Object B is closed by stainless steel hemispherical end-caps of the same thickness and radius as the cylindrical part. The soldering process used to

[a]Electronic mail: dominique.decultot@iut.univ-lehavre.fr

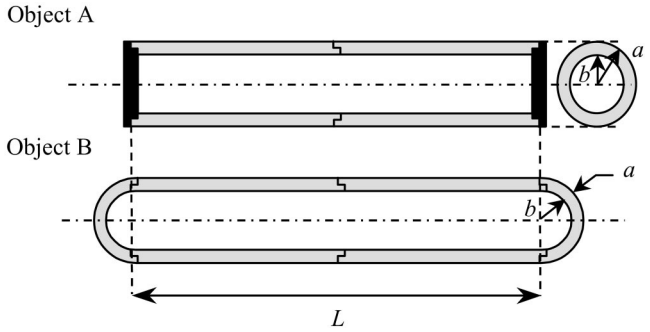


FIG. 1. Assembly of object A and object B.

join the hemispherical parts to the cylindrical one is identical to the process used for the median junction of cylindrical part.

The M.I.I.R. (Method of Isolation and Identification of Resonances) procedure[16] is used to obtain experimental data. The air-filled objects are horizontally immersed in water (sound velocity $C_1 = 1470\ \mathrm{m\,s^{-1}}$ and density $\rho_1 = 1000\ \mathrm{kg\,m^{-3}}$) and are then excited by a short acoustic pulse delivered from a transducer of central frequency equal to 200 kHz ($k_1 a = 25$, where k_1 is the wave number of the incident wave). The backscattered pressure from the object is received by the same transducer [Fig. 2(a)]. The broadband transducer (Panametrics V3507) enables signal investigation in the range $8 \leqslant k_1 a \leqslant 45$: time and frequency characteristics of the transducer are given in Fig. 2(b). The diameter of its circular radiating surface is equal to 50 mm. Throughout the experiment, the distance between the geometric center of the target and the transducer remains equal to 1.5 m. Echo wave forms obtained from objects A and B, at variable incidence angles comprised between 90° (normal incidence to the shell main axis) and 60°, at 1° increments, are cascaded to form Figs. 3(a)–(b), respectively. In the next section, a survey of the echo arrangement analysis from Refs. 14–15 is proposed in order to show the influence of the end-caps.

III. ANALYSIS AND DISCUSSION

In the case of a target end-capped by hemispherical shells (object B), it has been shown that two distinct angular zones can be identified depending on whether the excitation is on the hemispherical or cylindrical portion of the shell.[15] The first zone, from 0° to 50°, concerns excitation on the hemispherical end-cap : the acoustic response of the object is similar to that of axial incidence. The second zone, from 55° to 90°, concerns the excitation of the object on the cylindrical part. In this case, surface waves follow helical travel paths around the cylindrical part and cross through to the hemispherical parts. During the propagation, waves are continuously reradiated into the surrounding fluid. However, due to the monostatic setup, they can only be received by the transducer when r, the number of times the wave crosses hemispherical parts, is odd ($r = 1, 3, ...$). In the same way, considering flat-ended targets (object A), it has been shown that successive series of echoes from the propagation of helical waves may be observed.[14] The corresponding travel paths on the shell include odd numbers r of reflections at the extremities.

Travel time measurements from object A [Fig. 3(a)] are compared to those from object B [Fig. 3(b)]. Different echoes are identified by their arrival times. The geometrical approach[14,15,17,18] takes into consideration wave types, as well as traveled paths in water, on the cylindrical part and on the hemisphere(s) (for shell B only). The general expressions of echo arrival times can be found in Ref. 14 for flat-ended shells, and in Ref. 15 for shells ended by hemispherical caps (the resulting curve networks giving the echo loci in the time-angle representations are not presented here).

A. T_0 wave helical propagation

For flat-ended and hemispherical-ended shells, echoes due to the propagation of the T_0 wave (quasi-shear wave of lowest order) are, respectively, observable on the 60°–90° angular windows of Figs. 3(a)–(b). This angular zone corresponds to observation angles of the T_0 wave on the cylindrical shell. Then, close to 90° the observation of T_0 wave echoes is difficult because the coupling between water and the shell is weak. The lower angular limit 61.7° corresponds to the cut off condition of the observation of T_0 wave which is obtained by the relation $90° - \sin^{-1}(C_1/C_s)$; where C_s is the speed of shear waves in stainless steel. The experimental observation and modeling of helical wave propagation on

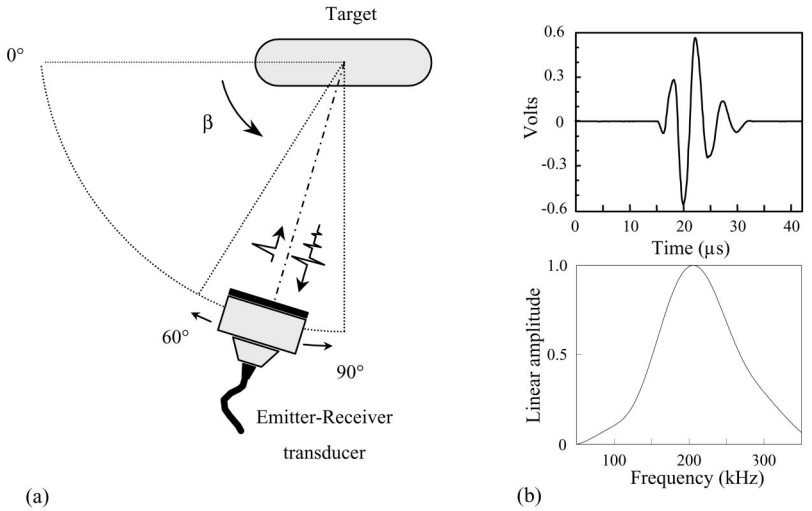


FIG. 2. (a) Experimental monostatic setup; (b) time and frequency characteristics of the Panametrics V3507 transducer.

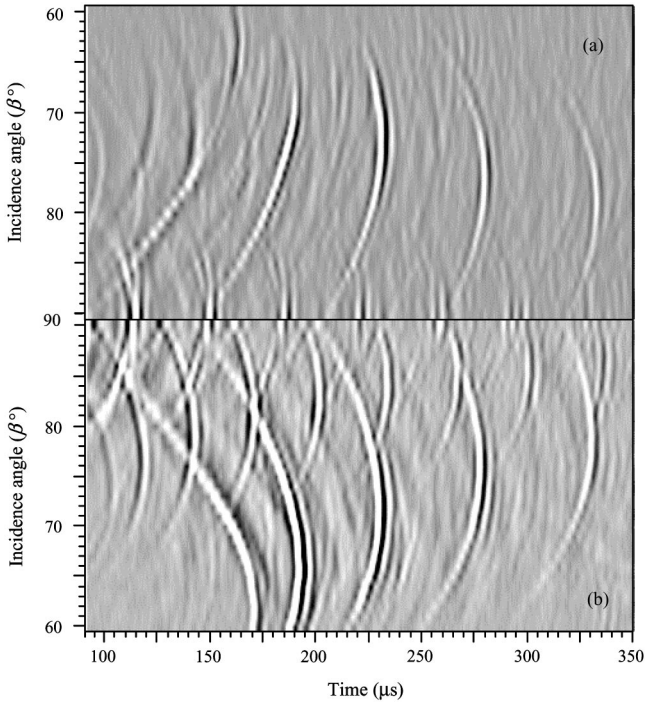


FIG. 3. Evolution of experimental echo arrival times as a function of the incidence angle from object A (a) and object B (b). Amplitude increases linearly from black to white zone. The time origin is fixed by the arrival time of the specular reflection echo at normal incidence ($\beta=90°$).

finite cylindrical shell have previously been achieved in Refs. 3–4, 6–7, 14–15. All echoes in Figs. 3(a)–(b) belong to the first pattern: they have been received from travel paths comprising one reflection at the extremity (shell A), or a single crossing to the hemispherical part (shell B); ($r=1$). For a given incidence angle, the interval between two successive echoes of the pattern is identical in both objects, since it only depends on the radius of the shell. On the contrary, echo arrival times are not identical in both objects. The delay Δt of echo arrival times from shell B with those from shell A has been evaluated using the geometrical approach of Refs. 14–15:

$$\Delta t = r\pi a\left(\frac{\cos\beta}{C_1\tan\alpha}+\frac{1}{C_{\mathrm{gr}}^{\mathrm{sphere}}}-\frac{1}{C_{\mathrm{gr}}^{\mathrm{cylinder}}\sin\alpha}\right). \qquad (1)$$

In relation (1), the incidence angle β and the propagation angle α are measured from the cylinder axis and are linked as follows: $\cos\alpha=\cos\beta/\sin\theta$. The excitation angle is given by $\theta=\sin^{-1}(C_1/C_{\mathrm{ph}}^{T_o})\approx 28.3°$ ($C_{\mathrm{ph}}^{T_o}$: phase velocity of the T_0 wave). Within the frequency range of investigation, the group velocities along the cylindrical and on the spherical surfaces can be taken as $C_{\mathrm{gr}}^{\mathrm{cylinder}}=C_{\mathrm{gr}}^{\mathrm{sphere}}\approx 3110\ \mathrm{m\,s^{-1}}$.[15] For $r=1$, Fig. 4 displays the evolution of Δt against the incidence angle. Calculated time values are comparable to the echo wave form duration (few microseconds). Hence, the discrepancy is too small to be clearly noticed from the observation of Figs. 3(a)–(b). In addition to that, Fig. 4 shows that the delay increases with the deviation from normal incidence. However, echoes of the T_0 wave are not observed when $\beta\leqslant 65°$ in the case of the flat-ended shell [Fig. 3(a)]. This angular limit is slightly above the cut off condition for

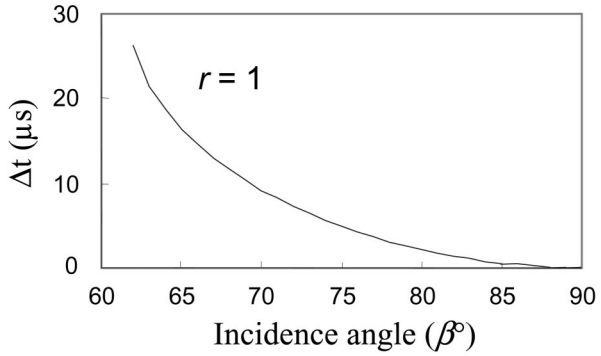


FIG. 4. Evolution of Δt, the calculated delay of echoes from shell B with echoes from shell A for $r=1$, as a function of the incidence angle.

the T_0 wave observation (61.7°); but this feature has already been observed.[14] For the hemispherical ended shells, the angular limit of observation of first T_0 wave echoes is below the cut off condition [Fig. 3(b)]. This feature is certainly due to the T_0 wave excitation at the threshold of the hemispherical part of cylindrical/hemispherical junctions. A comparison between two echo wave forms extracted from Figs 3(a)–(b) at an incidence angle 80° is made in Fig. 5. In both objects, echoes due to the T_0 wave have almost the same arrival time.

B. S_0 wave propagation

In this work two types of echoes related to the S_0 wave (quasi-compressional wave of the lowest order) can be observed depending on whether the propagation is along the axis or not.

(1°) It can be noticed that echoes of the S_0 wave, observable in the range $87°\leqslant\beta\leqslant 90°$, show a similar arrange-

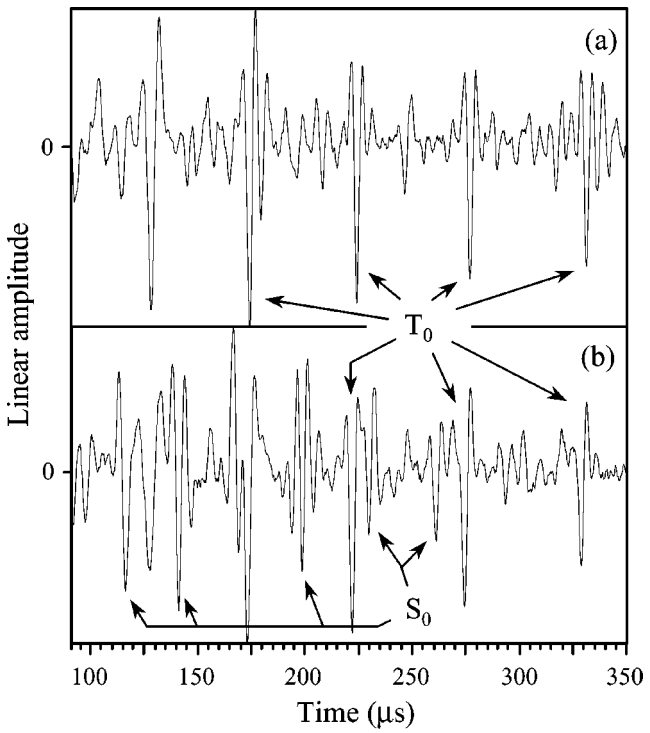


FIG. 5. Echo wave forms extracted from Figs. 3(a)–(b) at incidence angles equal to 80° (upper and lower parts, respectively).

ment in both objects: there is no propagation along the shell axis. These echoes correspond to circumferential propagation around the cylindrical part.

(2°) A significant difference highlighted in comparing Figs. 3(a)–(b) concerns the S_0 wave with helical propagation. Echoes linked to this propagation are only observed in the case of shell B: the continuity of propagation through to the hemispherical parts is observed between 74° and 90°. The difference is made clearer in comparing echo wave forms in Fig. 5.

In conclusion, this study shows that the influence of the type of the extremities of finite cylindrical shells on the scattered pressure significantly depends on the wave type. First, echoes related to the S_0 helical wave are observed only in the case of hemispherical caps. Second, T_0 wave echoes are observed for both hemispherical end-caps and flat-ended shells. But, the differences between the arrival times of echoes are small.

ACKNOWLEDGMENTS

The authors are grateful to Dr. J. G. Chiumia for its careful reading of this manuscript and useful suggestions.

[1] V. V. Muzychenko and S. A. Rybak, ''Low-frequency resonance scattering of sound by finite cylindrical shells (review),'' Sov. Phys. Acoust. **34**, 325–333 (1988).
[2] M. Tran-Van-Nhieu, ''Scattering from a finite cylindrical shell,'' J. Acoust. Soc. Am. **91**, 670–679 (1991).
[3] M. L. Rummerman, ''Contribution of membrane wave reradiation to scattering from finite cylindrical steel shells in water,'' J. Acoust. Soc. Am. **93**, 55–65 (1993).
[4] S. F. Morse, P. L. Marston, and G. Kaduchak, ''High-frequency backscattering enhancements by thick finite cylindrical shells in water at oblique incidence: Experiments, interpretation and calculations,'' J. Acoust. Soc. Am. **103**, 785–794 (1998).
[5] B. H. Houston, J. A. Bucaro, and D. M. Photiadis, ''Broadband acoustic scattering from a ribbed shell,'' J. Acoust. Soc. Am. **98**, 2851–2853 (1995).
[6] F. J. Blonigen and P. L. Marston, ''Leaky helical flexural wave backscattering contributions from tilted cylindrical shells in water: Observations and modeling,'' J. Acoust. Soc. Am. **112**, 528–536 (2002).
[7] S. F. Morse and P. L. Marston, ''Backscattering of transients by tilted truncated cylindrical shells: Time-frequency identification of ray contributions from measurements,'' J. Acoust. Soc. Am. **111**, 1289–1294 (2002).
[8] L. Haumesser, D. Décultot, F. Léon, and G. Maze, ''Experimental identification of finite cylindrical shell vibration modes,'' J. Acoust. Soc. Am. **111**, 2034–2039 (2002).
[9] N. D. Veksler, A. Lavie, and B. Dubus, ''Peripherical waves generated in a cylindrical endcaps by a plane acoustic wave at axial incidence,'' Wave Motion **31**, 349–369 (2000).
[10] D. Décultot, F. Lecroq, G. Maze, J. Ripoche, B. Dubus, and A. Lavie, ''Resonances interpretation with surface waves propagation on cylindrical shells bounded by hemispherical endcaps'' *Ultrasonics International 93 Conference Proceedings* (Butterworths, London, 1993), pp. 479–482.
[11] B. Dubus, N. D. Veksler, and A. Lavie, ''Acoustic wave scattering from a composed shell reinforced by a single rib: characteristics of peripherical waves,'' Ultrasonics **38**, 838–841 (2000).
[12] N. Touraine, ''Diffusion acoustique par des coques cylindrico-hémisphèriques immergées en incidence variable (Acoustic scattering from submerged cylindrical shells enclosed by hemispherical caps for various incidences),'' Ph.D. dissertation, Le Havre University, 2000.
[13] I. T. Lu and Y. Q. Yao, ''Ray shooting, eigenray search, and complex resonances for submerged structures,'' J. Acoust. Soc. Am. **101**, 66–76 (1997).
[14] L. Haumesser, A. Baillard, D. Décultot, and G. Maze, ''Behavior of first guided wave on finite cylindrical shells of various lengths: Experimental investigation,'' J. Acoust. Soc. Am. **109**, 583–590 (2001).
[15] N. Touraine, L. Haumesser, D. Décultot, G. Maze, A. Klauson, and J. Metsaveer, ''Analysis of the acoustic scattering at variable incidences from an extra thin cylindrical shell bounded by hemispherical endcaps,'' J. Acoust. Soc. Am. **108**, 2187–2196 (2000).
[16] G. Maze, and J. Ripoche, ''Méthode d'isolement et d'identification des résonances (M.I.I.R.) de cylindres et de tubes soumis à une onde acoustique plane dans l'eau,'' ''Method of isolation and identification of resonances (M.I.I.R.) of cylinders and cylindrical shells insonified by a plane acoustic wave in water,'' Rev. Phys. Appl. **18**, 319–326 (1983).
[17] X. L. Bao and H. Überall, ''Experimental study of acoustic resonances of elastic spheres and hemispherically endcapped cylinders,'' *Acoustic Resonance Scattering*, edited by H. Überall (Gordon and Breach, New York, 1992), Chap. 8, pp. 161–180.
[18] X. L. Bao, ''Echoes and helical surface waves on finite elastic cylinder excited by sound pulses in water,'' J. Acoust. Soc. Am. **94**, 1461–1466 (1993).

The influence of sexual orientation on vowel production (L)

Janet B. Pierrehumbert and Tessa Bent
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60208–4090

Benjamin Munson[a)]
Department of Speech–Language–Hearing Sciences, University of Minnesota, Minneapolis, Minnesota 55455

Ann R. Bradlow
Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60208–4090

J. Michael Bailey
Department of Psychology, Northwestern University, Swift Hall 102, 2029 Sheridan Road, Evanston, IL 60208–2710

(Received 12 February 2004; revised 30 June 2004; accepted 2 July 2004)

Vowel production in gay, lesbian, bisexual (GLB), and heterosexual speakers was examined. Differences in the acoustic characteristics of vowels were found as a function of sexual orientation. Lesbian and bisexual women produced less fronted /u/ and /ɑ/ than heterosexual women. Gay men produced a more expanded vowel space than heterosexual men. However, the vowels of GLB speakers were not generally shifted toward vowel patterns typical of the opposite sex. These results are inconsistent with the conjecture that innate biological factors have a broadly feminizing influence on the speech of gay men and a broadly masculinizing influence on the speech of lesbian/bisexual women. They are consistent with the idea that innate biological factors influence GLB speech patterns indirectly by causing selective adoption of certain speech patterns characteristic of the opposite sex. [DOI: 10.1121/1.1788729]

PACS numbers: 43.70.Fq, 43.70.Gf [AL] Pages: 1905–1908

I. INTRODUCTION

Sexual orientation, social identity, and language use are all matters of intense interest. These issues come together in the question of whether gay, lesbian, and bisexual (GLB) adults are able to use distinctive speech patterns that convey their social identity. In many cultures, a popular stereotype holds that there are systematic differences in speech production as a function of sexual orientation. Indeed, Carahaly (2000) and Linville (1998) both found that listeners can judge the sexual orientation of speakers at greater than chance levels based on speech samples alone.

A small number of instrumental studies have examined the acoustic characteristics that may cue these judgments. However, these studies are limited in their scope and have yielded mixed results. Linville (1998) reports differences in the duration and spectrum of /s/ for a small group of five gay and four heterosexual men. Gaudio (1994) found no differences in vocal pitch between gay men and heterosexual men. Avery and Liss (1996) report vowel differences relating to perceived effeminacy in men, but they did not obtain information about the actual sexual orientation of their subjects. Furthermore, the few existing instrumental studies of lesbian or bisexual women's speech report null results (Waksler, 2001).

This study goes substantially beyond prior work by examining acoustic characteristics of a small number of vowels produced by a considerably larger ($n=103$) cohort of GLB and heterosexual men and women than has been examined in previous research. Our analysis bears on the ongoing debate about the origins of same-sex attraction and its potential relationship to speech characteristics. Prior research on language acquisition and speech production suggests three alternative theories for how sexual orientation could influence speech patterns. One possibility is that an innate biological factor influences both sexual orientation and the anatomical structures that underlie speech production. A related and more sophisticated hypothesis is that sexual orientation relates to hormonal exposure *in utero*, and that the primary biological reflex in adults is sex-typical versus sex-atypical patterns of neural differentiation. This is the position of Bailey (2003a), who reviews a variety of research demonstrating same-sex attraction to be associated (at least at the group level) with hormonal environment *in utero*. These variants of biological determinism have in common the prediction that the speech patterns of gay men should be shifted toward female norms, compared to those of heterosexual men. The patterns of lesbian women should be shifted toward male norms, compared to those of heterosexual women. The shift could arise directly, if anatomical structures of GLB adults partially resemble those of opposite-sexed heterosexual adults. It could arise indirectly, if patterns of speech motor control resemble those of opposite-sexed heterosexual adults.

A second possibility is that an innate biological factor influences both sexual orientation and the trajectory of language acquisition. Distinctively GLB speech patterns would

[a)]Please direct queries to Benjamin Munson, Ph.D., Assistant Professor, Department of Speech–Language–Hearing Sciences, University of Minnesota, 115 Shevlin Hall, 164 Pillsbury Drive, SE, Minneapolis, Minnesota 55455. Voice: (612) 624-0304; fax: (612) 624-7586; electronic mail: Munso005@umn.edu

be due to the influence of this factor on higher-level aspects of language production, such as attention to adult models during acquisition. This explanation is consistent with theories of sexual orientation that posit a common biological basis for both sexual orientation and vocational and avocational choices. It differs from the first conjecture in that it posits that the common factor influences higher-level aspects of language acquisition and use, rather than anatomical structures.

Finally, GLB speech patterns may be completely learned as a special speech register for the GLB culture (Zwicky, 1997). The existence of such speech registers has been documented widely in other groups [e.g., Eckert, 2000]. This register learning would only begin when people came to identify with a GLB peer group. According to this third possibility, attention to peer models, rather than to opposite-sexed adult models, would be the crucial factor in question.

These theories differ in their predictions about how GLB vowels should pattern in comparison with those produced by heterosexual men and women. Under the first theory, speech patterns of GLB speakers should generally resemble those of opposite-sexed heterosexual speakers. Under the second and third theories, the vowels of GLB speakers could easily differ from those of heterosexual speakers in a way that cannot be characterized as a general displacement. We explore these alternatives by looking at the overall spacing of vowels in the F1/F2 space (a measure of how much GLB vowels are shifted in their ensemble toward those of opposite-sex adults) and vowel-space dispersion (a measure of articulatory effort and precision). We also examine the acoustic characteristics of individual vowels. It has been well established that the acoustic characteristic of vowels are related to both peripheral anatomical patterns (e.g., Lindblom and Sundburg, 1971) and to learned production patterns (e.g., Mendoza-Denton, 2003)

A baseline for understanding differences between GLB and heterosexual people is provided by general speech-production differences between adult males and females. These occur as a function of both anatomical differences and social factors. For males, the larynx becomes more massive at puberty, and its position is lowered. An increased mass causes the fundamental frequency (f0) range of men to be lower than for women. Larynx lowering causes the vocal tract to be longer and differently proportioned for men than for women, with derivable consequences for the vocal tract resonances (Stevens, 1998). On average, men have longer vocal tracts than women (Fitch and Giedd, 1999). As a consequence, their formants tend to be lower than those of women.

Socially conventional differences between male and female speech also exist. Some of these are exaggerations of the patterns that result from anatomical differences. For example, in some cultures women exaggerate the high f0 and breathy voice quality that typically result from their smaller, lighter laryngeal structures (Van Bezooijen, 1995). Young children adopt sex-specific speech traits even before sex-related anatomical differences begin to appear, and the sex of children as young as four years old can be accurately identified from speech (Perry *et al.*, 2001). Such differences are clearly learned through imitation of adult models. The amount of latitude in the system leaves ample room for learning of socially conventional patterns and for individual choices relating to personal identity.

II. METHOD

The data in this study were collected from a large group ($n=103$) of Chicago-area self-identified GLB and heterosexual women and men participating in a broad-based social psychology study of sexual orientation. There were 26 self-identified heterosexual men, 29 self-identified gay men, 16 self-identified heterosexual women, 16 self-identified lesbian women, and 16 self-identified bisexual women. The lesbian and female bisexual groups were combined because no significant differences between them were observed in initial data analyses; they are henceforth referred to as the LB (lesbian/bisexual) women. The speech samples for this study were gathered in the course of a survey of personal and social characteristics related to sexual orientation. As a component of the study, the speakers were recorded, reading a standard set of phonetically balanced sentences (IEEE, 1967). They were not given any instructions regarding speaking style. The talkers were recorded in an academic office using a Shure 10A microphone attached directly to the hard drive of a Celeron 667 mHz personal computer with a Soundblaster sound card. The recording quality was variable, limiting the measures that could be obtained for the present post-hoc analysis.

Four of these sentences (*It's easy to tell the depth of a well; Help the woman get back to her feet; Four hours of steady work faced us*; and *The soft cushion broke the man's fall*) were used as stimuli in a perception study (Bailey, 2003b). Bailey demonstrated that listeners had significant success in judging sexual orientation from this small speech sample. Here, 80 listeners listened to 4 of the sentences and rated each talker on a scale of 1 (''sounds totally straight'') to 7 (''sounds totally gay/lesbian''). Listeners were tested in a quiet university laboratory. Gay men were rated as significantly more-gay sounding than heterosexual men, and LB women as significantly more lesbian-sounding than the heterosexual women. The average value for gay men was 4.6 and for heterosexual men was 3.2; the average value for LB women being 4.3 and that for heterosexual women being 3.2. These differences were significant at the $\alpha<0.05$ level. These results strongly support the claim that speech traits can effectively cue the sexual orientation of many gay, lesbian, and bisexual adults, even in a very neutral communication situation. Objective acoustic differences must be present, at least on average, in the speech signal.

In this study, acoustic measurements of five vowels were made, to gauge the range of acoustic cues to which the listeners in Bailey (2003a) may have been attending. F1, F2, F3, and duration were measured for the vowels /ɑ/ in the word *bɑ̲x*, /i/ in *f̲ee̲t*, /eɪ/ in *mɑ̲kes*, /u/ in *blu̲e* and /æ/ in *bɑ̲ck*. The Praat signal-processing program (Boersma and Weenink, 2003) was used to make acoustic measurements. Formant measurements were taken from Linear Predictive Coding (LPC) formant tracks calculated by Praat using a 20-millisecond window and 10 coefficients. Measurements

were made blindly, without reference to self-reported or perceived sexual orientation of the speakers. Formant tracking errors were hand corrected by a trained phonetician (TB). The measurement used in these analyses was taken from the midpoint of the vowel. All formant measurements were converted from the Hertz scale to the Bark scale (Zwicker and Ternhardt, 1980) prior to statistical analyses. Bark measures allow perceptual distances between vowels to be compared. Four summary measures were taken for each speaker. The first was the mean duration for the five vowels. The second and third were average F1 and F2 values across the five vowels studied (e.g., the centroid of the vowels). The fourth was a measure of overall dispersion in the F1/F2 space. This was measured using a method from Bradlow *et al.* 1996, as the average Euclidian distance from the center of the speaker's F1/F2 space. This measure reflects overall clarity and effort in speaking. F3 data were also analyzed, but results are not reported here because no significant differences were found as a function of sexual orientation.

III. RESULTS

The first analysis focused on vowel-space shift. Mean F1 and F2 values were submitted to a two-factor (sex by sexual orientation) between-subjects MANOVA. There was a significant main effect of sex on F1 ($F[1,99]=37$, $p<0.01$, partial $\eta^2=0.28$) and F2 ($F[1,99]=133$, $p<0.01$, partial $\eta^2=0.58$). There was no significant main effect of sexual orientation. Moreover, there was a small but significant sex by sexual orientation interaction for F2 ($F[1,99]=6.1$, $p<0.01$, partial $\eta^2=0.06$), and a marginal interaction for F1 ($F[1,99]=2.7$, $p=0.10$, partial $\eta^2=0.03$). In *post-hoc* tests of significant main effects, heterosexual women produced higher formant frequencies than heterosexual men, reflecting their shorter vocal tracts ($F[1,41]=21.7$, $p<0.01$ for F1; $F[1,41]=91.8$, $p<0.01$ for F2). LB women produced average F1 and F2 values that were significantly lower than heterosexual females' values ($F[1,46]=4.7$, $p<0.05$ for F1, $F[1,46]=5.7$, $p<0.05$ for F2). However, Fig. 1 shows that this effect is primarily due to the back vowels /ɑ/ and /u/. The F1 and F2 values for /ɑ/ and the F1 value for /u/ all differed significantly as a function of sexual orientation ($F[1,46]>5$, $p<0.01$ for all tests); the F2 value for /u/ differed marginally ($F[1,46]=3$, $p=0.08$). The F1 and F2 frequencies vowels /i/, /e/, and /æ/ were comparable for LB and heterosexual women.

In contrast, average vowel formant values for gay and heterosexual men were not statistically significantly different ($F[1,53]<1$, $p>0.05$ for F1; $F[1,53]=1.3$, $p>0.05$ for F2). Figure 2 shows that gay men produced vowel spaces that were different from those of heterosexual men, but that the direction of the difference varied according to vowel. *Post-hoc* tests showed that gay men produced the vowel /ɑ/ with a significantly lower F2 value and a significantly higher F1 value than heterosexual men ($F[1,52]>4$, $p<0.05$ for both tests). The vowel /i/ had a higher F2 value and a lower F1 value in gay men than in heterosexual men; again, these differences were statistically significant ($F[1,52]>5$, $p<0.01$ for both tests). Finally, /æ/ had a significantly higher F2 and a marginally higher F1 in gay men than in heterosexual men ($F[1,52]>5$, $p<0.01$ for F2, $F[1,52]=3.8$, $p=0.06$ for F1). The vowels /u/ and /eɪ/ did not differ between the two groups of men.

TABLE I. Mean F1 and F2 values, duration, and dispersion for individual vowels produced by the four groups.

Group	Vowel	F1 (bark)	F2 (bark)	Duration (ms)	Dispersion (bark)
Heterosexual women	i	3.74	14.69	140.7	2.94
	e	5.16	13.73	90.1	1.38
	æ	7.21	12.32	126.8	1.61
	a	8.05	10.89	142.6	2.92
	u	4.08	10.84	109.0	2.39
	all	5.65	12.49	121.8	2.25
Heterosexual men	i	3.63	13.45	116.7	2.46
	e	4.72	12.46	84.2	1.20
	æ	6.42	11.23	118.5	1.51
	a	6.79	9.65	139.3	2.58
	u	3.64	10.57	94.3	1.88
	all	5.04	11.47	110.4	1.93
LB women	i	3.55	14.67	134.5	3.09
	e	5.07	13.79	90.3	1.65
	æ	7.13	12.20	122.3	1.78
	a	7.58	10.33	134.8	2.94
	u	3.82	10.26	107.8	2.63
	all	5.43	12.25	117.5	2.42
Gay men	i	3.28	13.77	120.8	2.87
	e	4.78	12.70	85.3	1.25
	æ	6.63	11.27	114.6	1.65
	a	7.10	9.58	132.4	2.88
	u	3.59	10.67	89.8	1.96
	all	5.08	11.59	108.7	2.12

The second analysis focused on vowel-space dispersion. This measure was examined in a two-factor (sex by sexual orientation), between-subjects ANOVA. A significant main effect of sex was found, $F[1,99]=30.9$, $p<0.01$, partial $\eta^2=0.24$. The vowel spaces produced by women were more dispersed than those produced by men. This is consistent with previous research on sex differences in speech clarity and precision (e.g., Bradlow *et al.*, 1996). Moreover, there was a significant main effect of sexual orientation, $F[1,99]=10$, $p<0.01$, partial $\eta^2=0.09$. Both gay men and LB women produced significantly more-expanded vowel spaces than heterosexual speakers. The two factors did not interact, $F[1,99]<1$, $p>0.05$. Inspections of Figs. 1 and 2 suggest that the effect of sexual orientation on vowel-space dispersion was due to different factors for men and women. The greater vowel-space dispersion in women was due to the LB women producing back vowels with lower F2 values. The difference between gay men and heterosexual men was more global. Three of the five vowels showed shifts toward more extreme values. (See Table I.)

An expanded vowel space can result either from a slower speech rate, which permits articulatory targets to be achieved more completely (Moon and Lindblom, 1994), or from greater articulatory precision and effort (Lindblom, 1990). Therefore, vowel durations were analyzed in relation to vowel space dispersion. A two-factor (sex by sexual orientation) between subjects ANOVA showed that women produce significantly longer vowels than men ($F[1,99]=6.2$, $p<0.05$. partial $\eta^2=0.06$). However, there was no significant effect of sexual orientation or interaction of sex with sexual

orientation. These results suggest that the relationship of sexual orientation to vowel-space dispersion reflects differences in articulatory precision and effort rather than speech rate.

IV. DISCUSSION

This analysis found reliable differences in vowel production between GLB and heterosexual men and women. The specific results are inconsistent with the hypothesis that gay men have shorter, more feminine vocal tracts than heterosexual men and that lesbian and bisexual women have longer, more masculine vocal tracts than heterosexual women. Although the vowels of GLB speakers did differ from those of heterosexual speakers, they were not uniformly shifted in the way that this hypothesis predicts. They are also inconsistent with the hypothesis that GLB speakers generically display speech motor control patterns of the opposite sex. Although the gay men did have an expanded vowel space, like women, LB women also displayed an expanded vowel space, contrary to the hypothesis. Furthermore, the expansion of the vowel space was not attributable to all vowels equally. For the LB women, the dominant contributor was more extreme back vowels. For gay men, three out of five vowels had more extreme values, with the effect on /æ/ being the dominant one.

One reasonable interpretation of this finding is that GLB speech patterns reflect learned manipulation of the phonetic space. They are consistent with the suggestion that GLB speakers learn to model the speech of opposite sexed speakers in specific respects. The values for the LB women were intermediate between male and female targets for /u/, and more back than men for /ɑ/. The use of backness in back vowels to convey social identity is not unprecedented. According to one sociolinguistic field study (Habick, 1991), the freedom with which English permits in the production of /u/ is exploited by adolescents to convey social identity. A back variant of /u/ was associated with membership in a group known for its ''tough'' stance. The notion that the LB women were using backness to convey social identity rather than overall masculinity is supported by our finding that they did not mimic the articulatory reduction that is typical of male speech.

Gay men produced vowel spaces with more dispersion than heterosexual men. Since greater precision is also widely reported for women's speech (Bradlow *et al.*, 1996), this could reflect selective learning of a female speech feature. It is noteworthy, however, that the vowels were far from uniformly affected. It is also noteworthy that gay men did not display any analog to the exaggerated diminutivity that has been reported for some female speech populations (Van Bezooijen, 1995). Specifically, we did not find the overall raising of formant values that would result from active articulatory maneuvers to shorten the vocal tract. Linville (1998) found no differences in the average spectrum; the adoption of a breathy, feminine, voice quality would affect this measure. Gaudio (1994) found that gay men do not have higher average f0 than heterosexual men. Thus, the gay men in these studies have at most adopted aspects of female speech that convey social engagement and emotional expressiveness, such as vowel-space dispersion, and not those that would convey diminutivity or subservience, such as a higher f0 or overall higher-scaled formants.

In summary, the distinct speech patterns of GLB speakers do not reflect the direct impact of biological factors on speech production. Instead, they appear to be learned. They could in principle be learned in adolescence as a special speech register that the speakers acquire when they begin to identify with a GLB peer group. However, our results are equally consistent with the idea that young people predisposed to becoming GLB adults (perhaps through a genetic disposition or difference in prenatal environment) selectively attend to certain aspects of opposite-sex adult models during early language acquisition. Future research should examine this question more directly.

Avery, J., and Liss, J. (**1996**). ''Acoustic correlates of less-masculine sounding speech,'' J. Acoust. Soc. Am. **99**, 3738–3748.

Bailey, J. M. (**2003a**). *The Boy Who Would Be Queen: The Science of Gender-Bending and Transsexualism* (Joseph Henry Press, Lincoln, NE).

Bailey, J. M. (**2003b**). *Gender Nonconformity and Sexual Orientation*, paper presented at the annual meeting of the Human Behavior and Evolution Society, 4–8 June, 2003.

Boersma, P., and Weenink, D. (**2003**). *Praat v. 4.1.7 [Computer Software]*, Institute of Phonetic Sciences, Amsterdam.

Bradlow, A., Torretta, G., and Pisoni, D. (**1996**). ''Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics,'' Speech Commun. **20**, 255–272.

Carahaly, L. (**2000**). ''Listener accuracy in identifying the sexual orientation of male and female speakers,'' unpublished M.A. thesis, Department of Speech and Hearing Science, The Ohio State University, Columbus, OH.

Eckert, P. (**2000**). *Linguistic Variation as Social Practice: the Linguistic Construction of Identity in Belten High* (Blackwell, Malden, MA).

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.

Gaudio, R. (**1994**). ''Sounding gay: Pitch properties in the speech of gay and straight men,'' Am. Speech **69**, 30–57.

Habick, T. (**1991**). ''Burnouts versus rednecks: effects of group membership on the phonemic system,'' in *New Ways of Analyzing Sound Change*, edited by P. Eckert (Academic Press, San Diego).

IEEE (**1967**). Standards Publication No. 297, IEEE Recommended Practice for Speech Quality Measurements, au-17 (3), pp. 225–246.

Lindblom, B. (**1990**). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W. J., and Marchal, A., in *Speech Production and Speech Modeling* (Kluwer Academic, Dordrecht), pp. 403–439.

Lindblom, B., and Sundburg, J. (**1971**). ''Acoustical consequences of lip, tongue, jaw, and larynx movement,'' J. Acoust. Soc. Am. **50**, 1166–1179.

Linville, S. (**1998**). ''Acoustic correlates of perceived versus actual sexual orientation in men's speech,'' Pholia Phoniatrica et Logopaedica **50**, 35–48.

Mendoza-Denton, N., Hay, J., and Jannedy, S. (**2003**). ''Probabilistic sociolinguistics: beyond variable rules,'' in *Probabilistic Linguistics*, edited by R. Bod, J. Hay, and S. Jannedy (MIT Press, Cambridge, MA).

Moon, B., and Lindblom, B. (**1994**). ''Interaction between duration, context, and speaking style in English stressed vowels,'' J. Acoust. Soc. Am. **96**, 40–55.

Perry, T. L., Ohde, R., and Ashmead, D. (**2001**). ''The acoustic bases for gender identification from children's voices,'' J. Acoust. Soc. Am. **109**, 2988–2998.

Stevens, K. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Van Bezooijen, R. (**1995**). ''Sociocultural aspects of pitch differences between Japanese and Dutch women,'' Lang Speech **38**, 253–265.

Waksler, S. (**2001**). ''Pitch range and women's sexual orientation,'' Word **52**, 69–77.

Zwicker, E., and Terhardt, E. (**1980**). ''Analytical expression for critical-band rate and critical bandwidth as a function of frequency,'' J. Acoust. Soc. Am. **68**, 1523–1525.

Zwicky, A. (**1997**). ''Two lavender issues for linguists,'' in *Queerly Phrased*, edited by K. Hall and A. Livia (Oxford University Press, New York), pp. 21–32.

Source parameter estimates of echolocation clicks from wild pygmy killer whales (*Feresa attenuata*) (L)

P. T. Madsen[a)]
Ocean Alliance, 191 Western Rd., Lincoln, Massachusetts 01773; and Department of Zoophysiology, Institute of Biological Sciences, University of Aarhus, DK-8000 Aarhus C, Denmark

I. Kerr and R. Payne
Ocean Alliance, 191 Western Rd., Lincoln, Massachusetts 01773

(Received 9 April 2004; revised 30 June 2004; accepted 12 July 2004)

Pods of the little known pygmy killer whale (*Feresa attenuata*) in the northern Indian Ocean were recorded with a vertical hydrophone array connected to a digital recorder sampling at 320 kHz. Recorded clicks were directional, short (25 μs) transients with estimated source levels between 197 and 223 dB $re.$ 1 μPa (pp). Spectra of clicks recorded close to or on the acoustic axis were bimodal with peak frequencies between 45 and 117 kHz, and with centroid frequencies between 70 and 85 kHz. The clicks share characteristics of echolocation clicks from similar sized, whistling delphinids, and have properties suited for the detection and classification of prey targeted by this odontocete. [DOI: 10.1121/1.1788726]

PACS numbers: 43.80.Ev, 43.80.Ka, 43.80.Jz [WWA] Pages: 1909–1912

I. INTRODUCTION

Most of the knowledge on odontocete biosonars stems from an intense research on a few delphinid species. It has been demonstrated that dolphins use biosonar to probe their environment, and their sonar system has, at ranges up to 100 m, detection and discrimination capabilities surpassing the performance of manmade analogs (Au, 1993). While such investigations have shed essential light on the basic performance of odontocete biosonar systems, it is not clear if data on the transmission system from trained animals studied in captivity are representative of the signals free ranging animals produce while using their biosonar for orientation and food finding in natural habitats (Au, 1993; Au and Herzing, 2003).

In this paper we report the first data on source properties of clicks from free ranging pygmy killer whales (*Feresa attenuata*). The pygmy killer whale (Feresa) among the smaller dolphins with a body length of around 2.3 m and a weight around 150 kg. They are usually found in off-shore tropical waters in groups of 5–30 animals, where they forage on a variety of food items including, fish, cephalopods and apparently also other small delphinids (Ross and Leatherwood, 1994). Most of the knowledge on Feresa is limited to morphometrics and stomach contents collected from stranded specimens, and there is little or no data on the ecology, behavior, life history and acoustics of this odontocete species (for a review, see Ross and Leatherwood, 1994).

We quantify and discuss characteristics of Feresa clicks and compare them to the properties of clicks of other odontocetes with biosonar recorded in captivity and in natural habitats.

[a)]Present address: Woods Hole Oceanographic Institution, Wood Hole, Massachusetts 02543. Electronic mail: pmadsen@whoi.edu

II. MATERIALS AND METHODS

A. Platform and recording gear

The recording gear consisted of a vertical array of three hydrophones deployed 5–7 m from the research vessel (for details see Madsen *et al.*, 2004).

Signals were digitized with a Wavebook 512 (IOtech), 12 bit ADC, sampling at 320 kHz on each of the three channels. Each recording session lasted 20 s with an additional 5 s of off-load time from the WBK30 memory. The LP filter, acting as an analog anti-alias filter, was compensated for during analysis yielding a flat ($\pm$2 dB) frequency response of the recording system between 1 and 160 kHz.

B. Signal analysis

Analysis was performed with Cool Edit Pro (*Syntrillium*) and custom written routines in Matlab 6.0 (*Mathworks*) (for details see Madsen *et al.*, 2004). Signal duration (τ, μs) was determined by 97% the relative signal energy derived by integrating the squared pressure over an interpolated (10 steps) 64 point window symmetrical around the peak of the signal envelope (Fig. 1). Received rms sound pressure level (dB $re.$ 1 μPa rms) was calculated by integrating the square of the instantaneous pressure as a function of time over the time window τ relative to the same integral over the same time τ of a calibration signal. Energy flux density (dB $re.$ 1 μPa^2 s) was defined as the rms sound pressure level $+10\log(\tau)$ (*sensu* Au, 1993).

The spectral characteristics of the signals were quantified from a 256 point Fast Fourier Transform (FFT) on Hanning windowed data symmetrical around the peak of the signal envelopes. The peak frequency (f_p, kHz), centroid frequency (f_0, kHz), -3 dB BW (kHz), -10 dB BW (kHz), and centralized root mean square bandwidth (rms-BW, kHz) [Fig. 1(c)] were derived *sensu* Au (1993).

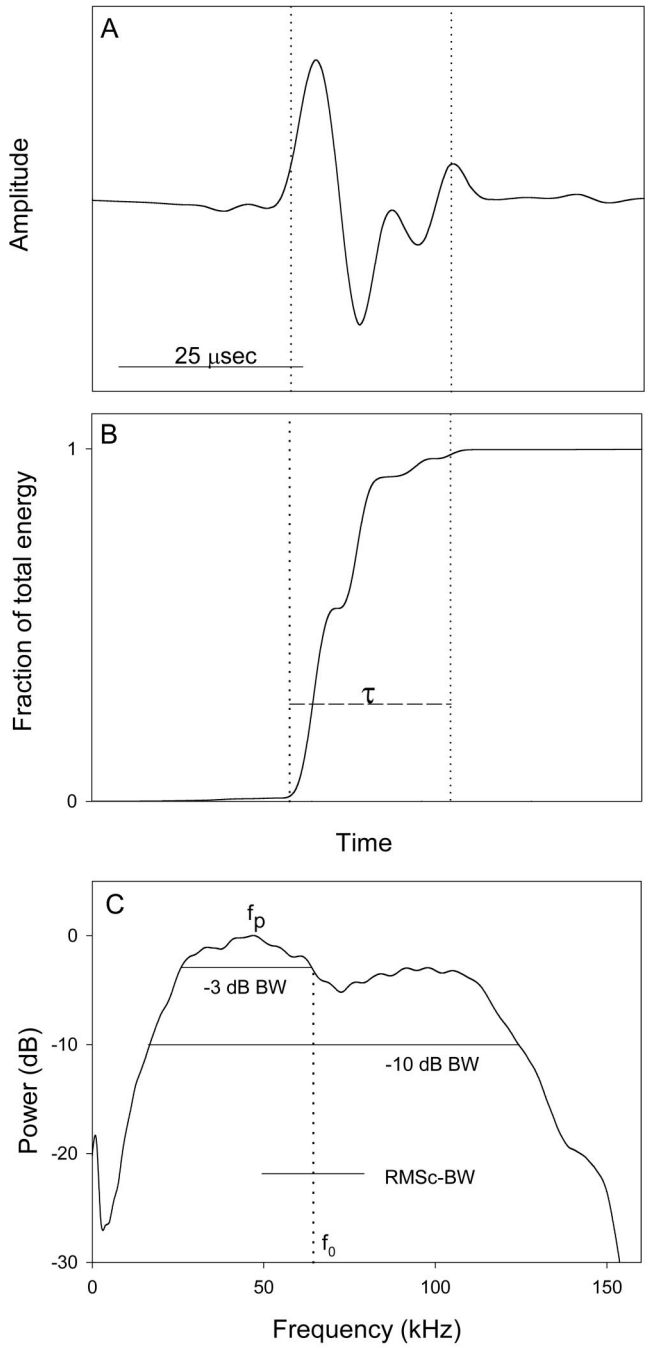


FIG. 1. (a) The waveform of the ultra-short waveform of a Feresa click. (b) The relative energy in a 64-point frame as a function of time derived as the cumulative squared pressure of the waveform displayed in (a). Signal duration, τ, is defined as the window (indicated by the dashed lines) containing 97% of the energy in the 64-point window. (c) Power spectrum of the click in (a) calculated with a 256-point FFT on Hanning windowed data. The bin width is 1.25 kHz. The relevant parameters describing the properties of the spectrum are displayed.

C. Sound source localization and estimation of source parameters

The location of the sound source was estimated from time of arrival differences (TOAD) at the three receivers by using a trigonometric approach (see, e.g., Lammers and Au, 2003). The range between the source and the receivers was calculated from the Pythagorean theorem in a localization routine implemented in Matlab (courtesy of M. Wahlberg).

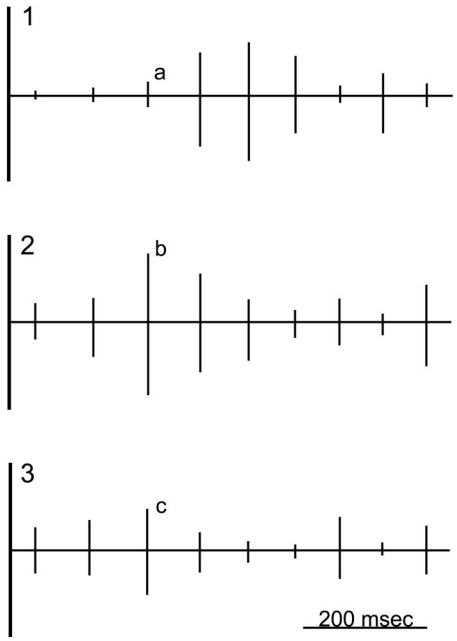


FIG. 2. Sounds tracks of the three hydrophones deployed at 4 (1), 8 (2), and 12 (3) meters depth. The same click train recorded in three different aspects is displayed in the three tracks. The full amplitude on the *y* axis corresponds to a tone with a sound pressure level of 213 dB//1 μPa (pp). Note how the ASL anomaly shifts between the receivers as the likely result of a directional sound source scanning different parts of the array. Click b is an example of a click classified as being recorded on or close to the acoustic axis of the sound beam, and (a) and (c) are recorded at angles of 15° and 7° off-axis, respectively. The tracks have been high pass filtered at 5 kHz (20 dB/octave).

Source levels (SL) were calculated from the following equation: SL=RL+TL. Transmission loss (TL) was estimated by $TL=20\log(R)+R\alpha$, where α is the frequency-dependent absorption at the centroid frequency of the received click. The term apparent source level (ASL, *sensu* Møhl *et al.*, 2000) is used to emphasize that RL+TL equals the back-calculated sound pressure level one meter from a directional source of unknown orientation. The term source level (SL) can only be used where the recording aspect equals the axis of the sound beam. Source properties derived from hydrophones in a position off the acoustic axis have little relevance for the performance of the sonar system and it is therefore important to report properties measured on or close to the acoustic axis along with reliable SL estimates (Au and Herzing, 2003).

As can be seen in Fig. 2, the ASL of the same click recorded with different hydrophones from different aspects varies considerably. These profound amplitude changes over time on each of the hydrophones are presumably the result of scanning movements of a directional sound beam ensonifying the array. It is therefore clicks with the highest ASL's in such click trains that are most likely to represent the properties of sonar signals close to or on the acoustic axis of the clicking animal. All recordings were carefully examined manually, and only signals with maximal, relative amplitude on the center hydrophone compared to the other two hydrophone tracks in ensonifications were classified as being close to or on the acoustic axis of the clicking animal (e.g., see click b of Fig. 2).

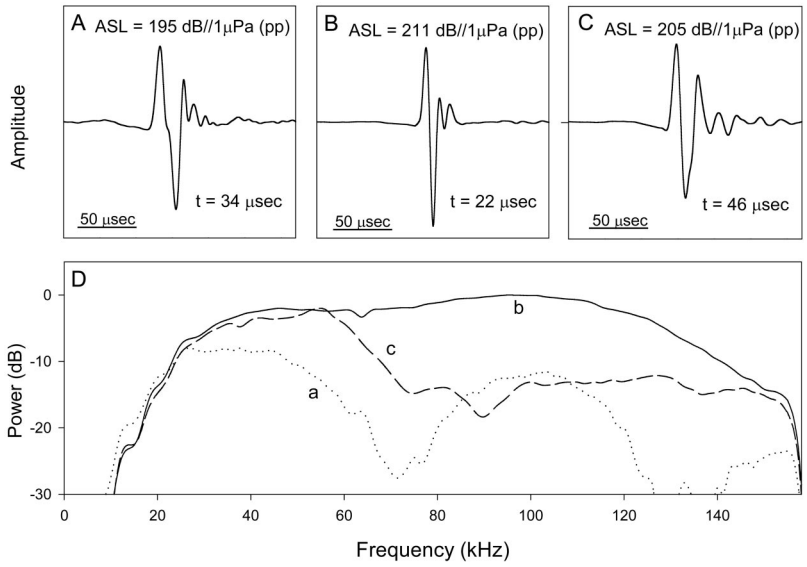


FIG. 3. Waveforms (A), (B), (C) of the click displayed as (a), (b), and (c) in Fig. 2, along with apparent source levels and durations. The ASL of (B) is likely to be the source level for that click. Figure (D) shows the power spectra of the click displayed in three different aspects in (A), (B), (C). The spectra were computed with a 256-point FFT on Hanning windowed data. The bin width is 1.25 kHz. Note that the click recorded on or close to the acoustic axis (b) is broadband with high-frequency components compared to the same click recorded at increasing angles off-axis [(c), 7°: (a), 15°], where it suffers from a low pass filter effect and notches in the spectrum.

III. RESULTS

The first of the recordings commenced on 8 March 2003, at position 1°19N/73°57E off the Huvadu Atoll in the Maldivian archipelago, where the water depth is some 1800 m. A group of 8–12 Feresa circled the research vessel and recording gear for approximately 30 min during which 49 sessions, containing more than 5000 clicks, were recorded. The second recording took place on 28 April 2003, south of Dondra head, Sri Lanka, at position 5°37N/80°43E were the water depth is roughly 2500 m. A group of 10–15 Feresa approached the research vessel, and circled it for 10 min. A total number of 18 sessions, containing some 1100 clicks, were recorded.

Using the previously outlined selection criteria, we identified a total of 26 click trains where animals ensonified the recording gear with their directional sound beams (see, e.g., Fig. 2). The clicks were part of long click trains fading in and out of the background noise with interclick intervals (ICI) between 50 and 120 ms, corresponding to instantaneous repetition rates of 8–20 click/s. On axis clicks have back-calculated source levels between 197 and 223 dB $re.$ 1 μPa (pp). The rms SLs are 12–14 dB lower. The waveforms have short durations of 20–40 μs, leading to energy flux density SLs between 130–165 dB $re.$ 1 μPa^2 s. Clicks recorded off the acoustic axis are of longer duration, have lower ASLs and show low-pass filtered, distorted spectra compared to the same click recorded on or close to the acoustic axis [Fig. 3].

The spectra of on-axis clicks are broadband with a -10 dB BW around 100 kHz, rms BW of around 32 kHz, and Q values between 2 and 3. Click spectra are bimodal with a stable peak around 40 kHz and a more variable peak around 100 kHz. The peaks are within ± 6 dB of each other with the high-frequency peak dominating for clicks with high source levels [Fig. 3(c)], and the low-frequency peak dominating clicks with lower source levels (Fig. 1). Therefore, peak frequency is not a good measure of the frequency emphasis of broadband spectra (Au *et al.*, 1995). The centroid frequency is a more robust measure, and it appears that Feresa clicks have centroid frequencies between 70 and 85 kHz.

IV. DISCUSSION AND CONCLUSION

A problem arising during an analysis of ultrasonic, directional clicks from free ranging odontocetes of unknown orientation in relation to the hydrophones is determining whether a signal is recorded on the acoustic axis. Since clicks recorded off the acoustic axis are distorted and not representative of the properties of the on-axis signal used by the animal for biosonar it is critical to determine if the signals have been recorded on or close to the acoustic axis of the clicking animal. As outlined previously in the materials and methods section, we have classified as the best available candidates for on-axis signals, clicks with maximum amplitudes in scans registered on the center hydrophone. When working with free-ranging animals an inherent problem with this approach is that, at least theoretically, none of the clicks may actually have been recorded on the acoustic axis. On the other hand, if an animal continuously ensonifies the recording system with a series of clicks with low source levels, they will not be classified as being on axis. The following discussion is made with these reservations since, when working with free-ranging animals, there is, as yet, no analytical method that can provide a rigid on–off classification for broadband clicks.

Feresa emits short duration, broadband signals similar to a large number of delphinids that have been measured in captivity (Au *et al.*, 1974) and in the wild (Rasmussen *et al.*, 2002, Au and Herzing, 2003; Schotten *et al.*, 2003; Madsen *et al.*, 2004, Au *et al.*, 2004). The centroid frequencies are higher than was has been reported for the larger, free ranging false killer whale (*Pseudorca crassidens*) (Madsen *et al.*, 2004), but are comparable to the centroid frequencies and properties of high SL clicks from similar sized dolphins such as *Tursiops*, *Lagenorhyncus*, and *Stenella*, and also the larger *Grampus*.

When recorded simultaneously on the acoustic axis and at different angles off it, the ASLs of off-axis clicks are much lower than the estimated SL defined by the on-axis version of the same clicks. This directionality is also seen in the frequency domain where the spectra of off-axis clicks are low

pass filtered with increasing azimuth, and where step null regions are seen to begin forming at lower and lower frequencies (Fig. 3). This off-axis distortion is consistent with a directional signal whose transfer function is that of a broadband transient signal radiating from a piston of finite aperture. Pistons have successfully been used to model the transmitting part of the sonar systems in other odontocetes (Au, 1993), and it appears that the transmitting system of Feresa operates in a similar fashion. Given that transmitting aperture scales with the size of the nasal structures (Au *et al.*, 1999), it may be surmised that the transmitting aperture of a Feresa is larger than that of Phocoena, but smaller than that of Tursiops. If so, the equation of Au *et al.* (1999) suggests that the directionality of Feresa clicks is larger than Phocoena (DI=22 dB) but smaller than Tursiops (DI=26 dB). Hence, it can be concluded that Feresa clicks possess the needed directionality to reduce clutter, and to render generation of high source levels energetically feasible for a biological sound source.

Spectra are not only affected by directionality, but also by the acoustic output. Figure 1(a) displays an on-axis click with a source level of 203 dB $re.$ 1 μPa (pp), and the computed spectrum is depicted in Fig. 1(c). It is seen that the lower-frequency peak dominates the spectrum. This is in contrast to the spectrum of a click with a source level of 212 dB $re.$ 1 μPa shown in Fig. 3(b), in which the high-frequency peak dominates. The relationship is consistent with the sound production in other delphinids producing short, broadbanded clicks (Au, 1993).

Even though this study does not demonstrate echolocation, it seems reasonable to conclude that the directional, ultrasonic clicks, and the context in which they are used, is consistent with a biosonar function. The whales were localized at ranges between 13 and 52 m from the array, and under the assumption that they echolocated on the deployed recording gear, it is seen that the click intervals are in the same interval of 40–100 ms as used by trained Tursiops echolocating at similar target ranges (Au, 1993).

For prey localization during foraging it cannot be assured that the animals would use clicks with the same properties as measured when they echolocate on recording gear. Nevertheless, the broadband, powerful click properties derived in the present study should provide good temporal resolution and discrimination capabilities in detection of prey targets (Au, 1993). Assuming that the detection and signal processing of Feresa equals that of other delphinids, it can be conjectured that the source properties, including SLs up to 223 dB//1 μPa (pp), would allow for echolocation of relevant cephalopod and fish prey at ranges of 50 to 200 m, as has been estimated for other free ranging delphinids (Au *et al.*, 2004, Madsen *et al.*, 2004).

The present study, along with other array-based investigations (Møhl *et al.*, 1990; Au and Herzing, 2003; Au *et al.*, 2004; Møhl *et al.*, 2000; Rasmussen *et al.*, 2002; Schotten *et al.*, 2003; Madsen *et al.*, 2004), show that basic source parameters of clicks can be derived from free ranging odontocetes in coastal and off-shore habitats, given that the implemented on-axis selection criteria render a representative sample of clicks to be measured.

ACKNOWLEDGMENTS

M. Preedy, the Odyssey crew, the staff of Ocean Alliance, N. Kristiansen, M. Bjørn, M. Wahlberg, A. Surlykke, S. B. Pedersen, B. K. Nielsen, and B. Møhl provided technical support and/or constructive criticism Ocean Alliance and *Voyage of the Odyssey* funded ship time. PTM was funded by postdoctoral grants from the Ocean Alliance and the Danish Natural Science Research Council. Gear was funded by the Oticon Foundation and the Novo Nordisk Science Foundation Dr. W. W. L. Au kindly provided constructive criticism and access to a paper in press. This work was carried out under NFMS Permit No. 751-1641, Research Permit No. FA-A/33/2003/01 from Fisheries Research in the Maldives, and Research Permit No. WL/3/2/1/18 from National Aquatic Resources Agency/The Wildlife Conservation Office of Sri Lanka.

Au, W. W., Floyd, R. W., Penner, R. H., and Murchison, A. E. (**1974**). ''Measurement of echolocation signals of the Atlantic bottlenose dolphin, Tursiops truncatus Montagu, in open waters,'' J. Acoust. Soc. Am. **56**, 1280–1290.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer Verlag, New York).

Au, W. W. L., Kastelein, R. A., Rippe, T., and Schooneman, N. M. (**1999**). ''Transmission beam pattern and echolocation signals of a harbor porpoise (Phocoena phocoena),'' J. Acoust. Soc. Am. **106**, 3699–3705.

Au, W. W., and Herzing, D. L. (**2003**). ''Echolocation signals of wild Atlantic spotted dolphin (Stenella frontalis),'' J. Acoust. Soc. Am. **113**, 598–604.

Au, W. W. L., Ford, J. K., Horne, J. K., and Allman, K. A. N. (**2004**). ''Echolocation signals of free-ranging killer whales (Orcinus orca) and modelling of foraging for Chinook salmon (Oncorhyncus tshawytscha),'' J. Acoust. Soc. Am. **115**, 1–9.

Lammers, M. O., and Au, W. W. L. (**2003**). ''Directionality in the whistles of Hawaiian spinner dolphins (Stenella longirostris): A signal feature to cue direction of movement?'' Marine Mammal Sci. **19**, 249–264.

Madsen, P. T., Kerr, I., and Payne, R. (**2004**). ''Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: False killer whales (*Pseudorca crassidens*) and Risso's dolphins (*Grampus griseus*),'' J. Exp. Biol. **207**, 1811–1823.

Møhl, B., Surlykke, A., and Miller, L. A. (**1990**). ''High intensity Narwhal click,'' *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 295–304.

Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (**2000**). ''Sperm whale clicks: directionality and source level revisited,'' J. Acoust. Soc. Am. **107**, 638–648.

Rasmussen, M. H., Miller, L. A., and Au, W. W. L. (**2002**). ''Source levels of clicks from free-ranging white-beaked dolphins (Lagenorhynchus albirostris Gray 1846) recorded in Icelandic waters,'' J. Acoust. Soc. Am. **111**, 1122–1125.

Ross, G. J. B., and Leatherwood, S. (**1994**). ''Pygmy killer whale Feresa attenuata Gray, 1874,'' *Handbook of Marine Mammals*, edited by S. H. Ridgway and R. Harrison (Academic, London), pp. 387–404.

Schotten, M., Au, W. W. L., Lammers, M. O., and Aubauer, R. (**2003**). ''Echolocation recordings and localizations of wild spinner dolphins (*Stenella longirostris*) and pantropical spotted dolphins (*Stenella attenuata*) using a four hydrophone array,'' in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. F. Moss, and M. Vater (University of Chicago Press, Chicago).

Scattering matrix properties with evanescent modes for waveguides in fluids and solids

V. Pagneux[a)]
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, Av. Olivier Messiaen, 72085 Le Mans Cedex 9, France

A. Maurel[b)]
Laboratoire Ondes et Acoustique, UMR CNRS 7587, Ecole Supérieure de Physique et de Chimie Industrielles, 10 rue Vauquelin, 75005 Paris, France

(Received 23 May 2003; revised 11 June 2004; accepted 24 June 2004)

Reciprocity, energy conservation, and time-reversal invariance are three general properties of the wave fields that imply algebraic scattering matrix properties. In this paper, these scattering matrix properties are established for waveguides when evanescent modes are taken into account. The situations correspond to guided acoustic pressure waves in fluids and Lamb waves in solids treated with the same formalism. The relations between the three properties verified by the scattering matrix are then discussed, and it is found that, as soon as two properties are verified, the third is also verified. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1786293]

PACS numbers: 43.20.Fn, 43.20.Gp, 43.20.Mv [MO] Pages: 1913–1920

I. INTRODUCTION

A very convenient quantity to characterize the scattering of a wave by a scattering region is the so-called scattering matrix (S matrix). The versatility of this formulation is due to the fact that it relates the outgoing wave to the ingoing wave directly, with respect to the scattering region, which are intuitive quantities.[1–4] The scattering problem is then to determine the S matrix and not to determine the wave field in the whole space.[5,6] From an experimental point of view, the S matrix is a very convenient tool since the measurements have to be performed outside of the scattering region only and do not have to be intrusive.

In waveguides, when the interest is only in the far field of the scattering region, the S matrix is restricted to the propagative components. Then, the fundamental properties of the wave propagation (i.e., reciprocity, energy conservation, and time-reversal symmetry) are very simply translated into relationships verified by the S matrix.[7] For instance, the energy conservation implies that the S matrix is unitary with proper normalization. Besides their fundamental interest, these algebraic properties of the S matrix can be very useful in experimental and numerical works, where they provide a convenient way to check if the fundamental properties of the wave propagation are verified. Nevertheless, as soon as the near field is considered, the S matrix has to include the evanescent waves; such a need can exist for example if several scattering regions are taken into account and if they are close enough. Then, the usual properties of the S matrix, correct for the propagative waves only, are not correct anymore.[8,9]

In this paper, we investigate the general relationships verified by the S matrix with evanescent waves in the cases of the propagation in 2D waveguides in fluids or solids. The first case corresponds to the guided propagation of a scalar wave in a fluid with hard wall and the second case corresponds to the guided propagation of either a scalar wave (SH wave) or a vectorial wave (Lamb wave) in an elastic waveguide with free boundaries. We treat the waveguides in fluids and solids with the same formalism that appears to be useful to consider the S matrix.

The plan of the paper is as follows. In Sec. II, we define the scattering matrix for the waveguides. Then, the method used to project the acoustic fields on the waveguide modes is presented in Sec. III. In Sec. IV, we find the three properties verified by the S matrix due to reciprocity, time-reversal symmetry, and energy conservation of the wave propagation, when evanescent modes are taken into account. Finally, in Sec. V, we summarize the three properties and discuss the relation between the three properties of the S matrix: we find that as soon as two properties are verified the third relation is automatically verified.

II. DEFINITION OF THE SCATTERING MATRIX

We consider the problem corresponding to Fig. 1. It consists of a bidimensional waveguide of longitudinal axis x with constant height for $x \leqslant 0$ and $x \geqslant L$. In the scattering region, $0 < x < L$, the waveguide is of variable height $h(x)$. The harmonic time dependence with pulsation ω is $e^{-i\omega\tau}$ and it will be omitted in the following.

Outside of the scattering region, the wave fields can be expressed as a sum over the transverse modes of the homogeneous waveguide with coefficients depending on x. These coefficients can be split into right-going components $\mathbf{A}$ and left-going components $\mathbf{B}$, that is, $\mathbf{A}$ and $\mathbf{B}$ are vectors whose components are the projections of the wave field on the wave modes. For the sake of clarity, we note in the following $\mathbf{A}^I$ (respectively, $\mathbf{B}^I$) at $x=0$ and $\mathbf{A}^{II}$ (respectively, $\mathbf{B}^{II}$) at $x=L$. We define $\mathbf{\Psi}_{\text{in}}$ as the ingoing waves and $\mathbf{\Psi}_{\text{out}}$ as the outgoing waves, with respect to the scattering region

a)Electronic mail: vincent.pagneux@univ-lemans.fr
b)Electronic mail: agnes.maurel@espci.fr

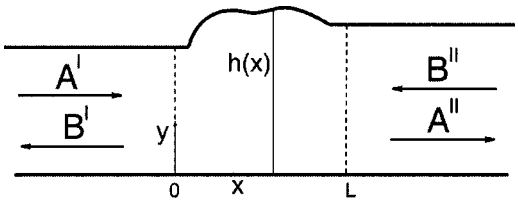


FIG. 1. Geometry of the waveguide.

$$\mathbf{\Psi}_{\text{in}}=\begin{pmatrix}\mathbf{A}^I\\ \mathbf{B}^{II}\end{pmatrix},\quad \mathbf{\Psi}_{\text{out}}=\begin{pmatrix}\mathbf{B}^I\\ \mathbf{A}^{II}\end{pmatrix}. \tag{2.1}$$

Then, the scattering matrix links together the outgoing waves $\mathbf{\Psi}_{\text{out}}$ and the ingoing waves $\mathbf{\Psi}_{\text{in}}$ at both extremities of the scattering region

$$\mathbf{\Psi}_{\text{out}}=\mathsf{S}\mathbf{\Psi}_{\text{in}}. \tag{2.2}$$

Actually, S is defined with the reflection and transmission matrices R^I, T^I (respectively R^{II}, T^{II}) corresponding to wave incident from the left (respectively, from the right)

$$\mathsf{S}=\begin{pmatrix}\mathsf{R}^I & \mathsf{T}^{II}\\ \mathsf{T}^I & \mathsf{R}^{II}\end{pmatrix}. \tag{2.3}$$

Here, R and T are matrices linking ingoing and outgoing wave components. Note that the scattering matrix could be used for the case of more than two terminating waveguides.

III. FORMALISM

A. System

Guided wave propagation in fluids and solids is considered. The fluid case corresponds to the Helmholtz equation on the pressure p associated with boundary condition $\partial_{\mathbf{n}}p=0$ on the walls $y=0$ and $y=h(x)$ ($\mathbf{n}$ denotes the vector normal to the wall). The solid case corresponds to the Navier equation on the displacement vector $\mathbf{w}$ with the boundary condition for a wall free of traction $\sigma\cdot\mathbf{n}=0$ on $y=0$ and $y=h(x)$, where $\sigma=\lambda\nabla\cdot\mathbf{w}I+\mu(\nabla\mathbf{w}+{}^t\nabla\mathbf{w})$ denotes the stress tensor, and (λ,μ) are the Lamé's constants. Results can be easily generalized to inhomogeneous media with variable ρ, λ, and μ and to other boundary conditions, e.g., $p=0$ for fluids or $\mathbf{w}=0$ for solids on the walls.

We choose to present both cases (fluid and solid) in the same formalism. This is done working with two quantities $\mathbf{X}$ and $\mathbf{Y}$ presented below. In addition to permitting a unified presentation, that formalism allows us to easily tackle the projection on the Lamb modes (for propagation in solids). The idea is to write the equations as an evolution equation (with respect to the axis x of the waveguide) on $\mathbf{X}$ and $\mathbf{Y}$ that leads to a canonical eigenvalue problem in the transverse direction when transverse modes are sought as in Sec. III B. For solids, this formulation is similar to the one presented recently in Ref. 10 in that it describes the evolution of a stress-displacement 4-vector, but here that 4-vector is suitably split in two 2-vectors that permit one to project easily on the transverse modes.[11]

For fluids, $\mathbf{X}$ and $\mathbf{Y}$ are scalar quantities defined by

$$\mathbf{X}=\frac{\partial}{\partial x}p \quad \text{and} \quad \mathbf{Y}=p.$$

For solids, they are 2-vectors

$$\mathbf{X}=\begin{pmatrix}u\\ t\end{pmatrix} \quad \text{and} \quad \mathbf{Y}=\begin{pmatrix}-s\\ v\end{pmatrix},$$

where

$$\mathbf{w}=\begin{pmatrix}u\\ v\end{pmatrix}$$

is the displacement vector and

$$\sigma=\begin{pmatrix}s & t\\ t & r\end{pmatrix}$$

the corresponding stress tensor.

It is shown in Appendix A that both sets of equations can be written in the same generic form

$$\frac{\partial}{\partial x}\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}=\begin{pmatrix}0 & \mathsf{F}\\ \mathsf{G} & 0\end{pmatrix}\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}, \tag{3.1}$$

with boundary conditions

$$\begin{aligned}&(\mathsf{C}_{\mathsf{X}}^0+h'\mathsf{C}_{\mathsf{X}}^1)\mathbf{X}+(\mathsf{C}_{\mathsf{Y}}^0+h'\mathsf{C}_{\mathsf{Y}}^1)\mathbf{Y}=0, \quad \text{at } y=h(x),\\ &\mathsf{C}_{\mathsf{X}}^0\mathbf{X}+\mathsf{C}_{\mathsf{Y}}^0\mathbf{Y}=0, \quad \text{at } y=0,\end{aligned} \tag{3.2}$$

where $h'=dh/dx$.

Expressions of F, G, $\mathsf{C}_{\mathsf{X}}^i$, and $\mathsf{C}_{\mathsf{Y}}^i$ $(i=0,1)$ in both cases are given in Appendix A. In the following, we use two properties of these matrices

(1) (Property 1): F, G, $\mathsf{C}_{\mathsf{X}}^i$, and $\mathsf{C}_{\mathsf{Y}}^i$ $(i=0,1)$ are real;
(2) (Property 2): $(\mathsf{F}\mathbf{Y}|\tilde{\mathbf{Y}})+(\mathbf{X}|\mathsf{G}\tilde{\mathbf{X}})=(\mathbf{Y}|\mathsf{F}\tilde{\mathbf{Y}})+(\mathsf{G}\mathbf{X}|\tilde{\mathbf{X}})+h'(\mathbf{Y}\cdot\tilde{\mathbf{X}}-\mathbf{X}\cdot\tilde{\mathbf{Y}})(h)$,

where $(\mathbf{X},\mathbf{Y})$ and $(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})$ are two solutions of Eq. (3.1) with boundary condition (3.2) and $(\mathbf{U}|\mathbf{V})=\int_0^{h(x)}\mathbf{U}(x,y)\cdot\mathbf{V}(x,y)dy$ denotes a bilinear form. Properties 1 and 2 are demonstrated in Appendix B. In Sec. IV, it will appear that property 1 is related to the time-reversal invariance, property 2 is related to the reciprocity, and both properties are related to energy conservation.

It can be noticed that both properties 1 and 2 involve the boundary conditions. For instance in the fluid case, if the walls were lined, property 2 would be conserved while property 1 would not.

B. Modal decomposition

The scattering matrix S links together the right-going components and the left-going components through (2.2). In this section, we expose the modal decomposition that permits us to define the right-going components $\mathbf{A}$ and left-going components $\mathbf{B}$.

The transverse modes used in the decomposition, denoted $\mathbf{X}_n(y)$ and $\mathbf{Y}_n(y)$, correspond to the natural basis in a uniform waveguide. They are associated with a wave number k_n and are defined by

$$ik_n\begin{pmatrix}\mathbf{X}_n\\ \mathbf{Y}_n\end{pmatrix}=\begin{pmatrix}0 & \mathsf{F}\\ \mathsf{G} & 0\end{pmatrix}\begin{pmatrix}\mathbf{X}_n\\ \mathbf{Y}_n\end{pmatrix}, \tag{3.3}$$

with boundary conditions corresponding to Eq. (3.2) setting $h'=0$

$$\mathsf{C}_{\mathsf{X}}^0\mathbf{X}_n+\mathsf{C}_{\mathsf{Y}}^0\mathbf{Y}_n=0, \quad \text{at } y=0 \quad \text{and } y=h(x). \tag{3.4}$$

These modes are well known: they are proportional to the cosine functions[12] in the fluid case and are the Lamb modes in the solid case[13] (their expressions are given in Appendix C). They are linked by a biorthogonality relation

$$(\mathbf{X}_n|\mathbf{Y}_m)=\mathcal{J}_n\delta_{nm}. \tag{3.5}$$

This relation corresponds to the orthogonality of the cosine functions in the fluid case and to the Fraser's biorthogonality relation in the solid case.[14] In the fluid case, $\mathcal{J}_n=ik_n$, and in the solid case, its expression can be found in Ref. 11. Here, this biorthogonality relation can be easily demonstrated using (3.3) and the symmetry property of F and G when they are applied to $\mathbf{X}_n$ and $\mathbf{Y}_n$: $(\mathsf{F}\mathbf{Y}_n|\mathbf{Y}_m)=(\mathbf{Y}_n|\mathsf{F}\mathbf{Y}_m)$ and $(\mathsf{G}\mathbf{X}_n|\mathbf{X}_m)=(\mathbf{X}_n|\mathsf{G}\mathbf{X}_m)$. The biorthogonality equation (3.5) is of interest because it permits one to easily project $\mathbf{X}$ and $\mathbf{Y}$ on the modes $\mathbf{X}_n$ and $\mathbf{Y}_n$.

In the following, k_n is indexed by $n>0$ when it refers to a right-going wave and $n<0$ when it refers to a left-going wave. Using the symmetry properties of the basis $\mathbf{X}_{-n}=-\mathbf{X}_n$ and $\mathbf{Y}_{-n}=\mathbf{Y}_n$, $\mathbf{X}$ and $\mathbf{Y}$ are decomposed as

$$\begin{aligned}\mathbf{X}(x,y)&=\sum_{n>0}(A_n(x)-B_n(x))\mathbf{X}_n(y),\\ \mathbf{Y}(x,y)&=\sum_{n>0}(A_n(x)+B_n(x))\mathbf{Y}_n(y).\end{aligned} \tag{3.6}$$

The S matrix is concerned with values of the components $\mathbf{A}$ and $\mathbf{B}$ at $x=0$ and $x=L$ (Fig. 1). At these two x positions, the values of $\mathbf{X}_n$, $\mathbf{Y}_n$, and $\mathcal{J}_n$ [Eq. (3.5)] are *a priori* distinct and we note in the following $\mathbf{X}_n^\alpha$, $\mathbf{Y}_n^\alpha$, and $\mathcal{J}_n^\alpha$ with $\alpha=I$, II for $\mathbf{X}_n$, $\mathbf{Y}_n$, and $\mathcal{J}_n$ on the cross section $x=0$, L, respectively. At each cross section, we define the matrices

$$\begin{aligned}\mathsf{J}^I_{mn}&=(\mathbf{X}^I_n|\mathbf{Y}^I_m)=\mathcal{J}^I_n\delta_{mn},\\ \tilde{\mathsf{J}}^I_{mn}&=(\mathbf{X}^I_n|\overline{\mathbf{Y}^I_m}),\end{aligned} \tag{3.7}$$

and the same for J^{II} and $\tilde{\mathsf{J}}^{II}$. Then, J and M are such that

$$\mathsf{J}=\begin{pmatrix}\mathsf{J}^I & 0\\ 0 & \mathsf{J}^{II}\end{pmatrix}, \quad \mathsf{J}\mathsf{M}=\begin{pmatrix}\tilde{\mathsf{J}}^I & 0\\ 0 & \tilde{\mathsf{J}}^{II}\end{pmatrix}. \tag{3.8}$$

By definition, J is a diagonal matrix [Eq. (3.7)]. In the fluid case, M is simply the identity matrix while in the solid case, M has a more complicated structure, as illustrated in Sec. V and detailed in Appendix C.

IV. PROPERTIES OF THE SCATTERING MATRIX

A. Reciprocity

The reciprocity relation corresponds to a relation between two solutions $(\mathbf{X},\mathbf{Y})$ and $(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})$ of (3.1)–(3.2). We show below that this relation can be written

$$(\mathbf{X}^I|\tilde{\mathbf{Y}}^I)-(\tilde{\mathbf{X}}^I|\mathbf{Y}^I)=(\mathbf{X}^{II}|\tilde{\mathbf{Y}}^{II})-(\tilde{\mathbf{X}}^{II}|\mathbf{Y}^{II}). \tag{4.1}$$

This comes from

$$\begin{aligned}&d_x((\mathbf{X}|\tilde{\mathbf{Y}})-(\tilde{\mathbf{X}}|\mathbf{Y}))\\ &=\left(\frac{\partial}{\partial x}\mathbf{X}\middle|\tilde{\mathbf{Y}}\right)+\left(\mathbf{X}\middle|\frac{\partial}{\partial x}\tilde{\mathbf{Y}}\right)-\left(\frac{\partial}{\partial x}\tilde{\mathbf{X}}\middle|\mathbf{Y}\right)-\left(\tilde{\mathbf{X}}\middle|\frac{\partial}{\partial x}\mathbf{Y}\right)\\ &\quad+h'(\mathbf{X}\tilde{\mathbf{Y}}-\tilde{\mathbf{X}}\mathbf{Y})(h)=(\mathsf{F}\mathbf{Y}|\tilde{\mathbf{Y}})+(\mathbf{X}|\mathsf{G}\tilde{\mathbf{X}})-(\mathsf{F}\tilde{\mathbf{Y}}|\mathbf{Y})\\ &\quad-(\tilde{\mathbf{X}}|\mathsf{G}\mathbf{X})+h'(\mathbf{X}\tilde{\mathbf{Y}}-\tilde{\mathbf{X}}\mathbf{Y})(h)=0.\end{aligned}$$

The result is deduced from property 2 since $(\mathbf{X},\mathbf{Y})$ and $(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})$ are solutions of (3.1)–(3.2). This form of the reciprocity relation is similar to the one found in Ref. 15 in the fluid case, where it is called a reciprocity theorem of the convolution type. Obviously, Eq. (4.1) can also be deduced from the usual integral representation of the reciprocity property.

Using the modal decomposition (3.6) and the biorthogonality (3.5), the reciprocity relation (4.1) takes the form

$${}^t\mathbf{A}^I\mathsf{J}^I\tilde{\mathbf{B}}^I+{}^t\mathbf{B}^{II}\mathsf{J}^{II}\tilde{\mathbf{A}}^{II}-{}^t\mathbf{B}^I\mathsf{J}^I\tilde{\mathbf{A}}^I-{}^t\mathbf{A}^{II}\mathsf{J}^{II}\tilde{\mathbf{B}}^{II}=0.$$

Here, we have used $\mathsf{J}^\alpha={}^t\mathsf{J}^\alpha$, $\alpha=I$, II. With Eqs. (2.1) and (3.8), this latter expression is equivalent to

$${}^t\boldsymbol{\Psi}_{\text{in}}\mathsf{J}\tilde{\boldsymbol{\Psi}}_{\text{out}}-{}^t\boldsymbol{\Psi}_{\text{out}}\mathsf{J}\tilde{\Psi}_{\text{in}}=0.$$

Eventually, using the scattering matrix S, we obtain

$$\mathsf{J}\mathsf{S}-{}^t\mathsf{S}\mathsf{J}=0, \tag{4.2}$$

which is the property of the scattering matrix induced by reciprocity of the propagation in the scattering region. It has to be noted that Eq. (4.2) has the same form with or without evanescent modes.

B. Time-reversal invariance

In both cases, fluid and solid, matrices F, G, $\mathsf{C}_{\mathsf{X}}^i$, and $\mathsf{C}_{\mathsf{Y}}^i$ $(i=0,1)$ in (3.1) are real (property 1). As a consequence, if $(\mathbf{X},\mathbf{Y})$ is a solution of (3.1), then $(\overline{\mathbf{X}},\overline{\mathbf{Y}})$ is also a solution of (3.1). This solution corresponds to the time-reversed solution, noted $(\mathbf{X}^R,\mathbf{Y}^R)$ in the following. Thus, the time-reversal invariance is translated in the harmonic regime by

$$\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}\text{solution}\Rightarrow\begin{pmatrix}\mathbf{X}^R=\overline{\mathbf{X}}\\ \mathbf{Y}^R=\overline{\mathbf{Y}}\end{pmatrix}\text{solution}. \tag{4.3}$$

To obtain a property for the S matrix from the time reversal, the idea is to use $\boldsymbol{\Psi}^R_{\text{out}}=\mathsf{S}\boldsymbol{\Psi}^R_{\text{in}}$ and to express $\boldsymbol{\Psi}^R_{\text{out}}$ and $\boldsymbol{\Psi}^R_{\text{in}}$ with $\overline{\boldsymbol{\Psi}_{\text{in}}}$ as $\boldsymbol{\Psi}^R_{\text{in}}=K(\mathsf{S})\overline{\boldsymbol{\Psi}_{\text{in}}}$ and $\boldsymbol{\Psi}^R_{\text{out}}=L(\mathsf{S})\overline{\boldsymbol{\Psi}_{\text{in}}}$; thereafter, the relation $L(\mathsf{S})=\mathsf{S}K(\mathsf{S})$ is deduced. We detail this calculation below.

According to the modal decomposition, we have (3.6)

$$\begin{aligned}&\mathbf{X}^R=\sum_n(A_n^R-B_n^R)X_n, \text{ and } \mathbf{Y}^R=\sum_n(A_n^R+B_n^R)Y_n,\\ &\overline{\mathbf{X}}=\sum_n(\overline{A_n}-\overline{B_n})\overline{X_n}, \text{ and } \overline{\mathbf{Y}}=\sum_n(\overline{A_n}+\overline{B_n})\overline{Y_n}.\end{aligned} \tag{4.4}$$

With $\mathbf{X}^R=\overline{\mathbf{X}}$ and $\mathbf{Y}^R=\overline{\mathbf{Y}}$, the equalities $(\mathbf{X}^R|\mathbf{Y}_n)=(\overline{\mathbf{X}}|\mathbf{Y}_n)$ and $(\mathbf{Y}^R|\mathbf{X}_n)=(\overline{\mathbf{Y}}|\mathbf{X}_n)$ become, using (3.5) and (4.4)

$$(A_n^R - B_n^R)\mathcal{J}_n = \sum_{m>0} (\overline{A_m} - \overline{B_m})(\overline{\mathbf{X}_m}|\mathbf{Y}_n),$$

$$(A_n^R + B_n^R)\mathcal{J}_n = \sum_{m>0} (\overline{A_m} + \overline{B_m})(\overline{\mathbf{Y}_m}|\mathbf{X}_n).$$

Summing and subtracting these relations, at $x=0$ ($\alpha=I$) and at $x=L$ ($\alpha=II$), leads to

$$2\mathsf{J}^\alpha \mathbf{A}^{R,\alpha} = (\tilde{\mathsf{J}}^\alpha + {}^t\overline{\tilde{\mathsf{J}}^\alpha})\overline{\mathbf{A}^\alpha} + (\tilde{\mathsf{J}}^\alpha - {}^t\overline{\tilde{\mathsf{J}}^\alpha})\overline{\mathbf{B}^\alpha},$$

$$2\mathsf{J}^\alpha \mathbf{B}^{R,\alpha} = (\tilde{\mathsf{J}}^\alpha - {}^t\overline{\tilde{\mathsf{J}}^\alpha})\overline{\mathbf{A}^\alpha} + (\tilde{\mathsf{J}}^\alpha + {}^t\overline{\tilde{\mathsf{J}}^\alpha})\overline{\mathbf{B}^\alpha},$$

for $\alpha=I$, II. Using the definitions (2.1) and (2.2), these relations become

$$2\mathsf{J}\boldsymbol{\Psi}_{\text{in}}^R = \mathsf{H}_p\overline{\boldsymbol{\Psi}_{\text{in}}} + \mathsf{H}_m\overline{\boldsymbol{\Psi}_{\text{out}}} = (\mathsf{H}_p + \mathsf{H}_m\bar{\mathsf{S}})\overline{\boldsymbol{\Psi}_{\text{in}}},$$

$$2\mathsf{J}\boldsymbol{\Psi}_{\text{out}}^R = \mathsf{H}_m\overline{\boldsymbol{\Psi}_{\text{in}}} + \mathsf{H}_p\overline{\boldsymbol{\Psi}_{\text{out}}} = (\mathsf{H}_m + \mathsf{H}_p\bar{\mathsf{S}})\overline{\boldsymbol{\Psi}_{\text{in}}},$$

where $\mathsf{H}_p = \mathsf{JM} + {}^t(\overline{\mathsf{JM}})$, $\mathsf{H}_m = \mathsf{JM} - {}^t(\overline{\mathsf{JM}})$, and $\overline{\boldsymbol{\Psi}_{\text{out}}} = \bar{\mathsf{S}}\overline{\boldsymbol{\Psi}_{\text{in}}}$ have been used.

Finally, knowing that $\boldsymbol{\Psi}_{\text{out}}^R = \mathsf{S}\boldsymbol{\Psi}_{\text{in}}^R$, we obtain the time-reversal invariance property for the scattering matrix

$$\mathsf{JSJ}^{-1}(\mathsf{H}_p + \mathsf{H}_m\bar{\mathsf{S}}) = \mathsf{H}_m + \mathsf{H}_p\bar{\mathsf{S}}. \tag{4.5}$$

This time-reversal invariance property of the S matrix with evanescent modes is different from the version without evanescent modes because of the extra terms involving H_p (see Sec. V).

C. Energy conservation

The energy conservation comes directly from the reciprocity relation and time-reversal invariance [taking in the reciprocity relation $(\mathbf{X},\mathbf{Y})$ and $(\mathbf{X}^R=\overline{\mathbf{X}}, \mathbf{Y}^R=\overline{\mathbf{Y}})$ as solutions]

$$(\mathbf{X}^I|\overline{\mathbf{Y}^I}) - (\overline{\mathbf{X}^I}|\mathbf{Y}^I) = (\mathbf{X}^{II}|\overline{\mathbf{Y}^{II}}) - (\overline{\mathbf{X}^{II}}|\mathbf{Y}^{II}). \tag{4.6}$$

This relation can be expressed, for $\alpha=I$, II, as the conservation of the energy flux proportional to $W_\alpha = (\mathbf{X}^\alpha|\overline{\mathbf{Y}^\alpha}) - (\overline{\mathbf{X}^\alpha}|\mathbf{Y}^\alpha)$

$$\begin{aligned} W_\alpha &= {}^t\mathbf{A}^\alpha(\tilde{\mathsf{J}}^\alpha - {}^t\overline{\tilde{\mathsf{J}}^\alpha})\overline{\mathbf{A}^\alpha} - {}^t\mathbf{B}^\alpha(\tilde{\mathsf{J}}^\alpha + {}^t\overline{\tilde{\mathsf{J}}^\alpha})\overline{\mathbf{A}^\alpha} \\ &\quad + {}^t\mathbf{A}^\alpha(\tilde{\mathsf{J}}^\alpha + {}^t\overline{\tilde{\mathsf{J}}^\alpha})\overline{\mathbf{B}^\alpha} - {}^t\mathbf{B}^\alpha(\tilde{\mathsf{J}}^\alpha - {}^t\overline{\tilde{\mathsf{J}}^\alpha})\overline{\mathbf{B}^\alpha} \\ &= \text{constant}. \end{aligned}$$

This can be written

$${}^t\boldsymbol{\Psi}_{\text{in}}\mathsf{H}_m\overline{\boldsymbol{\Psi}_{\text{in}}} + {}^t\boldsymbol{\Psi}_{\text{in}}\mathsf{H}_p\overline{\boldsymbol{\Psi}_{\text{out}}} - {}^t\boldsymbol{\Psi}_{\text{out}}\mathsf{H}_p\overline{\boldsymbol{\Psi}_{\text{in}}} - {}^t\boldsymbol{\Psi}_{\text{out}}\mathsf{H}_m\overline{\boldsymbol{\Psi}_{\text{out}}} = 0.$$

Finally, we obtain the energy conservation property for the scattering matrix

$$\mathsf{H}_m + \mathsf{H}_p\bar{\mathsf{S}} = {}^t\mathsf{S}(\mathsf{H}_p + \mathsf{H}_m\bar{\mathsf{S}}), \tag{4.7}$$

where H_p implies extra terms due to the presence of evanescent modes.

As for the reciprocity relation, Eq. (4.6) can also be deduced from an integral representation. It would permit one to generalize Eq. (4.6) to geometry with more than two terminating waveguides.

V. DISCUSSION

We have seen in both cases (fluid and solid) that the reciprocity relation, the time-reversal invariance, and the energy conservation lead to three relations on the scattering matrix S

$$\text{Reciprocity relation:} \quad \mathsf{JS} - {}^t\mathsf{SJ} = 0, \tag{5.1a}$$

Time-reversal invariance:

$$\mathsf{H}_m + \mathsf{H}_p\bar{\mathsf{S}} = \mathsf{JSJ}^{-1}(\mathsf{H}_p + \mathsf{H}_m\bar{\mathsf{S}}), \tag{5.1b}$$

$$\text{Energy conservation:} \quad \mathsf{H}_m + \mathsf{H}_p\bar{\mathsf{S}} = {}^t\mathsf{S}(\mathsf{H}_p + \mathsf{H}_m\bar{\mathsf{S}}). \tag{5.1c}$$

In these relations, the presence of evanescent modes is taken into account by the H_p matrix. Notably, the term involving H_p in the energy conservation represents the energy flux carried by evanescent modes.

Note that the same kind of relations have been shown in Ref. 9 in the angular spectrum representation, but in the scalar case only.

A. Structure of H_m and H_p

To gain some hints of the structures of H_p and H_m, let us take an example in both cases. In the fluid case, let us assume that we have, at $x=0$, one propagating mode $\mathcal{J}_1^I = jk_1^I$ (k_1^I real) and we account for one evanescent mode k_2^I (purely imaginary); at $x=L$, we suppose that we have two propagating modes k_1^{II} and k_2^{II} and we account for one evanescent mode k_3^{II}. With $\mathsf{M}=I$ (Appendix C), we have thus

$$\mathsf{H}_p = 2i\begin{pmatrix} 0 & & & & \\ & k_2^I & & & \\ & & 0 & & \\ & & & 0 & \\ & & & & k_3^{II} \end{pmatrix}, \qquad \mathsf{H}_m = 2i\begin{pmatrix} k_1^I & & & & \\ & 0 & & & \\ & & k_1^{II} & & \\ & & & k_2^{II} & \\ & & & & 0 \end{pmatrix}. \tag{5.2}$$

In the solid case, properties of $\mathcal{J}$ are given in Appendix B. Let us assume that, at $x=0$, we have one propagating mode ($\mathcal{J}_1^I$ purely imaginary) and we account for two evanescent modes such that $\mathcal{J}_3^I = \overline{\mathcal{J}_2^I}$; at $x=L$, we assume that we have only one propagating mode and one evanescent mode with a purely real $\mathcal{J}_2^{II}$. In this case, M takes the form

$$\mathsf{M}=\begin{pmatrix} 1 & & & & \\ & 0 & 1 & & \\ & 1 & 0 & & \\ & & & 1 & \\ & & & 1 & \end{pmatrix}. \tag{5.3}$$

Thus, we have

$$\mathsf{H}_p=2\begin{pmatrix} 0 & & & & \\ & 0 & \mathcal{J}_2^I & & \\ & \overline{\mathcal{J}_2^I} & 0 & & \\ & & & 0 & \\ & & & & \mathcal{J}_2^{II} \end{pmatrix},$$

$$\mathsf{H}_m=2\begin{pmatrix} \mathcal{J}_1^I & & & & \\ & 0 & 0 & & \\ & 0 & 0 & & \\ & & & \mathcal{J}_1^{II} & \\ & & & & 0 \end{pmatrix}. \tag{5.4}$$

B. Relationships without evanescent modes

A consequence of the results of the preceding section is that the three relationships on the S matrix can be simplified to

$$\text{Reciprocity relation:} \quad \mathsf{JS}-{}^t\mathsf{SJ}=0, \tag{5.5a}$$

$$\text{Time-reversal invariance:} \quad \mathsf{H}_m=\mathsf{JSJ}^{-1}\mathsf{H}_m\bar{\mathsf{S}}, \tag{5.5b}$$

$$\text{Energy conservation:} \quad \mathsf{H}_m={}^t\mathsf{SH}_m\bar{\mathsf{S}}, \tag{5.5c}$$

when the modes are restricted to propagating components, since in that case H_p is zero. Then, as it can be expected by a proper normalization, the reciprocity property implies that S is symmetric and the energy conservation implies that S is unitary.

C. Relations between reciprocity, time-reversal, and energy conservation

So far it has been assumed that the propagation is governed by the equation (3.1) with boundary conditions (3.2). It corresponds to the Helmholtz equation for fluids bounded by rigid walls or to the Navier equation for solids with traction free boundaries. Nevertheless, even if the governing equations in the scattering region are unknown, each equation in (5.1) can be used as a test of the corresponding physical property.

It is clear that if the reciprocity relation is verified ($\mathsf{JSJ}^{-1}={}^t\mathsf{S}$,), then time-reversal invariance and energy conservation are equivalent (i.e., time-reversal invariance implies energy conservation and vice versa).

It is more difficult to prove that the reciprocity relation comes from time-reversal invariance and energy conservation. To do that, we use the following procedure. The complex conjugate of the time-reversal invariance is written in the form

$$\bar{\mathsf{S}}\bar{\mathsf{J}}^{-1}(\bar{\mathsf{H}}_p+\bar{\mathsf{H}}_m\mathsf{S})=\bar{\mathsf{J}}^{-1}(\bar{\mathsf{H}}_m+\bar{\mathsf{H}}_p\mathsf{S}). \tag{5.6}$$

Then, the energy conservation is written

$$(\mathsf{H}_p-{}^t\mathsf{SH}_m)\bar{\mathsf{S}}=({}^t\mathsf{SH}_p-\mathsf{H}_m). \tag{5.7}$$

Finally, using (5.6) in (5.7) to eliminate $\bar{\mathsf{S}}$ yields a relation between ${}^t\mathsf{S}$ and S

$$(\mathsf{H}_p-{}^t\mathsf{SH}_m)\bar{\mathsf{J}}^{-1}(\bar{\mathsf{H}}_m+\bar{\mathsf{H}}_p\mathsf{S}) =({}^t\mathsf{SH}_p-\mathsf{H}_m)\bar{\mathsf{J}}^{-1}(\bar{\mathsf{H}}_p+\bar{\mathsf{H}}_m\mathsf{S}). \tag{5.8}$$

Thus, we have

$${}^t\mathsf{SH}_a\mathsf{S}-\mathsf{H}_a+{}^t\mathsf{SH}_b-\mathsf{H}_b\mathsf{S}=0, \tag{5.9}$$

with

$$\mathsf{H}_a=\mathsf{H}_p\bar{\mathsf{J}}^{-1}\overline{\mathsf{H}_m}+\mathsf{H}_m\bar{\mathsf{J}}^{-1}\overline{\mathsf{H}_p}, \tag{5.10}$$

$$\mathsf{H}_b=\mathsf{H}_p\bar{\mathsf{J}}^{-1}\overline{\mathsf{H}_p}+\mathsf{H}_m\bar{\mathsf{J}}^{-1}\overline{\mathsf{H}_p}. \tag{5.11}$$

Using that, by construction, J is diagonal, we obtain a simple expression of H_a and H_b

$$\mathsf{H}_a=2(\mathsf{JM}\bar{\mathsf{M}}-{}^t(\mathsf{M}\bar{\mathsf{M}})\mathsf{J}), \tag{5.12}$$

$$\mathsf{H}_b=2(\mathsf{JM}\bar{\mathsf{M}}+{}^t(\mathsf{M}\bar{\mathsf{M}})\mathsf{J}). \tag{5.13}$$

It is shown in Appendix C that, in both fluid and solid cases, $\mathsf{MM}=\mathrm{I}$. As a consequence, we find $\mathsf{H}_a=0$ and $\mathsf{H}_b=4\mathsf{J}$, so that Eq. (5.9) corresponds to the reciprocity relation.

We have demonstrated that as soon as two relations of (5.1) are verified, the third relation is also verified.

VI. CONCLUDING REMARKS

The main results of this paper are the relationships in Eq. (5.1) that are verified by the scattering matrix with evanescent modes. To the best of our knowledge, this is the first time that such general equations are found for the fluid case and the solid case. These results can be easily generalized to other boundary conditions, to waveguide with inhomogeneous media (λ and μ variable for instance), and to 3D geometry. They are exact whatever the distance from the scattering region, in contrast to usual relations involving only the propagating modes and assuming that one is in the far field of the scattering region. The equation (5.1) can be useful, in a numerical calculation, to test the energy conservation, time-reversal invariance, or the reciprocity of a scattering region. They could be useful also to test the results of an experiment if this latter is able to provide the evanescent components.

Besides, with evanescent modes taken into account, we have shown that, as soon as two physical properties among energy conservation, time-reversal invariance, and reciprocity are verified, the third is also verified. Consequently, this reduces the number of tests that have to be conducted with (5.1). These relationships can be also helpful to works related to time-reversal mirror in waveguides.[17]

APPENDIX A: EVOLUTION EQUATIONS IN WAVEGUIDES

1. Fluid case

We start from the Helmholtz equation

$$\nabla^2 p + \frac{\omega^2}{c^2} p = 0, \tag{A1}$$

where p denotes the acoustic pressure and c the acoustic wave speed in the considered fluid. In an x-axis waveguide of height $h(x)$, the boundary condition $\partial_{\mathbf{n}} p = 0$ (with $\mathbf{n}$ the normal to the wall) can be written

$$\begin{aligned} &-h'(x)\frac{\partial}{\partial x} p + \frac{\partial}{\partial y} p = 0, \quad \text{at } y = h(x), \\ &\frac{\partial}{\partial y} p = 0, \quad \text{at } y = 0. \end{aligned} \tag{A2}$$

With $\mathbf{X} = (\partial/\partial x)p$ and $\mathbf{Y} = p$, (A1) takes the form

$$\frac{\partial}{\partial x}\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} 0 & -\left(\frac{\partial^2}{\partial y^2} + \omega^2/c^2\right) \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}, \tag{A3}$$

and the boundary condition (A2)

$$\begin{aligned} &-h'\mathbf{X} + \frac{\partial}{\partial y}\mathbf{Y} = 0, \quad \text{at } y = h(x) \quad \text{and} \\ &\frac{\partial}{\partial y}\mathbf{Y} = 0, \quad \text{at } y = 0. \end{aligned} \tag{A4}$$

We thus identify $\mathsf{F} = -[(\partial^2/\partial y^2) + \omega^2/c^2]$, $\mathsf{G} = 1$, $\mathsf{C}^0_{\mathsf{X}} = 0$, $\mathsf{C}^1_{\mathsf{X}} = -1$, $\mathsf{C}^0_{\mathsf{Y}} = \partial/\partial y$, and $\mathsf{C}^1_{\mathsf{Y}} = 0$.

2. Solid case

With (u,v) the vector of displacements and σ the stress tensor, the elasticity equation can be written

$$-\rho\omega^2 \begin{pmatrix} u \\ v \end{pmatrix} = \operatorname{div} \sigma, \tag{A5}$$

where

$$\sigma = \begin{pmatrix} s & t \\ t & r \end{pmatrix}, \quad \text{with} \quad \begin{cases} s = \lambda \dfrac{\partial}{\partial y} v + (\lambda + 2\mu)\dfrac{\partial}{\partial x} u, \\ t = \mu\left(\dfrac{\partial}{\partial y} u + \dfrac{\partial}{\partial x} v\right), \\ r = (\lambda + 2\mu)\dfrac{\partial}{\partial y} v + \lambda \dfrac{\partial}{\partial x} u, \end{cases} \tag{A6}$$

where ρ denotes the density, (λ,μ) the Lamé's constants. Boundary condition of a wall free of traction $\sigma\cdot\mathbf{n} = 0$ ($\mathbf{n}$ denotes the normal to the wall) in an x-axis waveguide of height $h(x)$ can be written

$$\begin{aligned} &-h's + t = 0, \quad -h't + r = 0 \quad \text{at } y = h(x), \\ &t = 0, \quad r = 0 \quad \text{at } y = 0. \end{aligned} \tag{A7}$$

With $\mathbf{X} = \binom{u}{t}$ and $\mathbf{Y} = \binom{-s}{v}$, (A5)–(A6) take the form

$$\frac{\partial}{\partial x}\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} 0 & -\frac{\beta}{\lambda} & -\beta\frac{\partial}{\partial y} \\ & \beta\frac{\partial}{\partial y} & -\left(\rho\omega^2 + \alpha\frac{\partial^2}{\partial y^2}\right) \\ \rho\omega^2 & \frac{\partial}{\partial y} & \\ -\frac{\partial}{\partial y} & \frac{1}{\mu} & 0 \end{pmatrix} \times \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}, \tag{A8}$$

with $\alpha = 4\mu(\lambda+\mu)/(\lambda+2\mu)$ and $\beta = \lambda/(\lambda+2\mu)$. With $r = \alpha(\partial/\partial y)v + \beta s$, boundary conditions (A7) can be written

$$\begin{pmatrix} 0 & 1 \\ 0 & -h' \end{pmatrix}\mathbf{X} + \begin{pmatrix} h' & 0 \\ -\beta & \alpha\frac{\partial}{\partial y} \end{pmatrix}\mathbf{Y} = 0 \quad \text{at } y = h(x), \tag{A9}$$

and the same with $h' = 0$ at $y = 0$.

We thus identify

$$\mathsf{F} = \begin{pmatrix} -\frac{\beta}{\lambda} & -\beta\frac{\partial}{\partial y} \\ \beta\frac{\partial}{\partial y} & -\left(\rho\omega^2 + \alpha\frac{\partial^2}{\partial y^2}\right) \end{pmatrix},$$

$$\mathsf{G} = \begin{pmatrix} \rho\omega^2 & \frac{\partial}{\partial y} \\ -\frac{\partial}{\partial y} & \frac{1}{\mu} \end{pmatrix}, \quad \mathsf{C}^0_{\mathsf{X}} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

$$\mathsf{C}^1_{\mathsf{X}} = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}, \quad \mathsf{C}^0_{\mathsf{Y}} = \begin{pmatrix} 0 & 0 \\ -\beta & \alpha\frac{\partial}{\partial y} \end{pmatrix}, \quad \text{and}$$

$$\mathsf{C}^1_{\mathsf{Y}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

APPENDIX B: PROPERTIES OF F AND G

Property 1 is obviously verified in both fluid and solid cases, the expressions of F and G being given in Appendix A.

To obtain property 2, we calculate the products $(\mathsf{F}\mathbf{Y}|\tilde{\mathbf{Y}})$ and $(\mathsf{G}\mathbf{X}|\tilde{\mathbf{X}})$, where $(\mathbf{X},\mathbf{Y})$ and $(\tilde{\mathbf{X}},\tilde{\mathbf{Y}})$ verify relations (A8)–(A9). We have to distinguish here the fluid and solid cases, as detailed below.

1. Fluid case

It is obvious to see from the expressions of F and G given in Appendix A that

$$(\mathsf{F}\mathbf{Y}|\tilde{\mathbf{Y}}) = (\mathbf{Y}|\mathsf{F}\tilde{\mathbf{Y}}) + h'(\mathbf{Y}\tilde{\mathbf{X}} - \mathbf{X}\tilde{\mathbf{Y}})(h),$$

$$(\mathsf{G}\mathbf{X}|\tilde{\mathbf{X}}) = (\mathbf{X}|\mathsf{G}\tilde{\mathbf{X}}),$$

obtained simply by integration by parts and by using (A4). Subtracting these two relations, Property 2 is then clearly verified.

2. Solid case

We denote here (Z_1, Z_2) the two components of a vector $\mathbf{Z}=\mathbf{X}$, $\tilde{\mathbf{X}}$, $\mathbf{Y}$, $\tilde{\mathbf{Y}}$. We can calculate by integrating by parts

$$(\mathsf{F}\mathbf{Y}|\tilde{\mathbf{Y}}) = (\mathbf{Y}|\mathsf{F}\tilde{\mathbf{Y}}) + \left[Y_2\left(-\beta\tilde{Y}_1 + \alpha\frac{\partial}{\partial y}\tilde{Y}_2 \right) - \tilde{Y}_2\left(-\beta Y_1 + \alpha\frac{\partial}{\partial y} Y_2 \right) \right]_0^h,$$

$$(\mathsf{G}\mathbf{X}|\tilde{\mathbf{X}}) = (\mathbf{X}|\mathsf{G}\tilde{\mathbf{X}}) + [X_2\tilde{X}_1 - X_1\tilde{X}_2]_0^h,$$

which are valid for any vectors $\mathbf{X}$, $\tilde{\mathbf{X}}$, $\mathbf{Y}$, $\tilde{\mathbf{Y}}$. Using the boundary conditions (A9), we finally write these relations

$$(\mathsf{F}\mathbf{Y}|\tilde{\mathbf{Y}}) = (\mathbf{Y}|\mathsf{F}\tilde{\mathbf{Y}}) + h'(Y_2\tilde{X}_2 - X_2\tilde{Y}_2)(h),$$

$$(\mathsf{G}\mathbf{X}|\tilde{\mathbf{X}}) = (\mathbf{X}|\mathsf{G}\tilde{\mathbf{X}}) - h'(Y_1\tilde{X}_1 - X_1\tilde{Y}_1)(h).$$

Consequently, we have

$$(\mathsf{F}\mathbf{Y}|\tilde{\mathbf{Y}}) + (\mathbf{X}|\mathsf{G}\tilde{\mathbf{X}}) = (\mathbf{Y}|\mathsf{F}\tilde{\mathbf{Y}}) + (\mathsf{G}\mathbf{X}|\tilde{\mathbf{X}}) + h'(\mathbf{Y}\cdot\tilde{\mathbf{X}} - \mathbf{X}\cdot\tilde{\mathbf{Y}})(h), \tag{B1}$$

which constitutes property 2.

APPENDIX C: TRANSVERSE MODES AND STRUCTURE OF THE M MATRIX

1. Fluid case

The transverse modes, solutions of (3.3) with boundary conditions (3.4) are (for $|n|\geqslant 1$)

$$\mathbf{Y}_n(y) = \sqrt{\frac{2-\delta_{n1}}{h}}\cos\left(\frac{(n-1)\pi y}{h}\right), \quad \mathbf{X}_n(y) = ik_n\mathbf{Y}_n. \tag{C1}$$

Thus, we calculate for $n\geqslant 1$

$$(\mathbf{X}_n|\mathbf{Y}_m) = (\mathbf{X}_n|\overline{\mathbf{Y}}_m) = ik_n\delta_{nm}, \tag{C2}$$

from which it is easy to deduce that $\mathsf{M}=\mathsf{I}$, where M is defined in (3.8).

2. Solid case

For the sake of clarity, only symmetric Lamb modes are presented (for antisymmetric modes, see for instance Ref. 16). Then, $\mathbf{X}_n=\binom{T_n}{U_n}$ and $\mathbf{Y}_n=\binom{V_n}{-S_n}$, where

$$U_n = ik_n\phi_n + \frac{\partial}{\partial y}\psi_n,$$

$$V_n = \frac{\partial}{\partial y}\phi_n - ik_n\psi_n,$$

$$S_n = \mu\left[-(k_n^2 + 2\beta_n^2 - \alpha_n^2)\phi_n + 2ik_n\frac{\partial}{\partial y}\psi_n \right], \tag{C3}$$

$$T_n = \mu\left[2ik_n\frac{\partial}{\partial y}\phi_n + (k_n^2 + \alpha_n^2)\psi_n \right],$$

defined with the scalar potential $\phi_n=(k_n^2+\alpha_n^2)\cosh(\beta_n y)\sinh(\alpha_n h)$ and the potential vector $(0,0,\psi_n)$, with $\psi_n(x,y) = -2ik_n\beta_n\sinh(\alpha_n y)\sinh(\beta_n h)$ and with $\alpha_n=(k_n^2-k_t^2)^{1/2}$, $\beta_n=(k_n^2-k_l^2)^{1/2}$, $k_t=\omega/c_t=(\rho/\mu)^{1/2}\omega$, and $k_l=\omega/c_l=(\rho/(\lambda+2\mu))^{1/2}\omega$.

The structure of the spectrum in the solid case is quite more complicated than in the fluid case. The dispersion relation for symmetric modes is of the form[16]

$$\frac{(\alpha_n^2+k_n^2)^2}{\alpha_n}\sinh(\alpha_n h)\cosh(\beta_n h) - 4k_n^2\beta_n\sinh(\beta_n h)\cosh(\alpha_n h) = 0. \tag{C4}$$

We refer to a previous paper where $\mathcal{J}_n$ has been explicitly calculated[11] and we summarize below the main results we can extract from its calculation (results are considered for right-going waves, associated with wave number k_n, $\mathrm{Imag}(k_n)\geqslant 0$ in convention $e^{-i\omega\tau}$).

(i) Propagating modes correspond to real wave number k_n; in this case, $\mathcal{J}_n$ is purely imaginary.

(ii) Evanescent modes that correspond to complex wave number k_n with a nonzero real part give a imaginary $\mathcal{J}_n$ with nonzero imaginary part. Such a mode with k_n wave number can be associated with another one such that $k_m=-\overline{k_n}$. In this case, we have $\mathcal{J}_m=\overline{\mathcal{J}_n}$ and we choose a numbering such that $m=n+1$.

(iii) Finally, evanescent modes associated with purely imaginary wave number give a real value for $\mathcal{J}_n$.

The relation (for right-going modes) $(\mathbf{X}_n|\mathbf{Y}_m)=\mathcal{J}_n\delta_{mn}$ is used to calculate $(\mathbf{X}_n|\mathbf{Y}_m)$. Indeed, with $\mathbf{Y}_m=\mathbf{Y}(k_m)$, the latter product is nonzero only when $k_m=\pm k_n$. Thus, for real wave number (propagating modes), it is nonzero for $n=m$. For complex k_n with nonzero real part ($k_{n+1}=-\overline{k_n}$), it is nonzero for $m=n+1$. Symmetrically, considering k_{n+1}, it is nonzero for $m=n-1$. Finally, for purely imaginary wave number, it is nonzero again for $n=m$. The structure of the matrix $(\mathbf{X}_n|\mathbf{Y}_m)$ is thus the following:

$$\begin{pmatrix} \mathcal{J}_1 & 0 & 0 & 0 \\ 0 & 0 & \mathcal{J}_2 & 0 \\ 0 & \mathcal{J}_3=\overline{\mathcal{J}_2} & 0 & 0 \\ 0 & 0 & 0 & \mathcal{J}_4 \end{pmatrix}, \tag{C5}$$

where we have considered a propagating mode with wave number k_1 ($\mathcal{J}_1$ is purely imaginary), and three evanescent modes: two are associated with k_2 and $k_3=-\overline{k_2}$ with non-

zero real part and one is associated with a purely imaginary wave number k_4 ($\mathcal{J}_4$ is real).

We can now define the M in (3.8). It contains blocks of identity matrices for real or purely imaginary wavenumbers and blocks of the form $\left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)$ for two complex conjugate wave numbers. In the previous example, we would have

$$\mathsf{M}=\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}. \tag{C6}$$

An important property of M is that $\mathsf{M}^2=I$ with $\bar{\mathsf{M}}=\mathsf{M}$.

[1] E. Gerjuoy and D. S. Saxon, ''Variational principles for the acoustic field,'' Phys. Rev. **94**, 1445–1458 (1954).

[2] T. W. Dawson, ''Scattering matrix and boundary integral equation methods for long range propagation in an acoustic waveguide with repeated boundary deformations,'' J. Acoust. Soc. Am. **90**(3), 1560–1574 (1991).

[3] H. Hudde and U. Letens, ''Scattering matrix of a discontinuity with a nonrigid wall in a lossless circular duct,'' J. Acoust. Soc. Am. **78**(5), 1826–1837 (1985).

[4] V. Varatharajulu and Y.-H. Pao, ''Scattering matrix for elastic waves. I. Theory,'' J. Acoust. Soc. Am. **60**(3), 556–566 (1976).

[5] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics (Non-relativistic Theory)* (Pergamon, Oxford, 1965).

[6] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966).

[7] P. A. Mello, P. Pereyra, and N. Kumar, ''Macroscopic approach to multichannel disordered conductors,'' Ann. Phys. **181**(2), 290–317 (1988).

[8] C. Vassalo, *Théorie des guides d'ondes Electromagnétiques* (Eyrolles, Paris, 1985), Vol. 2.

[9] R. Carminati, J. J. Sáenz, J.-J. Greffet, and M. Nieto-Vesperinas, ''Reciprocity, unitarity, and time-reversal symmetry of the S-matrix of fields containing evanescent components,'' Phys. Rev. A **62**, 0127121–0127127 (2000).

[10] A. Folguera and G. H. Harris, ''Coupled Rayleigh surface in a slowly varying elastic waveguide,'' Proc. R. Soc. London, Ser. A **455**, 917–931 (1999).

[11] V. Pagneux and A. Maurel, ''Lamb wave propagation in inhomogeneous elastic waveguides,'' Proc. R. Soc. London, Ser. A **458**, 1913–1930 (2002).

[12] P. Morse and U. Ingard, *Theoretical Acoustics* (McGraw Hill, New York, 1952).

[13] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1987).

[14] W. B. Fraser, ''Orthogonality relation for the Rayleigh–Lamb modes of vibration of a plate,'' J. Acoust. Soc. Am. **59**, 215–216 (1976).

[15] C. P. A. Wapenaar, ''Reciprocity theorems for two-way and one-way wave vectors: A comparison,'' J. Acoust. Soc. Am. **100**, 3508–3518 (1996).

[16] I. A. Viktorov, *Rayleigh and Lamb Waves: Physical Theory and Applications* (Plenum, New York, 1967), Chap. 2.

[17] P. Roux, B. Roman, and M. Fink, ''Time reversal in an ultrasonic waveguide,'' Appl. Phys. Lett. **70**, 1811–1813 (1997).

Sound attenuation in lined bends

S. Félix[a] and V. Pagneux
Laboratoire d'Acoustique de l'Université du Maine, UMR-CNRS 6613, Université du Maine, Avenue Olivier Messiaen 72085 Le Mans Cedex 09, France

(Received 23 February 2004; revised 23 June 2004; accepted 12 July 2004)

In the present paper we are concerned with sound propagation and attenuation in two- or three-dimensional lined bends. First it is shown that the effect of locally reacting absorbing materials at the walls of a waveguide can easily be taken into account in the multimodal formulation proposed in earlier papers by the authors, and, for bends, algebraic solutions are carried out for the acoustic field and scattering properties. Then a study of the sound attenuation in lined bends is given using the multimodal formulation and the properties of such waveguides are shown and discussed, in particular, the presence of a plateau of attenuation at high frequencies and a whispering gallery effect that occurs in bends. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1788733]

PACS numbers: 43.20.Mv, 43.20.Hq, 43.20.Dk [LLT] Pages: 1921–1931

I. INTRODUCTION

In recent papers by the authors[1,2] on multimodal wave propagation in curved ducts, perfectly reflecting rigid walls and a lossless linear medium were assumed.

Our first aim in the present study is to introduce the effects of a lining of the walls in the multimodal formulation of the wave propagation in bends. One can include such mechanisms and extend the field of applications of the multimodal approach to more realistic problems, without increasing the complexity of the formulation. It is of particular interest, for both scientific and practical reasons, to take into account the joint effects of liners and curvature in ducts, since these properties are often present in practical duct systems, notably in engine noise suppressors.

In Sec. II, the multimodal equations governing the components of the pressure and velocity projected on the local transverse modes in a bend are modified to take into account localized impedance conditions at the walls, and algebraic solutions for the acoustic field and reflection and transmission matrices are given. Thus, this extended formulation allows us to propose a theoretical method to predict the sound attenuation in lined bends.

Two formulations of the sound attenuation are given in Sec. III, one of which is defined for a known incident wave and takes into account the mode coupling and the scattering at each end of the duct, while the other, deduced from the conductance, characterizes the attenuation in a more general scope, for an arbitrary incident wave field. In Sec. IV the calculation with the two expressions of the sound attenuation is done for bends with various configurations of the wall treatment, and the results, as well as the results and conclusions of the preceeding works on the subject,[3–8] are discussed.

II. FORMULATION

In the present section the formulation of the multimodal wave propagation is developed for a bend lined with an absorbing material. An adiabatic lossless linear medium is assumed. Both two-dimensional and three-dimensional bends are considered.

A. Two-dimensional bend

Consider a circularly curved section of a two-dimensional duct system of width h (Fig. 1). If the complex acoustic pressure is written as $\hat{p}=\rho_0 c_0^2 p\exp(j\omega t)$, with corresponding velocity $\hat{\mathbf{v}}=c_0\mathbf{v}\exp(j\omega t)$, where ρ_0 is the density of air and c_0 the speed of sound, then dimensionless pressure p and velocity $\mathbf{v}$ satisfy the linearized Euler equations,

$$-jk\mathbf{v}=\boldsymbol{\nabla} p \tag{1}$$

and

$$\boldsymbol{\nabla}\cdot\mathbf{v}=-jkp, \tag{2}$$

with frequency $k=\omega/c_0$. By eliminating the radial component v_r of the velocity, these equations becomes, in (r,s) coordinates (see Appendix A),

$$-jk(1-\kappa r)v_s=\frac{\partial p}{\partial s}, \tag{3}$$

$$-jk(1-\kappa r)p=\frac{\partial v_s}{\partial s}-\frac{1}{jk}\frac{\partial}{\partial r}\left((1-\kappa r)\frac{\partial p}{\partial r}\right), \tag{4}$$

with v_s the axial velocity and $\kappa=1/R_0$ the curvature of the bend axis.

The boundary conditions at the walls are

$$\left(\frac{\partial p}{\partial r}\right)_{r=h/2}=-jk\zeta_i p, \quad \left(\frac{\partial p}{\partial r}\right)_{r=-h/2}=jk\zeta_e p, \tag{5}$$

where $\zeta_i=\rho_0 c_0/z_i$ and $\zeta_e=\rho_0 c_0/z_e$ are reduced admittances, and $z_i=z_i(\omega)$ and $z_e=z_e(\omega)$ the surface impedances of the internal and external wall, respectively.

The pressure and axial velocity in the bend are now expressed using infinite series

[a]Electronic mail: simon.felix@univ-lemans.fr

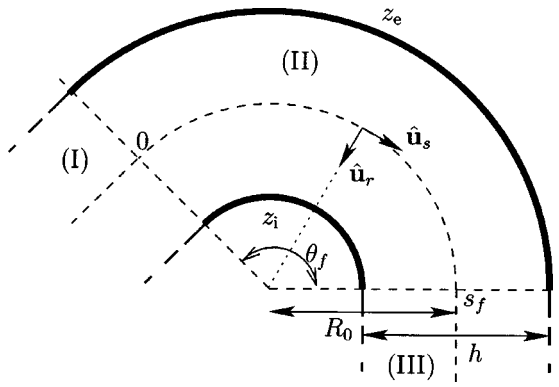


FIG. 1. Geometry of the 2-D duct system with a bend and system of coordinates. The wall treatment is characterized by localized impedances z_i and z_e.

$$p(r,s)=\sum_{\alpha} P_{\alpha}(s)\psi_{\alpha}(r), \tag{6}$$

$$v_s(r,s)=\sum_{\alpha} U_{\alpha}(s)\psi_{\alpha}(r), \tag{7}$$

where

$$\psi_{\alpha}(r)=\sqrt{2-\delta_{0\alpha}}\cos\left(\frac{\alpha\pi}{h}\left(r-\frac{h}{2}\right)\right), \quad \alpha\in\mathbb{N}, \tag{8}$$

are the *rigid* transverse eigenfunctions satisfying the homogeneous Neumann condition at the walls. *Rigid* eigenmodes are chosen instead of the eigenmodes that satisfy the boundary condition (5) for the simplicity of their calculation and use, notably when joining the solutions in the lined duct and in the rigid ducts upstream and downstream. Moreover, the *rigid* eigenmodes should not be calculated for each frequency. These eigenfunctions ψ_{α} and the pressure p do not satisfy the same boundary conditions; one can thus expect a lower convergence rate of the multimodal method, compared with the *rigid* case.[1]

The projection of Eq. (3) on the basis $(\psi_{\alpha})_{\alpha\in\mathbb{N}}$ gives the equation

$$\mathbf{P}'=-jkB\mathbf{U}, \tag{9}$$

independent of the conditions at the walls, and therefore unchanged compared with the *rigid* case.[1] B is a matrix depending on the geometrical parameters (see Appendix B). The projection of Eq. (4), however, gives a result modified by the new boundary conditions:

$$\forall(\alpha,\beta)\in\mathbb{N}^2,$$

$$U'_{\alpha}=\frac{1}{jk}\sum_{\beta}(C+KB)_{\alpha\beta}P_{\beta}+\frac{1}{jk}\frac{1}{h}\Big[(1-\kappa r)\frac{\partial p}{\partial r}\psi_{\alpha}\Big]_{-h/2}^{h/2}. \tag{10}$$

Hence, considering Eq. (5), this equation becomes

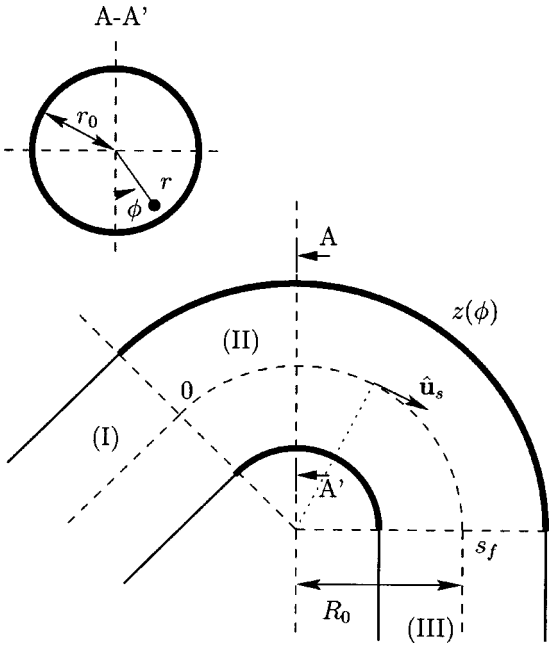


FIG. 2. Geometry of the 3-D duct system with a bend and system of coordinates. The wall treatment is characterized by the localized impedance $z(\phi)$.

$$U'_{\alpha}=\frac{1}{jk}(C+KB)_{\alpha\beta}P_{\beta}+\frac{1}{jk}\frac{1}{h}\left(\left(1-\frac{\kappa h}{2}\right)\times\left(-jk\zeta_i p\left(\frac{h}{2}\right)\right)\psi_{\alpha}\left(\frac{h}{2}\right)-\left(1+\frac{\kappa h}{2}\right)\times\left(jk\zeta_e p\left(-\frac{h}{2}\right)\right)\psi_{\alpha}\left(-\frac{h}{2}\right)\right), \tag{11}$$

with matrices K, B, and C given in Appendix B. Thus,

$$\mathbf{U}'=\frac{1}{jk}(C+KB+F)\mathbf{P}, \tag{12}$$

with

$$\forall(\alpha,\beta)\in\mathbb{N}^2,$$

$$F_{\alpha\beta}=-jk\frac{\sqrt{2-\delta_{0\alpha}}\sqrt{2-\delta_{0\beta}}}{h}\left(\zeta_i\left(1-\frac{\kappa h}{2}\right)+\zeta_e\left(1+\frac{\kappa h}{2}\right)(-1)^{\alpha+\beta}\right). \tag{13}$$

The elements $F_{\alpha\beta}$ being simply expressed as functions of k and $\zeta_{i,e}$, the matrix F can be calculated at each frequency without difficulty.

B. Three-dimensional bend of circular cross section

Consider now a three-dimensional bend, being part of a duct system of circular cross section of radius r_0 (Fig. 2). Equations (1) and (2) are expressed in toroidal coordinates (r,ϕ,s) and p and v_s then satisfy[2]

$$\frac{\partial p}{\partial s}=-jk(1-\kappa r\cos\phi)v_s, \tag{14}$$

$$\frac{\partial v_s}{\partial s}=\frac{1}{jk}\left[(1-\kappa r\cos\phi)k^2p+(1-\kappa r\cos\phi)\frac{\partial^2 p}{\partial r^2}\right.$$
$$+(1-2\kappa r\cos\phi)\frac{1}{r}\frac{\partial p}{\partial r}+(1-\kappa r\cos\phi)\frac{1}{r^2}\frac{\partial^2 p}{\partial\phi^2}$$
$$\left.+\frac{1}{r}\kappa\sin\phi\frac{\partial p}{\partial\phi}\right]. \tag{15}$$

Assuming an azimuthally varying wall treatment, given by an impedance $z=z(k,\phi)$, the boundary condition is

$$\left(\frac{\partial p}{\partial r}\right)_{r=r_0}=-jk\zeta(k,\phi)p, \tag{16}$$

where $\zeta=\rho_0 c_0/z$.

The pressure and axial velocity are projected on the orthonormal basis (ψ_α) of the transverse modes of a rigid straight duct of radius r_0. As detailed in Ref. 2, these modes are sorted in increasing order of their cutoff frequency, such that α is a simple integer index, referring to the triplet (m,n,σ) of the usual circumferential, radial, and symmetry indexes. However, due to the azimuthally varying wall impedance, the symmetry invoked in Ref. 2 to set apart symmetrical $(\sigma=1)$ and antisymmetrical $(\sigma=0)$ modes is no longer valid. All modes must be taken into account when sorted.

Again, the equation obtained by projection of the momentum conservation equation (14) on the modes ψ_α remains unchanged:

$$\mathbf{P}'=-jkB\mathbf{U}, \tag{17}$$

while equation of mass conservation (15) gives

$$\mathbf{U}'=\frac{1}{jk}\sum_\beta (C+KB)_{\alpha\beta}\mathbf{P}_\beta+\frac{1}{jk}\frac{1}{\pi r_0^2}$$
$$\times\int_0^{2\pi}\left[r(1-\kappa r\cos\phi)\frac{\partial p}{\partial r}\psi_\alpha\right]_0^{r_0}\mathrm{d}\phi, \tag{18}$$

with K, B, and C given in Appendix C. Considering Eq. (16), this equation becomes

$$\mathbf{U}'=\frac{1}{jk}(C+KB+F)\mathbf{P}; \tag{19}$$

the matrix F is given by

$$F_{\alpha\beta}=-jkr_0\zeta\frac{A_\alpha A_\beta}{\pi r_0^2}J_m(\gamma_{mn})J_\mu(\gamma_{\mu\nu})$$
$$\times\int_0^{2\pi}(1-\kappa r\cos\phi)\zeta(k,\phi)$$
$$\times\sin\left(m\phi+\frac{\sigma\pi}{2}\right)\sin\left(\mu\phi+\frac{\varsigma\pi}{2}\right)\mathrm{d}\phi, \tag{20}$$

where J_m is the mth-order Bessel function of the first kind and γ_{mn} the $(n+1)$th zero of J_m'. β refers to the triplet (μ,ν,ς) and A_α is given in Appendix B. In case of a uniform wall treatment, one can take $\sigma=\varsigma$ (there is no coupling between the symetrical and antisymetrical modes, which can therefore be considered separately), and F becomes

$$F_{\alpha\beta}=-jkr_0\zeta\frac{A_\alpha A_\beta}{\pi r_0^2}J_m(\gamma_{mn})J_\mu(\gamma_{\mu\nu})(\mathcal{I}_{m\mu}-\kappa r_0\mathcal{J}_{m\mu}), \tag{21}$$

where $\mathcal{I}_{m\mu}$ and $\mathcal{J}_{m\mu}$ are given in Appendix B. As in the two-dimensional case, the elements of F are simply expressed as functions of k and ζ.

C. Acoustic field, reflection and transmission matrices

The two infinite first-order differential equations (9) [resp., (17)] and (12) [resp., (19)] in a two-dimensional (resp., three-dimensional) bend have been constructed, assuming a duct of constant curvature and cross section and a wall lining invariant with s. Under these assumptions the coefficients in the differential equations are invariant along the axis of the bend, and algebraic solutions of these equations can be carried out to calculate the acoustic field in the bend and its reflection and transmission matrices.

Indeed, $\mathbf{P}$ obeys the equation $\mathbf{P}''+B(C+KB+F)\mathbf{P}=0$, of which a solution can be constructed in terms of the eigenvalues $\nu_\alpha^2(\alpha\in\mathbb{N})$ and eigenvectors $\mathbf{x}_\alpha$ of the matrix $B(C+KB+F)$:

$$\mathbf{P}=XD(s)\mathbf{C}_1+XD^{-1}(s)\mathbf{C}_2, \tag{22}$$

where $X=[\mathbf{x}_0,\mathbf{x}_1,\ldots]$, $D(s)$ is diagonal and given by $D_\alpha(s)=\exp(-j\nu_\alpha s)$. $\mathbf{C}_1$ and $\mathbf{C}_2$ are constant column vectors, functions of the conditions at each end of the bend.

Consider now a bend with a given radiation condition $\mathbf{U}(s_f)=Y(s_f)\mathbf{P}(s_f)$ at the outlet (the relation $\mathbf{U}=Y\mathbf{P}$ defines a generalized admittance Y within the context of the multimodal formalism) and a given pressure $\mathbf{P}(0)$ at the inlet. The continuity conditions at these two interfaces give the following system of four equations:

$$\mathbf{P}(0)=X\mathbf{C}_1+X\mathbf{C}_2, \tag{23}$$

$$Y(0)\mathbf{P}(0)=-HY(\mathbf{C}_1-\mathbf{C}_2), \tag{24}$$

$$\mathbf{P}(s_f)=X(D\mathbf{C}_1+D^{-1}\mathbf{C}_2), \tag{25}$$

$$Y(s_f)\mathbf{P}(s_f)=-HY(D\mathbf{C}_1-D^{-1}\mathbf{C}_2), \tag{26}$$

whose unknowns are the constant vectors $\mathbf{C}_1$ and $\mathbf{C}_2$, the admittance $Y(0)$ at the inlet of the bend, and the pressure $\mathbf{P}(s_f)$ at the outlet. $D=D(s_f)$, H and Y are functions of B, C, F, K, and X—the reader may refer to Ref. 1 for details.

Finding the solutions $Y(0)$ and $\mathbf{P}(s_f)$ in case where $Y(s_f)$ is the characteristic admittance Y_c, which is diagonal and given by $Y_{c\alpha}=k_\alpha/k$, where $k_\alpha^2=k^2-(\alpha\pi/h)^2$ (2D) or $k_\alpha^2=k^2-(\gamma_{mn}/r_0)^2$ (3D), allows us to define and calculate the reflection and transmission matrices of the bend. The algebraic calculation of these matrices have been developed in Ref. 1 for a rigid bend; with the appropriate matrices D, H, X, and Y taking into account the wall treatment as above, the same process leads to the reflection and transmission matrices of a lined bend.

With the solutions $\mathbf{C}_1$ and $\mathbf{C}_2$ one determines the pressure (22) in the bend. However, as dicussed in Ref. 1, the

matrix $D^{-1}(s)$ in (22) is a source of numerical problems of convergence. One thus defines $\tilde{\mathbf{C}}_2 = D^{-1}\mathbf{C}_2$ and the pressure in the bend is then written as

$$\mathbf{P} = X(D(s)\mathbf{C}_1 + D(s_f - s)\tilde{\mathbf{C}}_2), \tag{27}$$

and this formulation depends on D only, with positive arguments s and $s_f - s$, and does not depend on D^{-1}. The solutions $\mathbf{C}_1$ and $\tilde{\mathbf{C}}_2$ are

$$\mathbf{C}_1 = (1-\delta)^{-1}X^{-1}\mathbf{P}_0, \tag{28}$$

$$\tilde{\mathbf{C}}_2 = -(Y(s_f)X - HY)^{-1}(Y(s_f)X + HY) \times D(1-\delta)^{-1}X^{-1}\mathbf{P}_0, \tag{29}$$

where $\delta = D(Y(s_f)X - HY)^{-1}(Y(s_f)X + HY)D$.

This algebraic formulation, established for a duct lined with a longitudinally constant wall impedance, is also suitable for a piecewise constant impedance.[9]

When the coefficients in the matricial differential equations governing $\mathbf{P}$ and $\mathbf{U}$ are not invariant along the bend, i.e., if the curvature, cross section, or wall admittance vary, algebraic solutions cannot be carried out for the acoustic field or the reflection and transmission matrices. However, as developed in Refs. 1–2 and 10–11, the acoustic field and reflection matrix can be obtained numerically by first calculating the impedance matrix along the bend. The impedance obeys a Riccati equation, which can be integrated numerically after truncation at a sufficient number of modes (see Appendix D).

III. SOUND ATTENUATION IN A LINED BEND

A. Energy flux in the bend

The attenuation of an acoustic wave propagating in a waveguide can be defined as the decrease of the energy flux:

$$W(s) = \frac{1}{S}\int_S \frac{1}{2}\mathrm{Re}(p\mathbf{v}^*)\mathrm{d}\vec{S} = \frac{1}{2}\mathrm{Re}({}^t\mathbf{P}\mathbf{U}^*) \tag{30}$$

between the entrance and the exit of the waveguide. S is the cross-sectional area.

Since a lossless medium is assumed, only the wall treatment may cause an attenuation of the wave. Indeed, we verify that the multimodal equations lead to a constant energy flux $W(s)$ in a rigid bend, when the wall admittance is equal to zero: the variations of $W(s)$ are then given by

$$\frac{\mathrm{d}W(s)}{\mathrm{d}s} = W'(s) = \frac{1}{2}\mathrm{Re}({}^t\mathbf{P}'\mathbf{U}^* + {}^t\mathbf{P}\mathbf{U}'^*), \tag{31}$$

with $\mathbf{P}$ and $\mathbf{U}$ satisfying[1,2] $\mathbf{P}' = -jkB\mathbf{U}$ and $\mathbf{U}' = (1/jk)(C + KB)\mathbf{P}$. Thus,

$$W'(s) = \frac{1}{2}\mathrm{Re}\left(-jk\,{}^t\mathbf{U}B\mathbf{U}^* - \frac{1}{jk}\,{}^t\mathbf{P}(C+KB)\mathbf{P}^*\right). \tag{32}$$

Since B and $C+KB$ are real symmetrical matrices, the terms $-jk\,{}^t\mathbf{U}B\mathbf{U}^*$ and $(-1/jk)\,{}^t\mathbf{P}(C+KB)\mathbf{P}^*$ are purely imaginary numbers and $W'(s) = 0$: the energy flux is constant. However, if the wall admittance is finite—not equal to zero—the previous calculus leads to

$$W'(s) = \frac{1}{2}\mathrm{Re}\left(-\frac{1}{jk}\,{}^t\mathbf{P}F\mathbf{P}^*\right); \tag{33}$$

F being a complex symmetrical matrix, $W'(s)$ is not equal to zero, and consequently the energy flux may vary along the lined bend.

B. Attenuation of a given incident wave $\mathbf{P}^{(i)}$

The sound attenuation is given in decibels by

$$A_{\mathrm{dB}} = -10\log_{10}\left(\frac{W_{\mathrm{trans.}}}{W_{\mathrm{inc.}}}\right), \tag{34}$$

where $W_{\mathrm{inc.}}$ and $W_{\mathrm{trans.}}$ are the incident and transmitted energy fluxes, respectively.

Considering an incident wave $\mathbf{P}^{(i)}$, emitted by a source upstream from the bend [region (I) in Figs. 1 and 2], with corresponding incident velocity $\mathbf{U}^{(i)}$, the incident energy flux is $W_{\mathrm{inc.}} = 1/2\,\mathrm{Re}({}^t\mathbf{P}^{(i)}\mathbf{U}^{(i)*})$. Since for the incident wave $\mathbf{U}^{(i)} = Y_c\mathbf{P}^{(i)}$, this equation becomes

$$W_{\mathrm{inc.}} = \frac{1}{2}\mathrm{Re}(\mathbf{P}^{(i)\dagger}Y_c\mathbf{P}^{(i)}), \tag{35}$$

where † denotes the adjoint operator.

If T is the transmission matrix of the lined bend, the transmitted wave is $\mathbf{P}^{(t)} = T\mathbf{P}^{(i)}$, with corresponding velocity $\mathbf{U}^{(t)} = Y_c T\mathbf{P}^{(i)}$. The transmitted energy flux is then

$$W_{\mathrm{trans.}} = \frac{1}{2}\mathrm{Re}(\mathbf{P}^{(i)\dagger}T^\dagger Y_c T\mathbf{P}^{(i)}). \tag{36}$$

Then, as the transmission matrix T is known (see Sec. II C), the attenuation (34) can be calculated for any given incident wave $\mathbf{P}^{(i)}$ by using Eqs. (35) and (36):

$$A_{\mathrm{dB}} = -10\log_{10}\left(\frac{\mathrm{Re}(\mathbf{P}^{(i)\dagger}T^\dagger Y_c T\mathbf{P}^{(i)})}{\mathrm{Re}(\mathbf{P}^{(i)\dagger}Y_c\mathbf{P}^{(i)})}\right). \tag{37}$$

C. The case of an incoherent source

The result given above allows us to calculate the attenuation of a given incident wave in a lined bend. Thus some particular cases for which the incident wave is known can be studied, with the mode coupling taken into account. However, in many cases the sound source upstream from the lined duct system is not known—it may be a number of noise sources—and such an approach is then unadapted since the possibly statistical aspect of the distribution of incident modes is not taken into account. It could thus be useful to define a quantity characterizing the attenuation in a more general scope, for an arbitrary source.

For a given frequency k, we assume that the N_p propagative modes ψ_α upstream from the bend are excited by as many incoherent unit fluxes. The transmission of such a mode distribution in a waveguide can be characterized with the conductance[12]

$$G = \mathrm{Tr}(\hat{T}^\dagger\hat{Y}_c\hat{T}), \tag{38}$$

where Tr is the trace and $\hat{T}$ and $\hat{Y}_c$ are the transmission matrix and characteristic admittance matrix, respectively, restricted to the N_p propagative modes.

Indeed, the transmitted energy flux is $W_{\text{trans.}} = 1/2\,\text{Re}(\mathbf{P}^{(i)\dagger} T^\dagger Y_c T \mathbf{P}^{(i)})$, that is

$$W_{\text{trans.}} = \frac{1}{2}\text{Re}\left(\sum_{\alpha,\beta} (T^\dagger Y_c T)_{\alpha\beta} P_\alpha^{(i)*} P_\beta^{(i)} \right). \tag{39}$$

Assuming N_p incoherent unit fluxes implies $\langle P_\alpha^{(i)*} P_\beta^{(i)} \rangle = \delta_{\alpha\beta}$ if $\alpha < N_p$ and $\beta < N_p$, and $\langle P_\alpha^{(i)*} P_\beta^{(i)} \rangle = 0$ if $\alpha \geqslant N_p$ or $\beta \geqslant N_p$. Hence

$$W_{\text{trans.}} = \frac{1}{2}\text{Re}\left(\sum_{\alpha=0}^{N_p-1} (T^\dagger Y_c T)_{\alpha\alpha} \right), \tag{40}$$

that is

$$W_{\text{trans.}} = \frac{1}{2}\text{Tr}(\hat{T}^\dagger \hat{Y}_c \hat{T}). \tag{41}$$

The attenuation is then

$$A_{\text{dB}} = -10\log_{10}\left(\frac{W_{\text{trans.}}}{W_{\text{inc.}}} \right) = -10\log_{10}\left(\frac{G}{\text{Tr}(\hat{Y}_c)} \right). \tag{42}$$

IV. RESULTS

Expressions (37) and (42) are valid for both two-dimensional and three-dimensional ducts, as long as the transmission matrix of the lined duct can be calculated. However, for simplicity, a two-dimensional bend is considered in all of the results to be described in the following, except where otherwise stated.

For comparison, results on the attenuation in straight lined ducts are given in the present section, together with the results for bends. It can be easily shown that the algebraic formulation of the reflection and transmission matrices that has been formulated for bends (see Ref. 1 and Sec. II C) can be used for straight ducts by simply taking $\kappa = 0$. Therefore, the same expressions (37) and (42) of the attenuation are used in the following for both bent and straight ducts, with the appropriate transmission matrix.

The dimensions of the bend are $h = 0.2\,\text{m}$, $R_0 = 0.9\,\text{m}$, and the overall angle $\theta_f = s_f/R_0 = 60°$ (Fig. 1). The walls are lined with a rigid porous material. Assuming a locally reacting material, the surface impedance of the wall treatment is taken as the impedance at normal incidence of a plane porous plate having the same thickness:

$$z(\omega) = -\frac{j}{\phi}(\rho K)^{1/2} \cot\left(\omega \left(\frac{\rho}{K} \right)^{1/2} d \right), \tag{43}$$

where d is the thickness of the porous layer, ϕ the porosity, ρ the effective density, and K the effective bulk modulus of air in the material.[13] The porous material is characterized with the following quantitites: porosity $\phi = 0.98$, tortuosity $\alpha_\infty = 1.2$, flow resistivity $\sigma = 40000\ \text{N m}^{-4}\text{s}$, characteristic dimensions $\Lambda = 2 \times 10^{-4}\,\text{m}$ and $\Lambda' = 4 \times 10^{-4}\,\text{m}$, thickness $d = 0.05\,\text{m}$.

A. Incident plane wave

The following results are obtained assuming an incident plane wave, emitted by a source upstream from the bend. Thus one use the expression (37) to calculate the attenuation, with an incident mode 0, $P_\alpha^{(i)} = \delta_{\alpha 0}$.

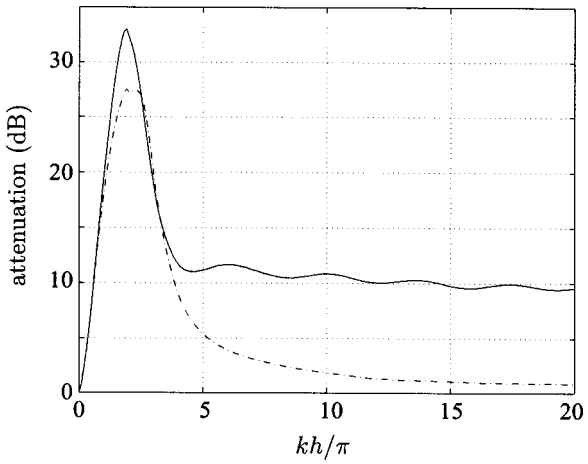


FIG. 3. Attenuation in a 2-D bend (—) and in a straight duct having the same axis length s_f (–·–), both lined with a porous material, for a plane incident wave.

1. General results

Figure 3 shows the attenuation as function of the frequency in the lined bend described above and in a straight lined duct having the same axis length and the same wall treatment. The attenuation in a straight lined duct is known to follow the typical curve shown in this figure: the attenuation is weak at low frequencies; it reaches a maximum when the wavelength is of the order of the duct width ($kh/\pi \sim 2$), and then decreases. The attenuation at high frequencies is poor, due to a "ray effect:"[3] the incident acoustic wave being regarded as rays propagates through the duct without "colliding" the walls.

A way, first pointed out by empirical means,[3] to avoid this ray effect and increase the efficiency of the system is to use a curved duct. A plateau appears then, keeping the attenuation at a high level (for $kh/\pi = 100$, the attenuation is still greater than 9 dB). Thus, although the curvature is weak ($\kappa h \sim 0.2$) in our example, it yields a substantial increase in the attenuation at high frequencies.

At low frequencies, the effects of the curvature vanish and the attenuation curves for the bend and the straight duct converge to tend toward zero. Rostafinski[7] concludes his study of lined bends at very low frequencies (the wavelength is at least two orders of magnitude larger than the width h of the bend, that is, $kh/\pi < 10^{-2}$) by stating that the attenuation in a bend is generally lower (by 2% to 7%) than in a straight duct. None of the configurations that have been considered in the present study have led to this conclusions; in all cases, the difference between the results for a bend and a straight duct in this range of frequency was much lower than 1%.

More generally, for all frequencies, the attenuation level is globally higher when the duct is curved than when it is straight. However it can be locally, that is, in a small frequency range (see, e.g., for $kh/\pi \sim 3$), lower. If the behavior at low and high frequencies can be easily interpreted, it is not simple to describe the effect of the curvature at medium frequencies, and therefore to explain this latter behavior.

When considering a three-dimensional bend and straight duct with a circular cross section (Fig. 4), similar results as discussed above are obtained.

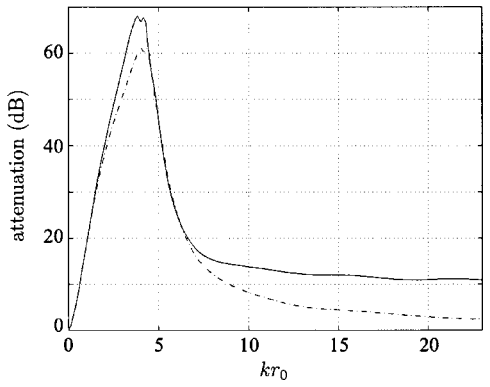


FIG. 4. Attenuation in a 3-D bend of circular cross section (—, $r_0 = 0.1$ m, $R_0 = 0.9$ m, $s_f/R_0 = 60°$), and in straight duct having the same axis length s_f (–·–), both lined with a porous material, for a plane incident wave.

Figure 5 shows the attenuation in a two-dimensional bend and a straight duct with walls that are lined with a fibrous sheet mounted on a locally reacting core with impervious backing, as studied by Ko and Ho[5] and Lafarge.[8] The reduced admittance of the lining is given by $\zeta(k) = [R(1 + jk/k_0) - j\cot(kd)]^{-1}$, where $R = 1.5$ is the reduced resistance, $k_0 = 70.12\,\mathrm{m}^{-1}$ the characteristic wave number, and $d = 0.1$ m the lining core depth. The attenuation curves describe a succession of absorption lobes and vanish each time the transverse velocity is null on the fibrous sheet, that is, since the thickness of the cores is $d = h/2$, for frequencies $kh/\pi = 4n$, $n \in \mathbb{N}$. For this configuration one can also consider that a plateau appears at high frequencies for the bend, since the decrease in the amplitude of the absorption lobes observed for the straight duct is no longer observed for the bend.

This particular behavior of the attenuation in bends at high frequencies—the appearence of a plateau—is thus not due to some properties of one particular kind of lining and any absorbing boundary condition would produce the same effect. As a matter of fact, Fig. 6 shows that the plateau is observed when considering an arbitrary constant admittance.

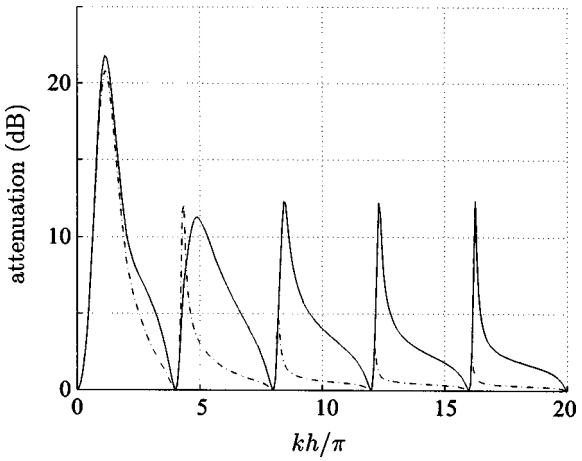


FIG. 5. Attenuation in a 2-D bend (—) and in a straight duct having the same axis length s_f (–·–), for a plane incident wave. The walls are lined with a fibrous sheet mounted on a locally reacting core.

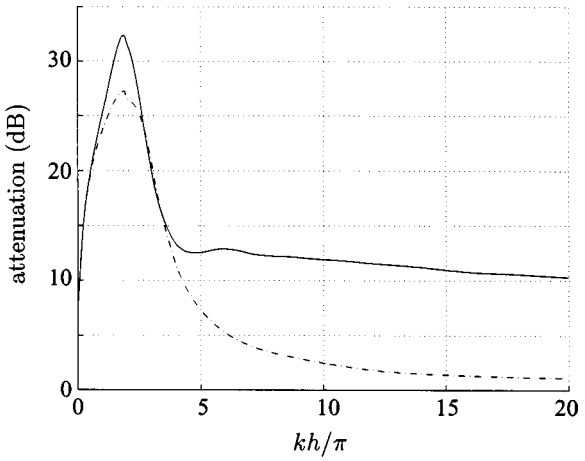


FIG. 6. Attenuation in a 2-D bend (—) and in a straight duct having the same axis length s_f (–·–), for a plane incident wave. The reduced wall admittance is constant: $\zeta_i = \zeta_e = 0.5 + j0.25$.

2. One-wall-lined bends

Figure 7 shows attenuation curves of one-wall-lined bends and a straight duct. The attenuation in the outer-wall-lined bend is much greater than in the inner-wall-lined duct. Furthermore, no plateau appears if the inner wall only is lined, and then attenuation in this case is lower than in a one-wall-lined straight duct. Thus, as remarked by Grigor'yan,[3] the curvature does not always yield an increase in the attenuation. Ko and Ho[5] suggest that the difference between the attenuations for inner-wall-lined and outer-wall lined bends may be due to the difference between the lengths of the walls in a bend. There is no doubt that an increase in the length of wall treatment generally increases the attenuation in a duct, but the difference between the length of the walls in a bend cannot explain on its own this significant difference. The attenuations for two bends with the same parameters h and R_0, one of which is inner-wall-lined while the other is outer-wall-lined, with axis lengths such that the length of the wall treatment is the same in both cases, are compared in Fig. 8. The results are clearly different, although the length that is lined is the same.

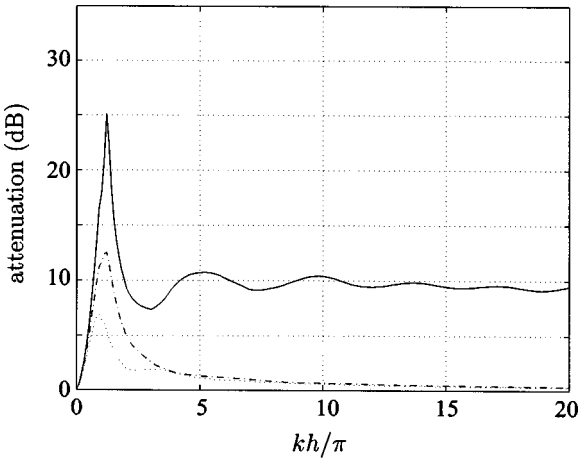


FIG. 7. Attenuation in one-wall-lined bends, for a plane incident wave; (—) outer wall lined, (···) inner wall lined, (–·–) straight duct with only one wall lined.

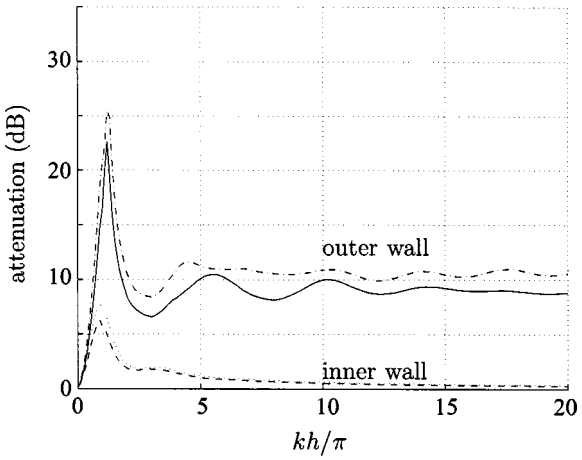


FIG. 8. Attenuation in one-wall-lined 2-D bends with various overall angles, for a plane incident wave; (—) outer wall lined, $\theta_f = 54°$, (– – –) inner wall lined, $\theta_f = 54°$, (–·–) outer wall lined, $\theta_f = 67.5°$, (···) inner wall lined, $\theta_f = 67.5°$. The lining length in the configurations (—) and (···) is the same.

3. Whispering gallery effect

More than a difference between the lengths of lining, the appearance of a whispering gallery effect may explain the difference of efficiency between the inner-wall-lined and outer-wall-lined bends. For sufficiently high frequencies, the incident wave "traverses the duct by glancing along the concave wall, the convex wall being left in the shadow."[3] The axial component j_s of the acoustic intensity $\mathbf{j} = 1/2\,\mathrm{Re}(p\mathbf{v}^*)$, calculated in a two-dimensional rigid bend for a plane incident wave, shows this whispering gallery effect (Fig. 9): the intensity, uniformly distributed on the cross section upstream from the bend, is progressively confined near the concave (outer) wall as the frequency is increased.

Since the intensity is at its minimum—and even locally null, see Fig. 9—near the convex (inner) wall, it is clear that the contribution of the latter to sound attenuation is poor. In contrary, lining the outer wall, where the intensity is at its maximum, may yield a substantial attenuation, even at high frequencies. Besides, the relatively uniform distribution of the intensity on the cross section of a straight duct may explain the intermediate position of the attenuation curve for a one-wall-lined straight duct in Fig. 7.

When the frequency is increased, the zone of the maximum of intensity near the outer wall becomes more narrow. Considering what is written above, a partial lining of the outer wall near this narrow zone should yield an attenuation similar to what is obtained when the two walls are wholly lined. Consider the attenuation in a two-dimensional bend with parameters $h = 0.2\,\mathrm{m}$, $R_0 = 0.9\,\mathrm{m}$, and $\theta_f = 60°$, as before, lined on 30° in the middle of its outer wall, for a plane incident wave (Fig. 10). The result at high frequencies is close to the one obtained for the wholly lined bend—the difference is of the order of 2 dB—and the attenuation remains much greater than in the wholly lined straight duct. If the lined zone is reduced to 20°, the attenuation plateau level is close to 7 dB. Besides, we verify that the position on the wall of a given length of wall treatment is not indifferent; indeed, the attenuation is at its maximum when the wall is lined where the maximum of energy flux is expected.

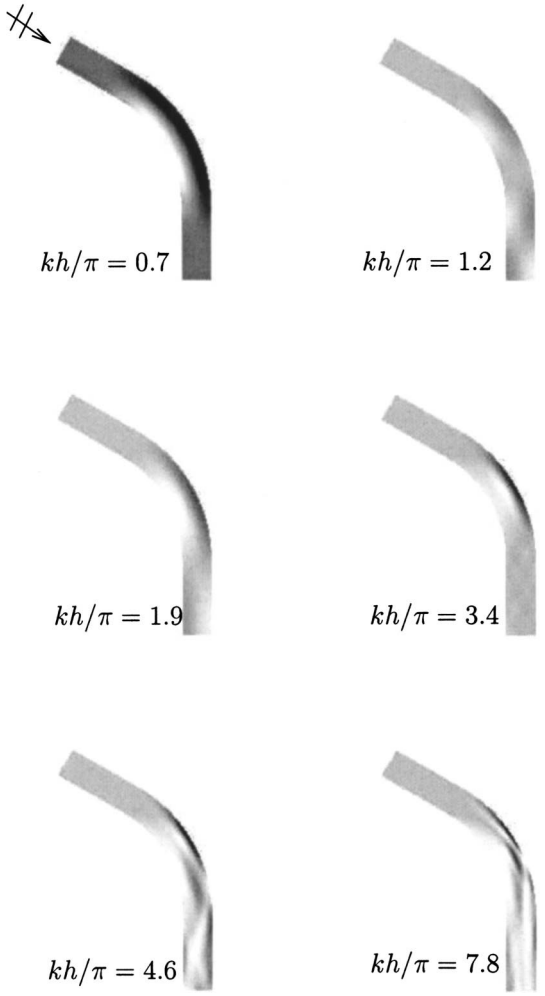


FIG. 9. Axial component j_s of the acoustic intensity $\mathbf{j} = 1/2\,\mathrm{Re}(p\mathbf{v}^*)$ in a 2-D rigid bend ($R_0/h = 4.5$, $\theta_f = 60°$), for six frequencies. A plane wave ($j_s = 1$, arbitrary unit) is emitted upstream from the bend. For all of the figures, except the first one ($kh/\pi = 0.7$), the gray scaling gives the values of j_s varying between 0 (white) and 5 (black). Since the variations of j_s in the first figure are lower than 7%, the gray scaling is different, varying between 0.93 (white) and 1.07 (black).

B. Incoherent sources

We now assume that the incident wave field upstream from the bend is produced by incoherent sources. Therefore the attenuation in the following results is calculated with the expression (42), deduced from the conductance (38).

Using Eq. (42) to calculate the attenuation in a two-dimensional bend lined on its two walls with a porous layer (Fig. 11) gives a result that is qualitatively close to the one obtained for an incident plane wave (Fig. 3), with notably the appearance of the plateau at high frequencies, after an absorption lobe at $kh/\pi \sim 2$. All of the propagative modes being taken into account and these modes being more attenuated than the plane wave mode, the level of the attenuation plateau if naturally higher.

However, in the straight duct, a significant change can be noticed. Since the plane wave mode is weakly attenuated at high frequencies in a straight duct, the curve shown in Fig. 3 is rapidly close to zero. The propagative modes being taken into account in the expression (42) of the attenuation, a plateau appears in Fig. 11, which is, however, still under the

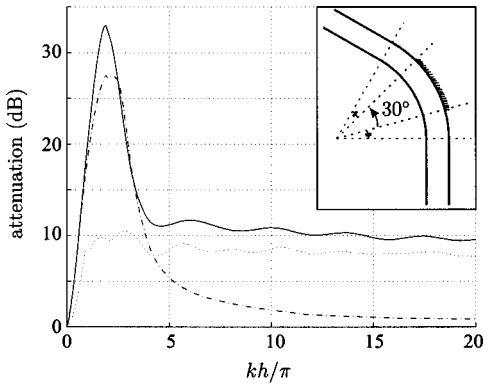


FIG. 10. Attenuation in a 2-D bend locally lined on the outer wall, for a plane incident wave. For comparison, the curves of Fig. 3 are also plotted.

plateau for the bend. The curvature still yields an increase in the attenuation.

Using Eq. (42) to calculate the attenuation in bends lined as in Fig. 5 and 6 or with other kinds of wall treatments leads to the same conclusions. The attenuation is globally increased when the duct is curved, and a plateau of attenuation appears in straight ducts.

As in the incident plane wave case, the contribution of the outer wall to the global attenuation is still more important than the inner wall contribution. However, the whispering gallery effect that has been observed in the previous paragraph is proper to the case of very low-order incident modes, and the distribution of the energy in the bend is not limited to the neighborhood of the outer wall when the incident wave field is complex. The ray method to be used in the following shows this effect of the higher-order modes on the energy distribution. One consequence of this spread distribution of the energy in the bend is the appearance of a plateau at high frequencies in inner-wall-lined bends.

A classical expression of the attenuation in waveguides and, in particular, in bend is[3,5,8]

$$A_{\mathrm{dB}} = -10\log_{10}\left(\frac{\sum_{n=0}^{N_p-1} e^{2\,\mathrm{Im}(\nu_n)\theta_f}}{N_p}\right), \tag{44}$$

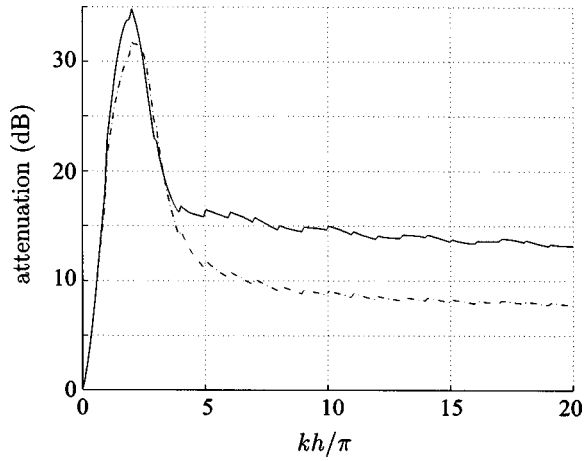


FIG. 11. Attenuation in a 2-D bend (—) and in a straight duct having the same axis length s_f (–·–), calculated with the expression (42).

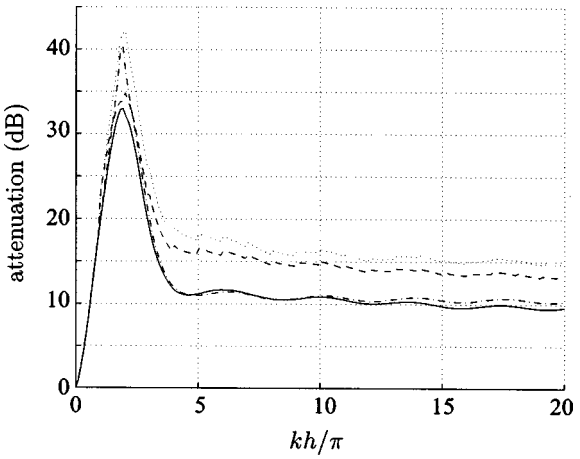


FIG. 12. Attenuation in a 2-D bend, calculated with (—) expression (37) with a plane incident wave, (– – –) expression (42), (···) expression (44), (–·–) expression (44) with $N_p=1$ (fundamental mode only).

where the angular wave numbers ν_n ($n\in\mathbb{N}$) are the solutions of the implicit dispersion relation,

$$\begin{aligned}
&\left[j\zeta_1 N_\nu\left(kR_0\left(1-\kappa\frac{h}{2}\right)\right)+N'_\nu\left(kR_0\left(1-\kappa\frac{h}{2}\right)\right)\right]\\
&\quad\times\left[j\zeta_2 J_\nu\left(kR_0\left(1+\kappa\frac{h}{2}\right)\right)+J'_\nu\left(kR_0\left(1+\kappa\frac{h}{2}\right)\right)\right]\\
&\quad-\left[j\zeta_1 J_\nu\left(kR_0\left(1-\kappa\frac{h}{2}\right)\right)+J'_\nu\left(kR_0\left(1-\kappa\frac{h}{2}\right)\right)\right]\\
&\quad\times\left[j\zeta_2 N_\nu\left(kR_0\left(1+\kappa\frac{h}{2}\right)\right)+N'_\nu\left(kR_0\left(1+\kappa\frac{h}{2}\right)\right)\right]=0.
\end{aligned} \tag{45}$$

The calculation of the angular wave numbers is often limited to the first mode,[3,7,8] since the fundamental mode in a straight or bent duct is generally the least attenuated.[3,5,8,14]

The result that is obtained with the expression (44) of the attenuation for the two-dimensional bend lined with porous material is close to the result obtained with Eq. (42) (Fig. 12). This figure also shows a comparison between the calculations of the attenuation with Eq. (44) with the first bend mode only and with Eq. (37). Again, both methods give similar results. In our example the bend is weakly curved, and therefore the energy of the incident plane wave mode is predominantly transferred to the first bend mode, which is the least attenuated. But if the bend is more strongly curved or if the wall impedance is high, Eqs. (44) and (42) no longer give similar results. In such cases the simple expression (44) of the attenuation in not a suitable measurement of the attenuation.

C. Rays to predict the attenuation

A mode ψ_α propagating in a two-dimensional straight waveguide can be regarded as two interfering sets of rays at angles $\theta_\alpha=\sin^{-1}(\alpha\pi/kh)$ and $-\theta_\alpha$. The aim of the present part is to use this interpretation in a simpler form—regarding

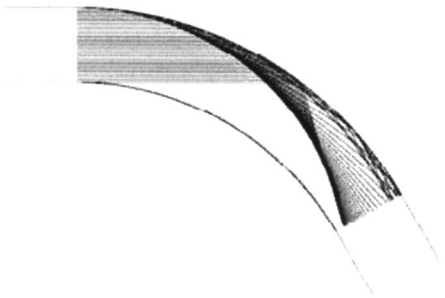


FIG. 13. Propagation in a 2-D bend of a plane incident wave using a ray method.

the mode propagating in the duct as two sets of rays with no phase information—to predict the zones of absorption of a given incident wave on the walls of a duct.

Consider first the propagation of an incident plane wave in a rigid bend joining two straight ducts (Fig. 13). As shown before with the calculations of attenuation and acoustic intensity (Sec. IV A 3), the inner wall, where none of the rays collide, is ''left in the shadow.''

Consider now the propagation of a higher order mode ($\alpha>0$) for a given frequency, as, e.g., the mode ψ_{12} with $kh/\pi=17.5$ (Fig. 14). Due to the finite angle of incidence of the rays ($\theta_{12}=43.3°$ at this frequency), the inner wall of the bend is not in the shadow anymore, even if the ''density'' of rays colliding this wall is still lower than the density on the outer wall. Figure 15 shows the mean attenuation in the bend drawn in Fig. 14, with $h=0.2$ m and the same porous lining as in the preceeding, in the frequency range $kh/\pi \in]17,18[$, for the incident mode ψ_{12}, and for five lining schemes. The attenuation in the outer-wall-lined case is greater than in the inner-wall-lined case, as expected with the ray approach.

Two zones can be distinguished on the outer wall in Fig. 14: the first 35°, characterized by a high density of rays colliding the wall, and the next 50°, where the density is much lower. Figure 15 shows that the attenuation when the first 35° of the outer wall are lined is greater that when the next 50° are lined, altough the length of lining is greater in the latter case.

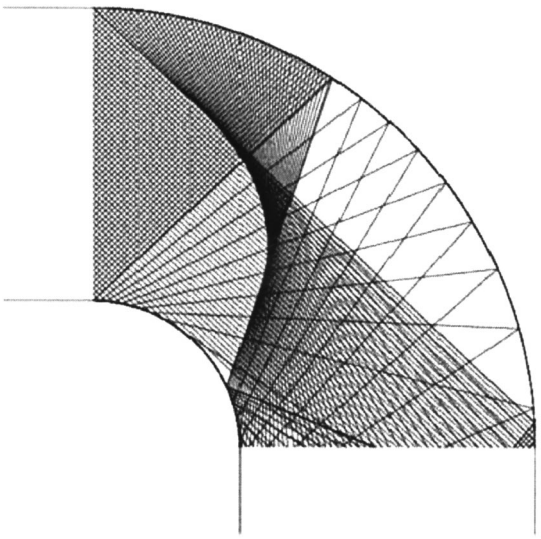


FIG. 14. Propagation in a 2-D bend of the incident mode ψ_{12} at the frequency $kh/\pi=17.5$ using a ray method ($R_0/h=1$, $\theta_f=90°$).

	attenuation (dB)		attenuation (dB)
	4.8 ± 0.1		3 ± 0.1
	4.3 ± 0.1		0.5 ± 0.05
	0.8 ± 0.1		

FIG. 15. Mean attenuation level in the frequency range $kh/\pi \in]17,18[$ for the bend shown in Fig. 14, with $h=0.2$ m, for five lining schemes (the lined zones on the walls are shown with bold strokes), for the incident mode ψ_{12}.

Thus, the simple observation of rays propagating in the duct allows us to expect the relative levels of attenuation for the different lining schemes. If one considers the practical problem of a noise source with a frequency emerging on a particular mode, a wall treatment localized on the efficient absorption zones can thus be proposed with this very simple approach.

V. CONCLUSION

The effects of a lining of the walls on wave propagation in curved ducts have been introduced in the multimodal formulation established in earlier papers by the authors. New equations giving the multimodal acoustic field and the scattering and impedance matrices have been carried out, extending the field of applications of the multimodal method to ducts lined with locally reacting absorbing materials.

Besides, two calculations have been proposed to evaluate the sound attenuation in a lined duct, for a given incident wave or a distribution of incoherent sources, and results with both calculations have been presented and discussed.

The known effects of the curvature on the sound attenuation—the increase of the global attenuation level and the appearance of a plateau of attenuation at high frequencies—have been shown systematically for various configurations of the wall treatment and incident wave field. For a two-dimensional bend, the whispering gallery effect of the fundamental mode that appears along the outer wall has been shown, which explains the great difference between sound attenuations induced by the inner and outer walls respectively. For a three-dimensional bend whose cross-section is such that one cannot consider disctinct inner and outer wall (circular or elliptical cross-section,...), one can expect that the whispering gallery effect still occurs, with the wave glancing along the external part of the wall. This effect of a larger attenuation on the external part of the duct wall is observed for any kind of source, even though the whispering gallery effect is less pronounced.

Finally, a simple study of ray propagation in a bend has been shown to allow a prediction of the zones of the wall where the sound absorption mainly occurs and further studies are planned on the characterization of waveguides with ray methods.[15]

APPENDIX A: 2D COORDINATE SYSTEM

Since (r,s) coordinates are orthogonal, we have $\mathrm{d}\mathbf{r} = h_r\,\mathrm{d}r\hat{\mathbf{u}}_r + h_s\,\mathrm{d}s\hat{\mathbf{u}}_s$, where the scale factors are $h_r=1$ and $h_s=1-\kappa r$. Thus, the gradient and divergence operators in this coordinate system are

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial r}\hat{\mathbf{u}}_r + \frac{1}{1-\kappa r}\frac{\partial f}{\partial s}\hat{\mathbf{u}}_s, \tag{A1}$$

$$\boldsymbol{\nabla}\cdot\mathbf{F} = \frac{1}{1-\kappa r}\frac{\partial}{\partial r}((1-\kappa r)F_r) + \frac{1}{1-\kappa r}\frac{\partial F_s}{\partial s}. \tag{A2}$$

APPENDIX B: MATRICES IN THE 2D MULTIMODAL EQUATIONS

The matrices K, B, and C are defined and calculated in Ref. 1. We give in the following their expression with the notations of the present paper:

$\forall(\alpha,\beta)\in\mathbb{N}^2$,

$$K_{\alpha\beta} = \left(k^2 - \left(\frac{\alpha\pi}{h}\right)^2\right)\delta_{\alpha\beta}, \tag{B1}$$

$$B_{\alpha\beta} = \begin{cases} 1 & \text{if } \alpha=\beta, \\ \sqrt{2-\delta_{0\alpha}}\sqrt{2-\delta_{0\beta}}\dfrac{\kappa h}{\pi^2}((-1)^{\alpha+\beta}-1)\dfrac{\alpha^2+\beta^2}{(\alpha^2-\beta^2)^2}, & \text{if } \alpha\neq\beta \end{cases} \tag{B2}$$

$$C_{\alpha\beta} = \begin{cases} 0 & \text{if } \alpha=\beta, \\ \sqrt{2-\delta_{0\alpha}}\sqrt{2-\delta_{0\beta}}\dfrac{\kappa}{h}((-1)^{\alpha+\beta}-1)\dfrac{\alpha^2}{\alpha^2-\beta^2}, & \text{if } \alpha\neq\beta \end{cases}. \tag{B3}$$

APPENDIX C: MATRICES IN THE 3D MULTIMODAL EQUATIONS

The integer index α refers to circumferential and radial indexes m and n, respectively, and β refers to μ and ν. For details about the following expressions, the reader may refer to Ref. 2:

$\forall(\alpha,\beta)\in\mathbb{N}^2$,

$$K_{\alpha\beta} = \left(k^2 - \frac{\gamma_{mn}^2}{r_0^2}\right)\delta_{\alpha\beta}, \tag{C1}$$

$$B_{\alpha\beta} = \delta_{\alpha\beta} - A_\alpha A_\beta \frac{\kappa}{\pi r_0^2}\mathcal{T}_{m\mu} \times \int_0^{r_0} r^2 J_m\left(\frac{\gamma_{mn}r}{r_0}\right)J_\mu\left(\frac{\gamma_{\mu\nu}r}{r_0}\right)\mathrm{d}r, \tag{C2}$$

$$C_{\alpha\beta} = A_\alpha A_\beta \frac{\gamma_{mn}\kappa}{\pi r_0^3}\mathcal{T}_{m\mu}\int_0^{r_0} r J_{m+1}\left(\frac{\gamma_{mn}r}{r_0}\right)J_\mu\left(\frac{\gamma_{\mu\nu}r}{r_0}\right)\mathrm{d}r, \tag{C3}$$

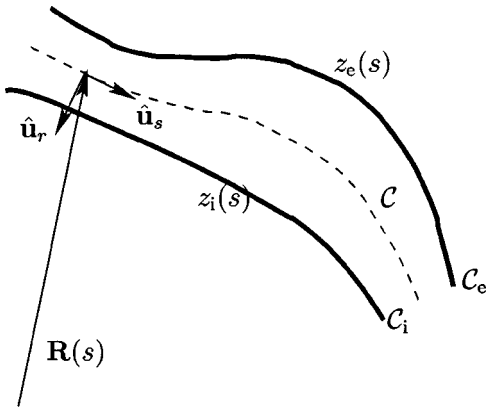


FIG. 16. Duct with varying curvature, cross section, and wall admittance.

$$-A_\alpha A_\beta m\frac{\kappa}{\pi r_0^2}(\mathcal{T}_{m\mu}-\mathcal{H}_{m\mu})\int_0^{r_0} J_m\left(\frac{\gamma_{mn}r}{r_0}\right)J_\mu\left(\frac{\gamma_{\mu\nu}r}{r_0}\right)\mathrm{d}r, \tag{C4}$$

where

$$A_\alpha = \begin{cases} 1/\sqrt{J_0^2(\gamma_{0n})}, & \text{if } m=0, \\ 1/\sqrt{\dfrac{1}{2}\left(1-\dfrac{m^2}{\gamma_{mn}^2}\right)J_0^2(\gamma_{mn})}, & \text{if } m>0; \end{cases} \tag{C5}$$

$$\mathcal{T}_{m\mu} = \int_0^{2\pi}\cos\phi\sin\left(m\phi+\frac{\sigma\pi}{2}\right)\sin\left(\mu\phi+\frac{\sigma\pi}{2}\right)\mathrm{d}\phi \tag{C6}$$

$$= \frac{\pi}{2}\delta_{|m-\mu|,1}(1+(\delta_{\sigma1}-\delta_{\sigma0})(\delta_{m0}+\delta_{\mu0})), \tag{C7}$$

and

$$\mathcal{H}_{m\mu} = \int_0^{2\pi}\sin\phi\cos\left(m\phi+\frac{\sigma\pi}{2}\right)\sin\left(\mu\phi+\frac{\sigma\pi}{2}\right)\mathrm{d}\phi \tag{C8}$$

$$= \frac{\pi}{2}(\delta_{\mu-m,1}(1-(\delta_{\sigma1}-\delta_{\sigma0})\delta_{m0}) - \delta_{m-\mu,1}(1+(\delta_{\sigma1}-\delta_{\sigma0})\delta_{\mu0})). \tag{C9}$$

The term $\mathcal{I}_{m\mu}$ in Eq. (21) is

$$\mathcal{I}_{m\mu} = \int_0^{2\pi}\sin\left(m\phi+\frac{\sigma\pi}{2}\right)\sin\left(\mu\phi+\frac{\sigma\pi}{2}\right)\mathrm{d}\phi = \pi\delta_{m\mu}(1+(\delta_{\sigma1}-\delta_{\sigma0})\delta_{m0}). \tag{C10}$$

APPENDIX D: DUCT WITH VARYING CURVATURE, CROSS SECTION AND WALL IMPEDANCE

In this appendix, the multimodal formulation is given for a duct with varying curvature, cross-section and locally reacting wall treatment. For convenience, a two-dimensional duct is considered, but this formulation can be carried out in a three-dimensional bend without difficulty.

The region in plane to be analyzed here is shown in Fig. 16, delimited by two lined duct walls, $\mathcal{C}_i$ and $\mathcal{C}_e$. Consider a curve $\mathcal{C}$ represented by a vector function $\mathbf{R}(s)$, where s is the

arclength measured from some reference point. $\hat{\mathbf{u}}_s$ and $\hat{\mathbf{u}}_r$ being the tangent and normal vectors of the curve $\mathcal{C}$, the inner wall $\mathcal{C}_i$ and outer wall $\mathcal{C}_e$ of the duct are represented by the vector functions

$$\mathbf{R}_i(s)=\mathbf{R}(s)+r_i(s)\hat{\mathbf{u}}_r, \tag{D1}$$

$$\mathbf{R}_e(s)=\mathbf{R}(s)+r_e(s)\hat{\mathbf{u}}_r. \tag{D2}$$

As in Sec. II, pressure p and axial velocity v_s in the duct are expressed using the infinite series

$$p(r,s)=\sum_\alpha P_\alpha(s)\psi_\alpha(r,s), \tag{D3}$$

$$v_s(r,s)=\frac{1}{h(s)}\sum_\alpha U_\alpha(s)\psi_\alpha(r,s), \tag{D4}$$

where $h(s)=r_i(s)-r_e(s)$. Note that a different development, more convenient in varying cross-section waveguides,[10] has been chosen for v_s compared to Eq. (7). The *rigid* transverse eigenfunctions ψ_α are

$$\psi_\alpha(r,s)=\sqrt{2-\delta_{0\alpha}}\cos\left(\alpha\pi\frac{r_i-r}{r_i-r_e}\right), \quad \alpha\in\mathbb{N}. \tag{D5}$$

Therefore, by projecting Eqs. (1) and (2), expressed in (r,s) coordinates, the following equations are obtained for $\mathbf{P}$ and $\mathbf{U}$:

$$\mathbf{P}'=-jkB\frac{1}{h}\mathbf{U}+\frac{h'}{h}{}^t\Pi\mathbf{P}, \tag{D6}$$

$$\mathbf{U}'=\frac{1}{jk}h(C+KB+F)\mathbf{P}-\frac{h'}{h}\Pi\mathbf{U}, \tag{D7}$$

where matrices B, C, K, and F are as given in previous sections and apendices, but with the width h, the curvature κ of $\mathcal{C}$, and the wall impedances z_i and z_e depending on s. The matrix Π is given by

$$\Pi_{\alpha\beta}=\begin{cases}(1-\delta_{0\alpha})/2, & \text{if } \alpha=\beta,\\ \sqrt{2-\delta_{0\alpha}}\sqrt{2-\delta_{0\beta}}\dfrac{r_i'-(-1)^{\alpha+\beta}r_e'}{r_i'-r_e'}\dfrac{\alpha^2}{\alpha^2-\beta^2}, & \text{if } \alpha\neq\beta\end{cases}. \tag{D8}$$

As Eqs. (D6) and (D7) cannot be integrated directly, the impedance matrix Z is defined, fulfilling $\mathbf{P}=Z\mathbf{U}$. Substituting this relation in Eqs. (D6) and (D7) leads to the Riccati equation

$$Z'=-jk\frac{1}{h}B+\frac{h'}{h}{}^t\Pi Z+Z\frac{h'}{h}\Pi-\frac{1}{jk}hZ\times(C+KB+F)Z, \tag{D9}$$

which is numerically workable. Hence, by first calculating the impedance in the duct, the acoustic field or reflection matrix can then be obtained.

[1] S. Félix and V. Pagneux, "Sound propagation in rigid bends: A multimodal approach," J. Acoust. Soc. Am. **110**, 1329–1337 (2001).
[2] S. Félix and V. Pagneux, "Multimodal analysis of acoustic propagation in three-dimensional bends," Wave Motion **36**, 157–168 (2002).
[3] F. E. Grigor'yan, "Soundproofing by means of ducts with curved porous walls," Akust. Zh. **16**, 229–235 (1970) [English translation: Sov. Phys. Acoust. **16**, 192–196 (1970)].
[4] M. K. Meyer and P. Mungur, "Sound propagation in curved ducts," Prog. Astronaut. Aeronaut. **44**, 347–362 (1976).
[5] S. H. Ko and L. T. Ho, "Sound attenuation in acoustically lined curved ducts in the absence of fluid flow," J. Sound Vib. **53**, 189–201 (1977).
[6] S. H. Ko, "Three-dimensionnal acoustic waves propagating in acoustically lined cylindrically curved ducts without fluid flow," J. Sound Vib. **66**, 165–179 (1979).
[7] W. Rostafinski, "Propagation of long waves in acoustically treated, curved ducts," J. Acoust. Soc. Am. **71**, 36–41 (1982).
[8] D. Lafarge, "Etude 2D de l'atténuation acoustique dans un guide semi-circulaire traité en parois par un matériau poreux," XXe Journée Technologique, C. T. T. M. Le Mans, France, 1998.
[9] W. Bi, V. Pagneux, D. Lafarge, and Y. Aurégan, "Sound propagation in nonuniform lined duct by the multimodal method," submitted to J. Sound. Vib.
[10] V. Pagneux, N. Amir, and J. Kergomard, "A study of wave propagation in varying cross-section waveguides by modal decomposition. I. Theory and validation," J. Acoust. Soc. Am. **100**, 2034–2048 (1996).
[11] N. Amir, V. Pagneux, and J. Kergomard, "A study of wave propagation in varying cross-section waveguides by modal decomposition. II. Results," J. Acoust. Soc. Am. **101**, 2504–2517 (1997).
[12] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, 1998).
[13] J.-F. Allard, *Propagation of Sound in Prous Media: Modelling Sound Absorbing Materials* (Chapman & Hall, London, 1993).
[14] P. E. Doak and P. G. Vaidya, "Attenuation of plane wave and higher order mode sound propagation in lined ducts," J. Sound Vib. **12**, 201–224 (1970).
[15] S. Félix and V. Pagneux, "Ray-wave correspondence in bent waveguides," to be published in Wave Motion (2004).

Acoustic directivity of rectangular pistons on prolate spheroids

Jeffrey E. Boisvert[a)]
NAVSEA Newport, Newport, Rhode Island 02841-5047

A. L. Van Buren
Middletown, Rhode Island 02842

(Received 2 December 2003; revised 10 June 2004; accepted 11 June 2004)

The directivity of acoustic radiation from a rectangular piston arbitrarily located on a rigid prolate spheroidal baffle is formulated. The piston is assumed to vibrate with uniform normal velocity and the solution is expressed in terms of a modal series representation in spheroidal eigenfunctions. The prolate spheroidal wave functions are obtained using computer programs that have been recently developed to provide accurate values of the wave functions at high frequencies. Results are presented in the form of far-field polar directivity patterns for various piston/spheroid acoustic sizes, piston locations on the spheroid, and spheroidal shapes. [DOI: 10.1121/1.1778840]

PACS numbers: 43.20.Rz, 43.30.Jx [LLT] Pages: 1932–1937

I. INTRODUCTION

The directivity of acoustic radiation from sources on baffles has been considered in the literature for various baffle geometries such as planes, spheres, and cylinders,[1,2] and is an important consideration in the design of baffled transducer arrays. For prolate spheroidal baffles, however, published results appear to be limited to radiation patterns from full spheroids.[3,4] In a previous work Boisvert and Van Buren[5] considered the radiation impedance of rectangular pistons mounted on a rigid prolate spheroidal baffle. The present work considers the far-field directivity of acoustic radiation from a rectangular piston, vibrating with uniform normal velocity, conformal to a rigid prolate spheroidal baffle.

II. MATHEMATICAL FORMULATION

A prolate spheroidal surface is generated by the rotation of an ellipse about its major axis. The prolate spheroidal coordinate system (ξ,η,ϕ) is illustrated in Fig. 1. The relationship to Cartesian coordinates is given by

$$x=\frac{d}{2}[(1-\eta^2)(\xi^2-1)]^{1/2}\cos\phi,$$

$$y=\frac{d}{2}[(1-\eta^2)(\xi^2-1)]^{1/2}\sin\phi, \quad z=\frac{d}{2}\eta\xi, \tag{1}$$

where d is the interfocal distance of the generating ellipse, and the ranges of variables are

$$1\leqslant\xi<\infty, \quad -1\leqslant\eta\leqslant 1, \quad 0\leqslant\phi\leqslant 2\pi. \tag{2}$$

The surface of constant ξ is an ellipsoid of revolution about the z axis with a major axis of length ξd and a minor axis of length $(\xi^2-1)^{1/2}d$. Varying the value for the shape parameter ξ produces a wide range of shapes for the prolate spheroid ranging from a straight line $(\xi=1)$ of length d to a sphere $(\xi\to\infty)$.

Consider a conformal region S_i that is assumed to vibrate with normal velocity v_i on an otherwise rigid prolate spheroidal baffle $(\xi=\xi_0)$, immersed in an unbounded fluid of density ρ and sound speed c as shown in Fig. 2. Assuming time harmonic $(e^{i\omega t})$ wave fields, the acoustic field exterior to $(\xi=\xi_0)$ is governed by the Helmholtz equation

$$\nabla^2\Psi+k^2\Psi=0, \tag{3}$$

where ∇^2 is the Laplacian in prolate spheroidal coordinates, $k=\omega/c$, ω is the angular frequency, and $\Psi(\xi,\eta,\phi)$ denotes the spatial portion of the acoustic velocity potential. The Helmholtz equation *is* separable in prolate spheroidal coordinates, and solutions are expressible in eigenfunction expansions of spheroidal wave functions,

$$\Psi_{ml}=R^{(4)}_{ml}(h,\xi)S^{(1)}_{ml}(h,\eta)\begin{Bmatrix}\sin m\phi\\ \cos m\phi\end{Bmatrix}, \tag{4}$$

where

$$R^{(4)}_{ml}(h,\xi)=R^{(1)}_{ml}(h,\xi)-i\,R^{(2)}_{ml}(h,\xi), \tag{5}$$

denotes the prolate spheroidal radial function of the fourth kind that satisfies the radiation condition for outgoing waves, and $h=kd/2$ is the size parameter. The definitions for the prolate spheroidal radial functions of the first and second kind appearing in Eq. (5) and the prolate spheroidal angle function of the first kind $S^{(1)}_{ml}(h,\eta)$ in Eq. (4) are found in Flammer.[6] Because of the asymmetry of region S_i in Fig. 2, both $\sin m\phi$ and $\cos m\phi$ eigenfunctions are required in the solution, hence,

$$\Psi(\xi,\eta,\phi)=\Psi_1(\xi,\eta,\phi)+\Psi_2(\xi,\eta,\phi), \tag{6}$$

where

$$\Psi_1(\xi,\eta,\phi)=\sum_{m=0}^{\infty}\sum_{l=m}^{\infty}A_{ml}R^{(4)}_{ml}(h,\xi)S^{(1)}_{ml}(h,\eta)\cos m\phi, \tag{7}$$

and

$$\Psi_2(\xi,\eta,\phi)=\sum_{m=1}^{\infty}\sum_{l=m}^{\infty}B_{ml}R^{(4)}_{ml}(h,\xi)S^{(1)}_{ml}(h,\eta)\sin m\phi. \tag{8}$$

a)Electronic mail: boisvertje@npt.nuwc.navy.mil

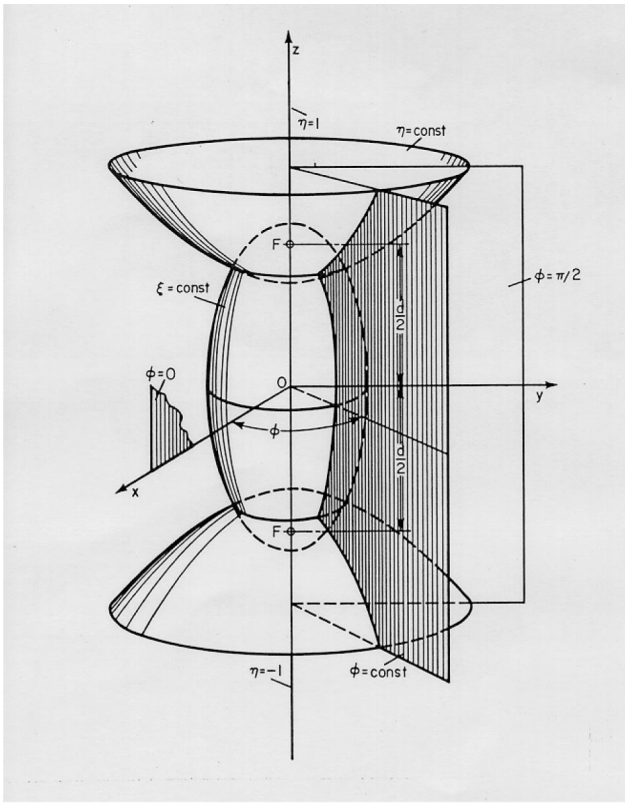


FIG. 1. Prolate spheroidal coordinate system.

As detailed in Ref. 5, the unknown coefficients A_{ml} and B_{ml} are obtained from the boundary condition of normal particle velocity at the spheroid–fluid interface

$$\frac{1}{\Delta_\xi}\left(\frac{\partial \Psi_{1,2}}{\partial \xi}\right)_{\xi=\xi_0} = \begin{cases} v_i(\eta,\phi) & \text{on } S_i \\ 0 & \text{elsewhere,} \end{cases} \tag{9}$$

where $\Delta_\xi = (d/2)(\xi^2-\eta^2)^{1/2}(\xi^2-1)^{-1/2}$ denotes the scale factor in the ξ direction. Since the acoustic pressure is related to the velocity potential by

$$p(\xi,\eta,\phi) = -\rho\frac{\partial \Psi}{\partial t} = -i\omega\rho\Psi(\xi,\eta,\phi), \tag{10}$$

the solution for the acoustic pressure may be written as[5]

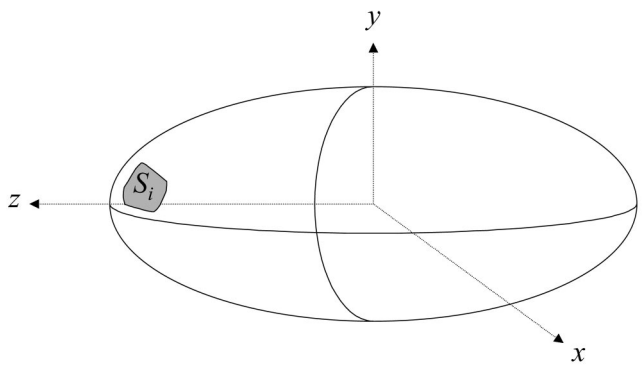


FIG. 2. Conformal region S_i vibrating on a rigid prolate spheroidal baffle.

$$p(\xi,\eta,\phi) = \frac{-ih}{2\pi}\frac{\rho c V}{(\xi_0^2-1)^{1/2}}\sum_{m=0}^{\infty}\sum_{l=m}^{\infty}\varepsilon_m \frac{R_{ml}^{(4)}(h,\xi)S_{ml}^{(1)}(h,\eta)}{R_{ml}^{(4)\prime}(h,\xi_0)N_{ml}} \times[\tilde{I}_{ml}^{s}\cos m\phi + \tilde{I}_{ml}^{c}\sin m\phi], \tag{11}$$

where

$$\tilde{I}_{ml}^{s} = \int\int_{S_i}(\xi_0^2-\eta^2)^{1/2}S_{ml}^{(1)}(h,\eta)\cos m\phi\, d\eta\, d\phi,$$
$$\tilde{I}_{ml}^{c} = \int\int_{S_i}(\xi_0^2-\eta^2)^{1/2}S_{ml}^{(1)}(h,\eta)\sin m\phi\, d\eta\, d\phi, \tag{12}$$

N_{ml} is the prolate spheroidal angle function normalization factor[6,5] and $\varepsilon_m = 1$ when $m=0$, and $\varepsilon_m = 2$ when $m \neq 0$.

It is noted that the integrals residing in Eq. (12) define the size, shape, and location of the piston on the spheroidal baffle, and that a piston is defined by the special case where the source on the baffle vibrates with uniform normal velocity, $v_i(\eta,\phi) = V$.

Upon employing the asymptotic form of the radial function of the fourth kind

$$R_{ml}^{(4)}(h,\xi) \xrightarrow[\xi\to\infty]{} \frac{e^{-i(kr-\pi(l+1)/2)}}{kr}, \tag{13}$$

and the relationship $\eta = \cos\theta$ into Eq. (11), the far-field acoustic pressure is given by

$$p^{\text{far}}(r,\theta,\phi) = \frac{-ih}{2\pi}\frac{\rho c V}{(\xi_0^2-1)^{1/2}}\frac{e^{-ikr}}{kr}f(\theta,\phi), \tag{14}$$

where (r,θ,ϕ) are the spherical coordinates. The directivity function in Eq. (14) is defined by

$$f(\theta,\phi) = \sum_{m=0}^{\infty}\sum_{l=m}^{\infty}\frac{\varepsilon_m S_{ml}^{(1)}(h,\cos\theta)}{R_{ml}^{(4)\prime}(h,\xi_0)N_{ml}}i^{l+1} \times[\tilde{I}_{ml}^{s}\cos m\phi + \tilde{I}_{ml}^{c}\sin m\phi]. \tag{15}$$

Using Eq. (15) the directivity of a piston on a spheroidal baffle is given by

$$F(\theta,\phi) = f(\theta,\phi)/f(\theta_0,\phi_0), \tag{16}$$

where θ_0, ϕ_0 specify the direction of maximum response. Finally the amplitude of the directional response may be expressed in decibels by

$$20\log|F(\theta,\phi)|\text{dB}. \tag{17}$$

III. NUMERICAL COMPUTATION FOR RECTANGULAR PISTONS

Of special interest in this paper is the case where the pistons are quasirectangular in shape, conformal to the surface of the spheroid, and relatively small compared to the spheroid. It is noted that the use of constant values of η and ϕ to define the piston sides would result in pistons that become progressively more wedge-shaped as they approach the tip of the spheroid. Following the methodology of Ref. 5, the center-point location (η_c, ϕ_c), width W, height H, and orientation angle α, of a *planar* rectangular piston, tangent to

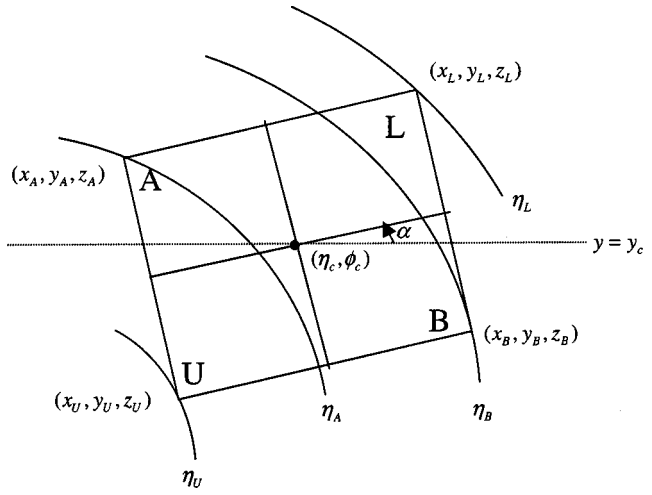


FIG. 3. Rectangular piston corner-points on the surface of the spheroid with $\eta_U > \eta_A > \eta_B > \eta_L$.

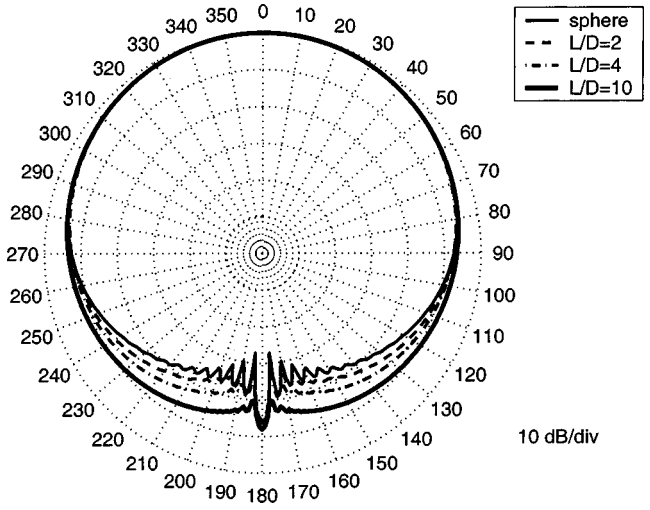


FIG. 5. Horizontal polar directivity pattern of a square piston located at the tip of the spheroidal baffle at $kH = \pi$.

the surface of the spheroid at the point $(\eta_c, \phi_c, \xi = \xi_0)$ is specified. The conformal piston is then defined by the projection of this rectangular piston onto the spheroidal surface $\xi = \xi_0$. Using the methods of analytic geometry, the projection in the $-\hat{\xi}$ direction of the (planar) piston onto the surface of the spheroid provides the (conformal) piston corner coordinates and the lines (curves) corresponding to the perimeter of the piston, as shown in Fig. 3. Here at each corner, the x,y,z value along with the corresponding value of η is designated, where $\eta_U > \eta_A > \eta_B > \eta_L$. Then the integrals over ϕ in Eq. (12) are evaluated numerically with limits that are functions of η as described in Ref. 5.

The calculation of acoustic directivity for a conformal rectangular piston by the use of Eq. (15) requires values of the prolate spheroidal wave functions $S_{ml}^{(1)}$ and $R_{ml}^{(4)\prime}$. Recently developed algorithms for the calculation of the radial functions of the first kind,[7] and second kind,[8] were employed along with an improved algorithm for the calculation of the angular functions in a new FORTRAN computer program called PROFCN[8] that calculates the prolate spheroidal wave functions and their first derivatives.

IV. RESULTS AND DISCUSSION

The equations have been used to determine the directivity amplitude of radiation from a rectangular piston conformal to a prolate spheroidal baffle as a function of acoustic size and piston location for various baffle shapes (elongations). The spheroidal baffle is defined by length, L, maximum diameter, D, and the piston is centered at (η_c, ϕ_c) with height, H, and width, W. The geometry of the spheroid with the far-field angles θ and ϕ is shown in Fig. 4. In all cases that follow, the piston is square $(H = W)$, is centered at $\phi_c = 0$, has an orientation angle of $\alpha = 0$, and unless noted otherwise, has a relative size given by $(H/D = 0.05)$. Four baffle shapes are chosen $L/D = 1.000\ 05$, 2, 4, and 10, where D is held constant and is the same in all cases. Since a prolate spheroid approaches a sphere as $L/D \to 1$, it is noted

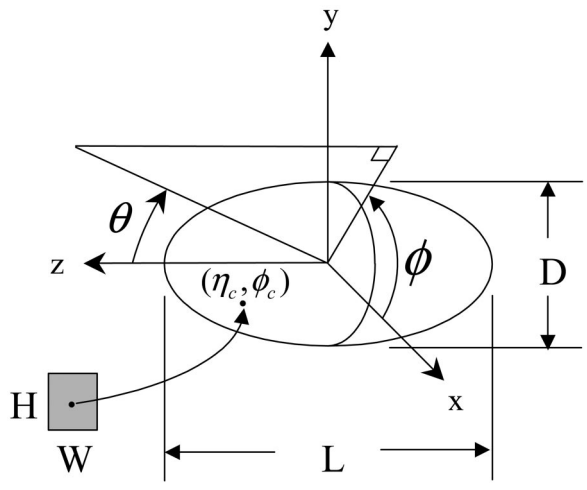


FIG. 4. Geometry of far-field radiation from a rectangular piston on a prolate spheroidal baffle.

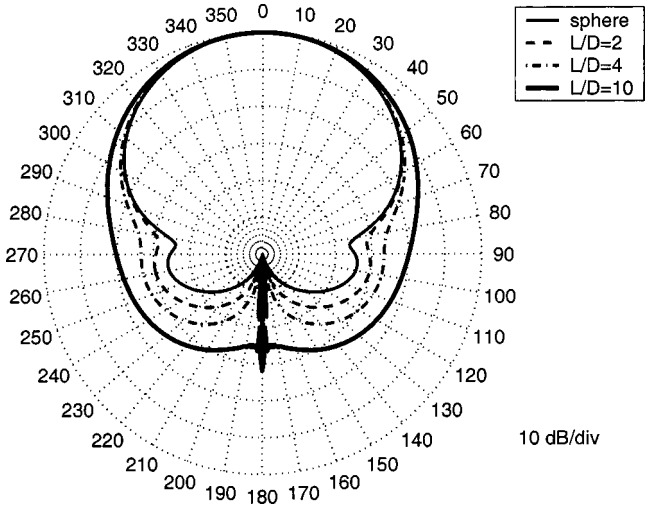


FIG. 6. Horizontal polar directivity pattern of a square piston located at the tip of the spheroidal baffle at $kH = 2\pi$.

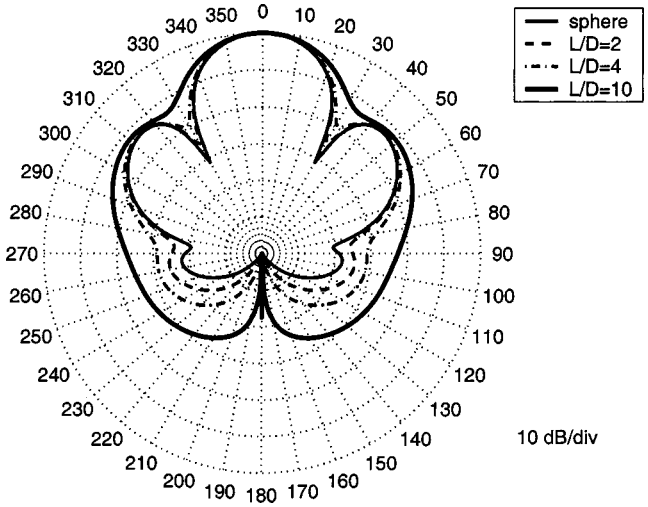


FIG. 7. Horizontal polar directivity pattern of a square piston located at the tip of the spheroidal baffle at $kH=4\pi$.

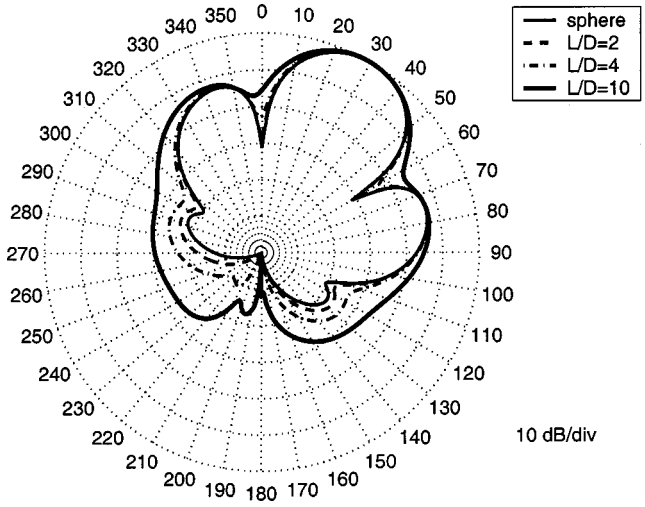


FIG. 9. Horizontal polar directivity pattern of a square piston located on the spheroidal baffle such that the maximum response is 30° at $kH=4\pi$.

that the results for $L/D=1.000\,05$ may be regarded with negligible error as the directivity of a piston on a spherical baffle.

With regard to the determination of convergence of the sums appearing in Eq. (15), for a given value of m, the summation over l is calculated beginning with $l=m$, until the relative contribution to this sum is less than 10^{-8}. This result is then added to the running cumulative sum over m. The procedure is repeated for successive values of m until the relative contribution for a given value of m is less than 10^{-8}.

The number of l and m values required for convergence of the sums appearing in Eq. (15) increases with the acoustic size of the spheroid, and also depends on the baffle shape L/D and the piston size H/D. In general, for a fixed spheroid size, smaller pistons require higher values of both m and l. Additionally the number of m values depends on the location of the piston on the spheroid. Higher values of m are required as the piston location approaches the equator. As an example, for a spheroidal baffle defined by $L/D=4$, and a dimensionless frequency of $kH=4\pi$, 145 values of m and a maximum of 350 values of l are required for a piston of size $H/D=0.05$ located at the equator, and 17 values of m and a maximum of 360 values of l are required for the same piston located at the tip. For the lower dimensionless frequency of $kH=\pi$, the required values of (m,l) are (9,101) for the piston at the tip. For a spheroidal baffle defined by $L/D=2$, and dimensionless frequency of $kH=4\pi$, the required values of (m,l) are (145, 211) for the piston at the equator, and (17, 215) for the same piston at the tip.

Figures 5–7 illustrate the horizontal directivity ($x-z$ plane) as a function of baffle shape for the piston located at the tip ($\eta_c=1$) of the spheroidal baffle, for dimensionless frequencies $kH=\pi,2\pi,4\pi$, respectively. It is noted that as the spheroid becomes more elongated, the results differ from those of the sphere ($L/D=1.000\,05$), however, the results for the short spheroid ($L/D=2$) are nearly identical to those of the sphere at the frequencies given. Below $kH=\pi$, the patterns become more omni-directional and exhibit less dependence on baffle shape.

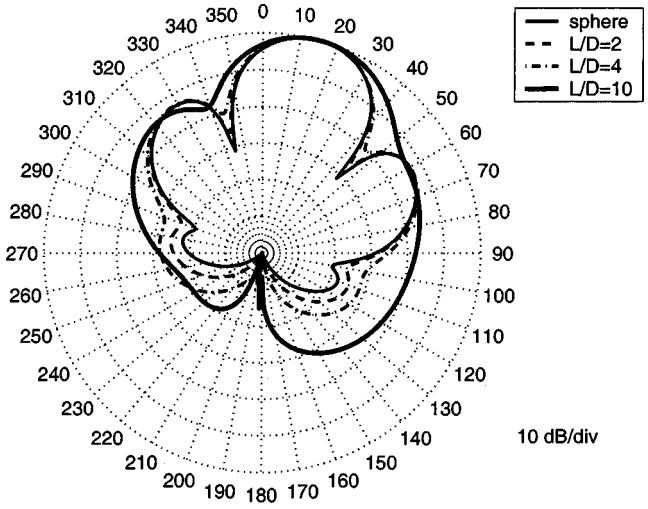


FIG. 8. Horizontal polar directivity pattern of a square piston located on the spheroidal baffle such that the maximum response is 15° at $kH=4\pi$.

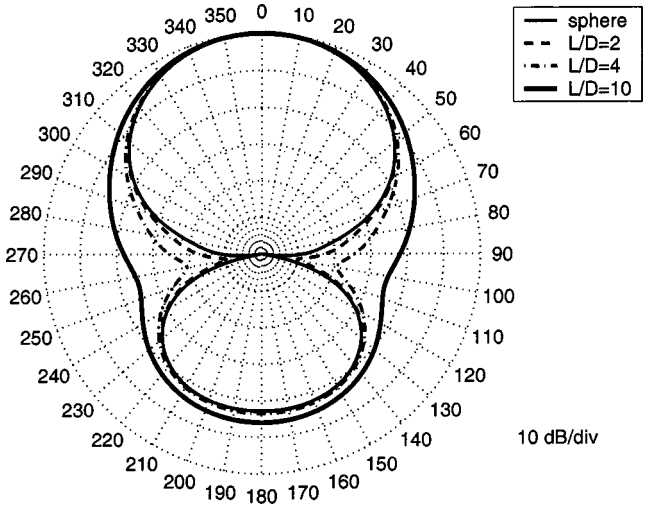


FIG. 10. Vertical polar directivity pattern of a square piston located on the spheroidal baffle such that the maximum response is 30° at $kH=4\pi$.

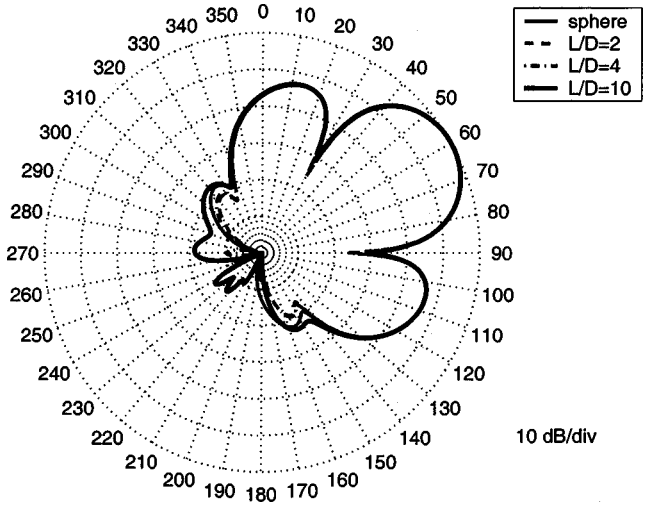


FIG. 11. Horizontal polar directivity pattern of a square piston located on the spheroidal baffle such that the maximum response is 60° at $kH=4\pi$.

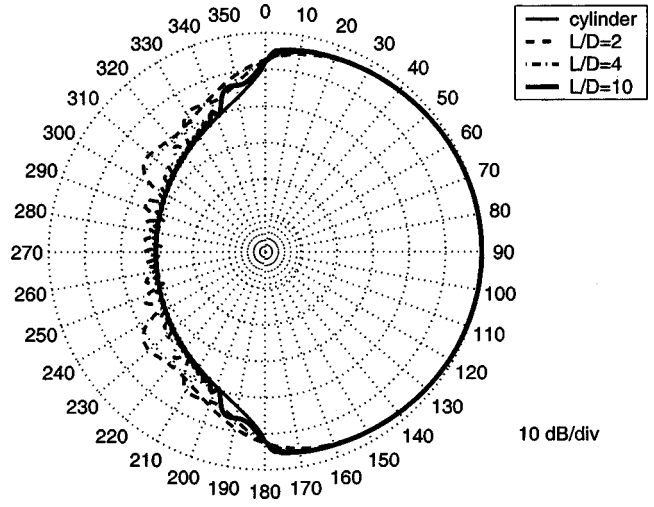


FIG. 13. Horizontal polar directivity pattern of a square piston located at the equator of the spheroidal baffle at $kH=\pi$.

To examine the behavior of a piston located away from the tip of the spheroid, Fig. 8 shows the case where the maximum response of the piston occurs at 15° on each baffle, for $kH=4\pi$. It is noted that the piston has a different value of η_c for each baffle shape, that is, the η_c location for each baffle was chosen such that the maximum response occurs at 15° in each case. Aside from the sphere (whose pattern is simply rotated 15° from the case at the tip), it is seen the elongated baffle ($L/D=2,4,10$) patterns become more asymmetrical as baffle elongation increases. Figure 9 gives the case where the maximum response of the piston occurs at 30°, for $kH=4\pi$. Figure 10 shows the corresponding vertical pattern ($x-y$ plane). Figures 11 and 12 show the horizontal and vertical patterns, respectively, for a piston maximum response of 60°. The figures presented show that as the piston is placed farther away from the tip of the baffle, the effect of baffle elongation (relative to a spherical baffle) diminishes. Now to examine the behavior of a piston located at the equator ($\eta_c=0$) of the spheroid, Figs. 13 and 14 show the horizontal directivity ($x-z$ plane), and Figs. 15 and 16

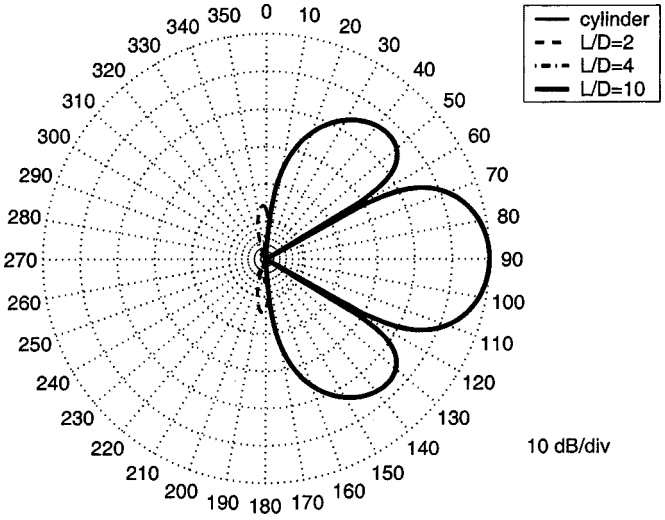


FIG. 14. Horizontal polar directivity pattern of a square piston located at the equator of the spheroidal baffle at $kH=4\pi$.

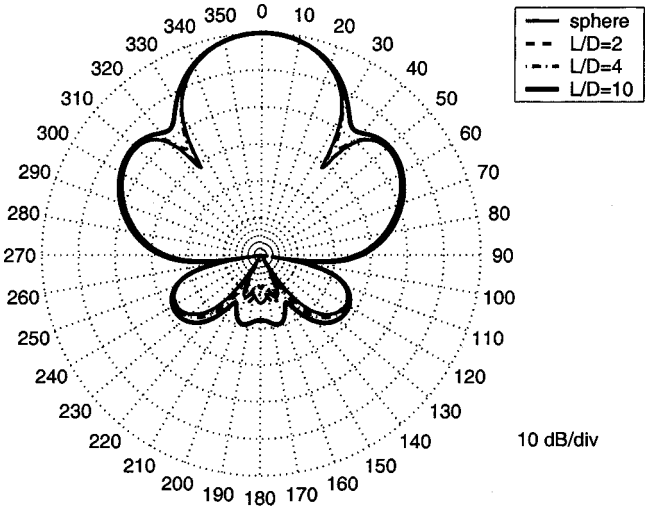


FIG. 12. Vertical polar directivity pattern of a square piston located on the spheroidal baffle such that the maximum response is 60° at $kH=4\pi$.

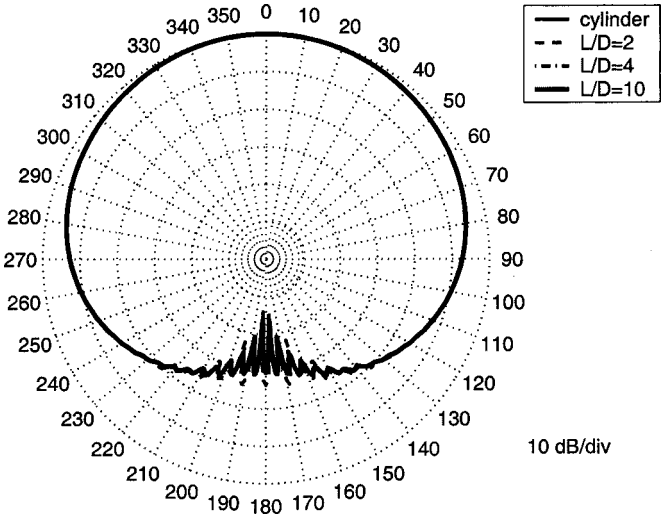


FIG. 15. Vertical polar directivity pattern of a square piston located at the equator of the spheroidal baffle at $kH=\pi$.

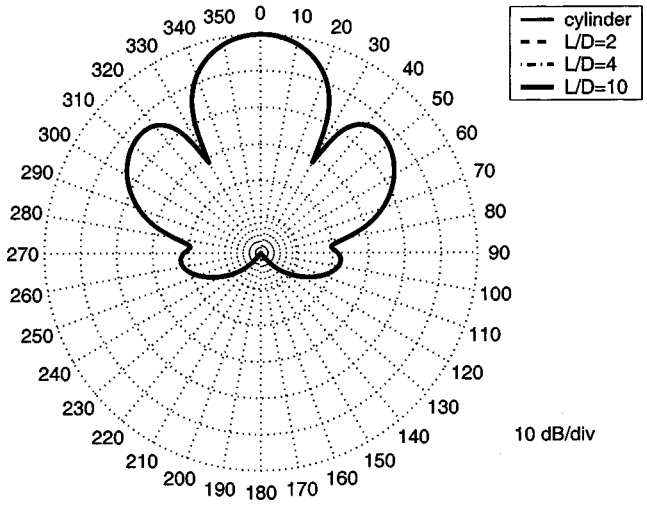


FIG. 16. Vertical polar directivity pattern of a square piston located at the equator of the spheroidal baffle at $kH=4\pi$.

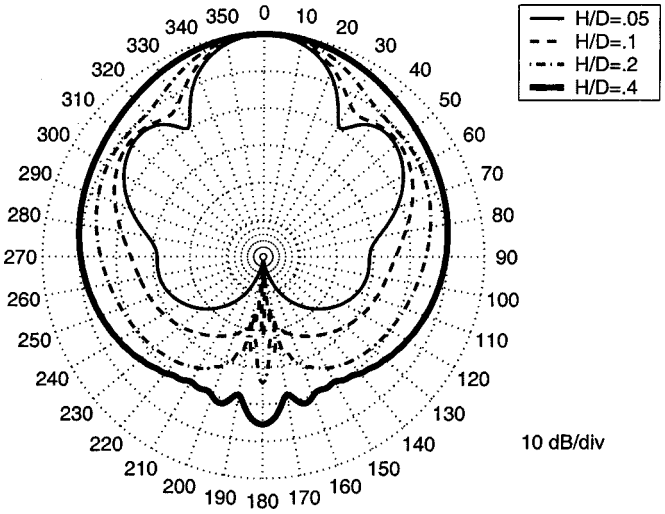


FIG. 17. Horizontal polar directivity pattern of square pistons of varying size located at the tip of an $L/D=4$ spheroidal baffle at $kH=4\pi$.

show the vertical directivity ($x-y$ plane) for $kH=\pi$ and 4π, respectively, for baffle shapes $L/D=2,4,10$, along with results for the piston on a circular cylinder with the same diameter. Figures 13 and 14 show that as the baffle elongation increases the horizontal directivity approaches that of a piston on a cylindrical baffle. Since the baffles under consideration have the same diameter at the equator, the curvature is the same, and hence, Figs. 15 and 16 show that the directivity in the vertical plane is virtually indistinguishable from the case of a piston on a cylindrical baffle of the same diameter.

Finally to illustrate the effect of larger piston size relative to a given baffle, Fig. 17 shows the horizontal directivity for various piston sizes all located at the tip of a spheroidal baffle defined by $L/D=4$. Here each piston has a conformal area equivalent to a square piston of normalized length H/D, where $H/D=0.05,0.1,0.2,0.4$, and the acoustic size is $kH=4\pi$ in all cases. These results show that for a given acoustic size, as the piston extends over more of the tip of the spheroid, its projected area becomes smaller and therefore the patterns become less directional.

V. SUMMARY

The expression for the acoustic directivity of radiation from a rectangular piston source vibrating on a rigid prolate spheroidal baffle has been developed and is given by Eqs. (15)–(17). Sample results have been presented that illustrate the behavior of the directivity (amplitude) for pistons on various spheroidal baffle shapes over a wide frequency range. It should be noted that the phase response is directly obtained from Eq. (15), although no results were presented. In addition, comparisons of the directivity of a rectangular piston on a spheroid versus on a sphere and on a cylinder with the same equatorial diameter D were presented. It was found that for a piston located at the equator of the spheroid, the directivity approaches that of a piston on a cylindrical baffle, and for the case of the piston located at the tip of the spheroid, the results approach those of a spherical baffle only for relatively short spheroids ($L/D\leqslant 2$). Although not shown, the directivity of an array of pistons conformal to a prolate spheroidal baffle is readily obtainable by superposition of the individual single element responses. It is noted that the acoustic directivity for pistons of other shapes on prolate spheroids is obtainable by the evaluation of Eq. (12) with limits of integration appropriate for the piston shape of interest.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research Code 321 and the Naval Undersea Warfare Center ILIR Program.

[1] P. M. Morse, *Vibration and Sound* (McGraw-Hill, New York, 1948).
[2] D. T. Laird and H. Cohen, "Directionality patterns for acoustic radiation from a source on a rigid cylinder," J. Acoust. Soc. Am. **24**, 46–49 (1952).
[3] V. A. Andebura, "Acoustical properties of spheroidal radiators," Sov. Phys. Acoust. **15**, 447–454 (1970).
[4] G. Chertock, "Sound radiation from prolate spheroids," J. Acoust. Soc. Am. **33**, 871–876 (1961).
[5] J. E. Boisvert and A. L. Van Buren, "Acoustic radiation impedance of rectangular pistons on prolate spheroids," J. Acoust. Soc. Am. **111**, 867–874 (2002).
[6] C. Flammer, *Spheroidal Wave Functions* (Stanford University Press, Stanford, CA, 1957).
[7] A. L. Van Buren and J. E. Boisvert, "Accurate calculation of prolate spheroidal radial functions of the first kind and their first derivatives," Q. Appl. Math. **60**, 589–599 (2002).
[8] A. L. Van Buren and J. E. Boisvert, "Improved calculation of prolate spheroidal radial functions of the second kind and their first derivatives," Q. Appl. Math. (in press).

Adhesion and nonlinear scattering by rough surfaces in contact: Beyond the phenomenology of the Preisach–Mayergoyz framework

Claudio Pecorari[a)]
Marcus Wallenberg Laboratory, Royal Institute of Technology, 100 44 Stockholm, Sweden

(Received 17 December 2003; revised 23 June 2004; accepted 4 July 2004)

Phenomenological models reproducing the elasticity and acoustic properties of geomaterials and materials with damage have been successfully developed. These models yield macroscopic stress–strain constitutive equations featuring hysteresis with end-point memory, and predict the efficient generation of higher harmonics accompanying the propagation of monochromatic waves. The assumption common to these models is that the material's microstructure is characterized by nonlinear compliant components of an unspecified nature which can exist in two states: ''open'' or ''closed.'' The density of the compliant units is defined on a mathematical continuum (the Preisach–Mayergoyz space) whose elements identify the dynamic behavior of the components. In this work, adhesion is shown to introduce hysteresis with end-point memory in the macroscopic behavior of an interface between two rough surfaces in contact, and, upon scattering, to generate higher harmonics bearing a striking similarity to those observed in wave propagation phenomena in media with distributed damage and in geomaterials. It appears, therefore, that two rough surfaces interacting via adhesion forces offer a meaningful example of macroscopic interface or bond with dynamics resembling that of the fictitious elements of the Preisach–Mayergoyz space, and acoustic nonlinear properties similar to those of rocks and damaged materials.
[DOI: 10.1121/1.1785616]

PACS numbers: 43.25.Ba, 43.25.Dc, 43.25.Jh [MFH] Pages: 1938–1947

I. INTRODUCTION

The elasticity of geomaterials and materials with damage at the scale of their microstructure has been extensively investigated for its peculiar (*nonclassical*) features, which cannot be reconciled with the *classical* nonlinearity of the potential governing the material interaction at the atomic level.[1,2] In particular, the hysteresis with end-point memory displayed by the stress–strain relationship of these materials has been shown to be accompanied by a variety of acoustic phenomena which, too, seem to be specific to this state of solid matter. Such phenomena include the unusually large amplitude of the higher harmonics—especially the odd harmonics—generated by a propagating wave,[3,4] the dependence of the third harmonic on the second power of the amplitude of the fundamental excitation,[5] and the dependence of the resonant frequency shift on the amplitude of the acoustic strain field.[6] Slow dynamics following conditioning of the material via dynamic loading or by stresses of thermal origin is also often observed in samples displaying the above-mentioned nonclassical nonlinearity.[7] However, whether the origin of slow dynamics resides in the hysteretic, nonlinear stress–strain constitutive relation of the material or results from the diffusion of heat generated by the interaction of the conditioning force with the material imperfection is still a matter of debate.[7,8]

In support of the laboratory efforts, a few phenomenological models have been developed, which successfully reproduce the main experimental findings, at least in a qualitative sense.[9–14] All these models assume the behavior of the physical system to be determined by a distribution of compliant units of an unspecified nature, which are allowed to exist in one of only two available states: ''up'' or ''down,'' or ''open'' or ''closed.'' The density of these units is defined on a mathematical continuum known as the Preisach–Mayergoyz (P-M) space.[1,9] Each dimension of a P-M space corresponds to a physical quantity which controls the dynamic behavior of a basic unit. The coordinates of each point of a P-M space, in particular, identify the values of the independent quantities at which the transitions between the two available states occur. Implicit in the model is also the assumption that the direction of the transition is determined by the history of the unit. This feature of the models leads to constitutive equations of the macroscopic state of the system, which display hysteresis with end-point memory. Within the context of the phenomenology considered in this work, the independent parameter is usually the applied stress, and the constitutive equations link the macroscopic stress and strain fields in the material. For this reason, a compliant element whose dynamics is defined by a point of a P-M space is often referred to as *hysteretic elastic unit* (HEU). Note that a HEU is a purely mathematical entity having *a priori* arbitrary physical and geometrical dimensions. Hence, the custom of attributing a ''mesoscopic'' scale to a HEU is misleading to some extent, and it is motivated only by the need to overcome a serious deficiency of the model's phenomenological nature: it can describe a phenomenon, but it cannot provide a causal explanation of it. In other words, the descriptive power of these models notwithstanding, their phenomeno-

a)Electronic mail: pecorari@kth.se

logical nature prevents the model from shedding light on the physical origin of the phenomena of interest. Therefore, a different approach must by adopted if any progress towards understanding the physical origin of the nonclassical nonlinear elastic behavior of geomaterials and materials with damage is to be made.

According to the received knowledge, microcracks and soft boundaries between adjacent mesoscopic material constituents (grains) are the loci where the mechanism(s) responsible for the dynamic phenomena of interest here acts. The earlier work in which the P-M space framework was developed[9] has borrowed its major ideas and terminology from Holcomb,[15] who introduced the concept of reversible Griffith crack to explain the dilatancy of rocks under compressive differential stresses. A reversible Griffith crack is one which extends its length when the local tensile stress exceeds a limit value, σ_o, and, upon reversing the load, returns to its original configuration when the applied stress decreases below a second threshold value, $\sigma_c < \sigma_o$. An essential and strong assumption for a crack to be reversible is the occurrence of healing during unloading, which can occur only if the faces of the crack conform to each other at the atomic level. Dilatancy appears at stresses of the order of 100 MPa.[15] Strain fields of the order of 10^{-4} in materials with Young's modulus lower than 100 GPa are accompanied by stresses lower than 10 MPa. This value is at least one order of magnitude below the stress threshold required for a rock to start dilating. That the reversible Griffith crack may provide an unlikely physical model for a HEU is supported also by recent experimental results[16] which establish the onset of nonclassical nonlinear acoustic phenomena in rocks at strain values of the order of 10^{-7}. The natural conclusion, therefore, is that the reversible Griffith cracks cannot be used as a physical model behind the nonlinear dynamics of rocks, and, thus, while the validity of the P-M model as a framework to describe these phenomena is not questionable, the identification of the physical nature of an HEU still remains an open issue.

This work examines the nonlinear dynamics and acoustic properties of imperfect interfaces in greater depth by extending a previous investigation[17] on the nonlinear interaction between acoustic waves and imperfect interfaces to include the effects of adhesion forces on the dynamics of a boundary between rough surfaces in contact. Other authors have argued that adhesion can explain some acoustic phenomena in rocks, such as the frequency independence of the attenuation of seismic waves.[18] To this end, however, the effect of adhesion on the dynamics of single contacts between spherical bodies, rather than that on the dynamics of interfaces or extended bonds, was examined. In this context, Burnham and Kulik's concluding remark to their review of adhesion and other surface forces provides a needed warning against easy extrapolations: "Of course, the gap between laboratory experiments that study one single contact and the real world where thousands or million of asperities are involved remains enormous."[19] In addition, the model considered by Sharma and Tutuncu[18] does not include the Tabor parameter, which is fundamental to a general description of the adhesion between asperities.

In the following, therefore, a micromechanical model developed by Greenwood and Johnson[20] (GJ model) to describe the contact between two spherical bodies interacting via both elastic forces and forces of adhesion is presented first. Next, the micromechanics of interacting asperities predicted by the GJ model is incorporated into the framework developed by Greenwood and Williamson[21] to derive the mechanics of two nominally flat, nonconforming rough surfaces in contact. The case of an interface subjected to a cyclic load is examined in detail. Adhesion between the surfaces in contact is shown to lead to hysteresis with end-point memory in the relationship between the applied stress and the relative approach of the two surfaces. The results of this section are then used to formulate effective boundary conditions to be enforced on the acoustic field of a longitudinal wave at normal incidence. This boundary value problem is finally solved by means of a classical perturbation approach in which two small parameters measuring the nonlinearity of the interface are used as perturbation parameters. The amplitude of higher harmonics is shown to display features distinguishing the spectra of waves propagating in geomaterials and materials with damage from those of waves supported by classical nonlinear materials. Other mechanisms have been recently examined which lead to the appearance of hysteresis with end-point memory in the dynamics of a solid–solid interface. This is the case of asperities in contact subjected to a tangential alternating force.[17] However, the magnitude and the dependence on the normal interfacial stiffness of the higher harmonics generated by such a mechanism can hardly be compared with those presented here and observed in geomaterials and materials with damage. In conclusion, this investigation identifies adhesion between contacts as a physical mechanism involving interfacial components with mesoscale dimensions, which leads to macroscopic nonlinear acoustic properties that are characteristic of systems described by the phenomenological models based on the P-M space. A series of critical remarks on the present work concludes this communication.

II. MICROMECHANICS OF ROUGH SURFACES IN CONTACT

A. Single contact

The interaction between two spheres, or a sphere and a flat surface, involving both elastic and adhesive forces is controlled by a single parameter, μ, known as the Tabor parameter. In this work, the definition of μ given by Greenwood and Johnson[20] is adopted: $\mu = \sigma_o (R/(E'^2 \Delta\gamma))^{1/3}$ (Ref. 22). The symbol σ_o is the maximum adhesive stress acting on the contact; R is the composite radius of curvature, which is defined by $R = (R_1^{-1} + R_2^{-1})^{-1}$, where R_1 and R_2 are the radii of curvature of the two spheres; E' is the reduced Young modulus of the contact, $E' = ((1-\nu_1^2)/E_1 + (1-\nu_2^2)/E_2)^{-1}$; and $\Delta\gamma$ is the surface energy. The symbols E_i and ν_i, with $i = 1, 2$, are the Young and the Poisson moduli of the two materials, respectively. In the GJ model, the interaction between the two bodies in contact is described by superimposing two Hertzian stress distributions of opposite sign. The compressive Hertzian stress, as in the original

TABLE I. Physical and geometrical parameters defining the contact.

μ	R (μm)	$\Delta\gamma$ (mJ m^{-2})	δ_c (nm)	F_c (10^{-6} N)
0.079	1	40.	0.37	0.25

model by Hertz, acts over a circular area of radius a. The tensile stress, which simulates the effect of adhesion, acts also over a surrounding circular annulus with external radius $c>a$. The Hertzian character of the two distributions allows their superposition to yield a uniform displacement of the points in contact over the area $r<a$. Therefore, like in the original theory by Hertz, the force applied to the centers of the spheres, $F=F(\delta|\mu)$,[23] can be related to the relative approach of such points, δ. However, no simple direct mathematical relationship exists between the two quantities. Rather, four equations are given to link δ and F, which involve four additional model parameters: a, c, μ, and k (see the Appendix). The latter parameter, which controls the intensity of the tensile stress, is introduced in the model to obtain a uniform displacement distribution of points for which $r<a$.

By varying the Tabor parameter between 0 and ∞, the whole spectrum of cases ranging from that of two rigid spheres in contact (the Bradley model,[24] $\mu=0$) to that contemplated by Johnson, Kendall, and Roberts[25] (the JKR model, $\mu\to\infty$) can be covered. The latter properly describes the case of a contact characterized by large values of R and surface energy, and/or between rather soft materials. For a thorough discussion and rigorous analysis of the interaction of two spheres the reader is referred to the work of Greenwood.[26]

Fuller and Tabor[27] described the adhesion between nominally flat, rough surfaces of Perspex and rubber by implementing the JKR model into the Greenwood and Williamson framework[21] that will be discussed later. Their results on a Perspex-rubber interface are used here to set an arbitrarily large reference value of μ, i.e., $\mu=100$, from which that of other material interfaces can be derived under the additional assumption that the maximum adhesive stress, σ_o, remains constant. In particular, a value as small as $\mu=0.079$ is obtained for contacts between Perspex and steel with relevant parameter values shown in Table I. This approach is motivated by the lack of sufficient information on the physical parameters characterizing the micromechanics of contacts considered in this work, that is to say, contacts with $\mu<1$. Figure 1 illustrates the dependence of the normalized force, $F^*(\delta^*|\mu)=F(\delta^*|\mu)/F_c$, on the normalized relative approach, $\delta^*=\delta/\delta_c$, where $F_c=(2\pi R\Delta\gamma)$, $\delta_c=(\beta^2/R)$, and $\beta=(R^2\Delta\gamma/E')^{1/3}$, for $\mu=0.079$. With this normalization, the maximum normalized tensile force, $\bar{F}$, varies from 1 for $\mu=0$ to $\bar{F}=0.75$ for $\mu\to\infty$. In particular, for $\mu=0.079$, $\bar{F}\cong 0.981$, that is to say, it differs from the value typical of a rigid contact by about 2%.

According to the GJ model, there is no long-range interaction. Thus, during approach, the first contact between the spheres is established when $\delta^*=0$. At that point, an attractive force draws the two bodies together, and a new equilibrium configuration characterized by a finite contact area, and, thus, a finite approach, is established by the balance between the attractive adhesive force and the elastic reaction to it. The application of an additional external compressive force increases the relative approach as in the Hertzian case. Upon unloading, the force–approach relationship retraces the loading curve in the opposite direction. At $\delta^*=0$, however, the area of the contact as well as its stiffness are still finite. Therefore, in order for a complete detachment to occur, the applied tensile force must be further increased and the contact must be stretched beyond the point at which it was established. If the load on the contact is transmitted by a device having infinite compliance, such as an acoustic wave incident on the interface, then the contact breaks when its stiffness becomes null, that is to say, when the tangent to the curve in Fig. 1 is parallel to the δ^* axis. At this point, $\delta^*=\delta$ and $F^*(\delta|\mu)=\bar{F}$. This situation closely resembles that occurring during a wave scattering event in which an interface is partially closed by an instrument controlling the load, and the contacts are formed and broken by the stress carried by the wave field.

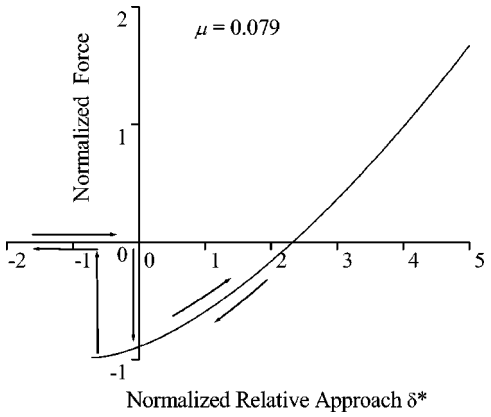


FIG. 1. Plot of the normalized force-displacement relationship for a contact between Perspex and steel spheres characterized by the parameters of Table I according to the Greenwood and Johnson model. The arrows indicate the path followed by the force as a relative approach varies.

B. Rough surfaces in contact

Greenwood and Williamson[21] developed a framework to evaluate the mechanical properties of an interface formed by a rough surface in contact with a flat rigid one. This approach is valid under the assumption that the contacts are independent of each other. Here, Brown and Scholz's[28] extension of their result to the case of two rough surfaces in contact is used. Following Brown and Scholz, the original problem is transformed into that of an auxiliary surface (the composite surface) pressed against an infinitely rigid flat (see Fig. 2). The relationship between the applied pressure, P, and the relative approach, Δ, between the mean planes of the rough surfaces in contact can be written as follows:

$$P(\Delta)=N\int_{Z_0-\Delta,\alpha_0,\beta_0}^{Z_0,\alpha_1,\beta_1} F(z-Z_o+\Delta|\lambda_1,\lambda_2,\ldots)\times\varphi(Z_o-z)\,\phi(\lambda_1)\,\eta(\lambda_2)\cdots dz\,d\lambda_1 d\lambda_2\cdots. \quad (1)$$

In Eq. (1), $\varphi(\cdot)$ is the probability density function of the peaks of the auxiliary, composite surface. The latter is de-

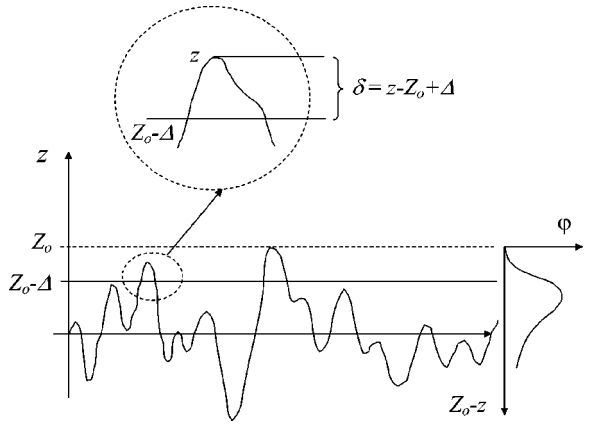


FIG. 2. Coordinate systems for the profile of the composite surface and for the probability density function of its asperities. The inset explains the relation between the coordinate of the flat rigid surface pressed against the composite one, the deformation of a given asperity, and the coordinates of the profile.

fined by an appropriate algebraic combination of the profiles of the two rough surfaces of interest, which maps the individual contacts between the asperities of the two surfaces into the peaks of the composite one.[28,29] Thus, $\varphi(Z_o-z)dz$, which gives the number of peaks with height between z and $z+dz$ above the mean plane of the composite surface, represents also the number of contacts formed in this interval. Z_o is the maximum height of the asperities of the composite surface, and, thus, $\varphi(Z_o-z)=0$ for $z>Z_o$. In Eq. (1), N is the number of contacts per unit area, and $F(\cdot)$ is the force law between asperities, which depends on their relative approach, $\delta=z-Z_o+\Delta$, as well as on additional contact parameters, λ_i, $i=1,2,\ldots$. The values of the latter for each contact are generally unknown, and, thus, are to be considered stochastic variables with probability density functions $\phi(\cdot)$, $\eta(\cdot),\ldots$, respectively. The integration over z is carried out between the actual position of the flat surface with respect to the mean plane of the composite surface, $Z_o-\Delta$, and the initial position of the same surface for $P=0$, Z_o. The integrals over the other stochastic variables are also evaluated over appropriate ranges of values.

Introducing the force law of the GJ model in Eq. (1), the latter becomes

$$P(\Delta^*)=2\pi N\int_{0,R_0,\Delta\gamma_0}^{\Delta^*,R_1,\Delta\gamma_1} R\Delta\gamma F^*(\Delta^*-t^*|\mu)\varphi(t^*)\times\phi(R)\,\eta(\Delta\gamma)dt^*dRd(\Delta\gamma). \quad (2)$$

In Eq. (2), the new nondimensional variable $t^*=(Z_o-z)/\delta_c$ has been introduced, and λ_i, $i=1,2$, have been identified with R and $\Delta\gamma$, respectively. Consistently, the nondimensional approach is $\Delta^*=\Delta/\delta_c$. For the sake of conciseness, the Tabor parameter has replaced R and $\Delta\gamma$ in the expression of the force law. The distribution of the values of R has been assumed to be independent of the asperity height z, and the surface energy $\Delta\gamma$, which depends only on the nature of the materials, has been taken to be the same for all the contacts. Thus, $\eta(\cdot)=\delta(\cdot)$, where $\delta(\cdot)$ is the delta of Dirac, and Eq. (2) becomes

$$P(\Delta^*)=2\pi N\Delta\gamma\int_{0,R_0}^{\Delta^*,R_1} R\times F^*(\Delta^*-t^*|\mu)\varphi(t^*)\phi(R)dt^*dR. \quad (3)$$

Fuller and Tabor[27] developed a similar model in which the force law of the JKR model was used. In their work, they implicitly assumed that all the contacts have the same composite radius of curvature, R. Although this is a rather drastic approximation, the effects of which will be discussed later, it will be adopted even in this work for the sake of simplicity. Therefore, setting $\phi(\cdot)=\delta(\cdot)$ in Eq. (3) yields

$$P(\Delta^*)=2\pi NR\Delta\gamma\int_0^{\Delta^*}F^*(\Delta^*-t^*|\mu)\varphi(t^*)dt^*. \quad (4)$$

Following Baltazar *et al.*,[30] the probability density $\varphi(t^*)$ is chosen to be a chi-squared probability density function

$$\varphi(t^*)=\left(\sqrt{\frac{2}{n}}\Sigma\right)^{-n/2}\frac{(t^*)^{(n-2)/2}}{\Gamma(n/2)}\exp\left(-\sqrt{\frac{n}{2}}\frac{t^*}{\Sigma}\right). \quad (5)$$

In Eq. (5), the parameter n, which is known as the *number of degrees of freedom* of the probability density, is an integer greater or equal to 2, while $\Sigma=\sigma/\delta_c$ is the normalized rms roughness of the composite profile. The latter, neglecting the normalization factor, is related to those of the two surfaces by $\sigma^2=\sigma_1^2+\sigma_2^2$. For $n=2$, $\varphi(t^*)$ is an exponential function, while for $n\to\infty$, it approaches a Gaussian distribution function.

Having brought the two surfaces to a maximum normalized approach $\Delta^*_{\max}$ at the end of loading phase of the first cycle, the relationship between the applied pressure, P, and the normalized relative approach, Δ^*, during unloading is given by

$$P(\Delta^*)=2\pi NR\Delta\gamma\int_0^{\Delta^*+D}F^*(\Delta^*-t^*|\mu)\varphi(t^*)dt^*, \quad (6)$$

where $D=\underline{\Delta}^*_{\max}-\Delta^*$, if $0<\Delta^*_{\max}-\Delta^*<\overline{\delta}$, and $D=0$, if $\Delta^*_{\max}-\Delta^*>\overline{\delta}$. The inclusion of D in the upper limit of integration is to account for the stretching of the peaks that had been formed last during the preceding loading phase of the cycle.

The main interest of this investigation is in the dynamic behavior of the interface when it is subjected to a cyclic loading. Thus, if $\Delta^*_{\min}$ is the relative approach at the end of the unloading phase of the cycle, the pressure–approach relationship during all the following loading phases is given by

$$P(\Delta^*)=2\pi NR\Delta\gamma\int_0^{\Delta^*+D}F^*(\Delta^*-t^*|\mu)\varphi(t^*)dt^*, \quad (7)$$

where $D=\overline{\delta}-\Delta^*+\Delta^*_{\min}$, if $0<\Delta^*-\Delta^*_{\min}<\overline{\delta}$, and $D=0$, if $\Delta^*-\Delta^*_{\min}>\overline{\delta}$. The upper limit of integration accounts for the effect of contacts that are under tension at the end of the unloading cycle, and are now progressively set under increasing compression again during the current compressive phase of the cycle.

Let the interface be subjected to a static load P_o upon which an oscillating component, ΔP, is superimposed. Fig-

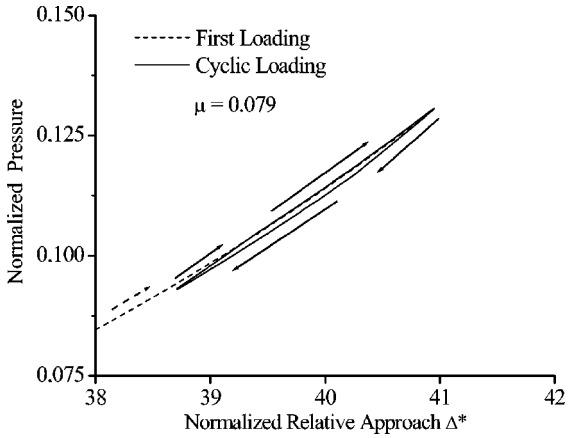


FIG. 3. An example of the hysteresis loop displayed by the normalized pressure-relative approach relationship of an interface between two rough surfaces in contact when subjected to a cyclic load. The interface separates two half-spaces of Perspex and steel, and the parameters characterizing the contacts are given in Table I.

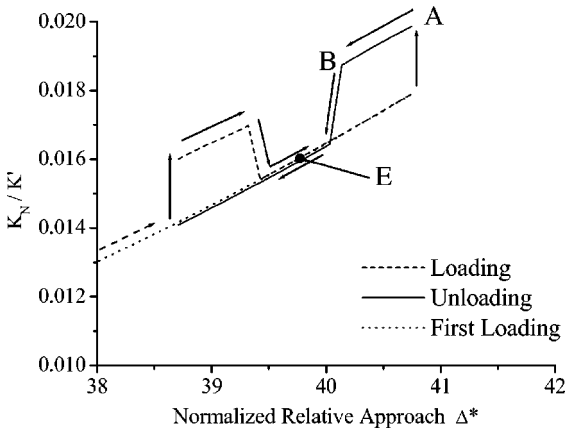


FIG. 4. Dependence of the normal interfacial stiffness on the interface relative approach. The normalization constant is $K'=Np_c/\delta_c$, where N is the number of contacts per unit area. The points A and B indicate the path along which the system evolves when the maximum variation of the interface opening displacement is smaller than $\Delta/2$, while E designates the configuration determined by the static load. The interface is that of the previous figure.

ure 3 illustrates an example of such a pressure–approach relationship for an interface between Perspex and steel. The Young modulus of the Perspex is $E_1=2.85\,\mathrm{GPa}$, and that of steel is $E_2=192\,\mathrm{GPa}$, while the values of the Poisson modulus are $\nu_1=0.4$ and $\nu_2=0.28$, respectively. The value of $\Delta\gamma=40\ \mathrm{mJ/m^2}$ for the surface energy change is used,[27] while that of the peak density, N, of the composite surface is obtained by employing the approximation $N=0.1/R\sigma$, which was found experimentally by Fuller and Tabor.[27] In addition, the rms roughness of the composite surface is $\sigma=\sqrt{\sigma_1^2+\sigma_2^2}=30\,\mathrm{nm}$, where σ_1 and σ_2 are the rms roughnesses of the two surfaces. A number of degrees of freedom, N, equal to 10 completes the characterization of the probability distribution density of the composite asperities. The parameter values characterizing the individual contacts are those reported in Table I. In Fig. 3, the pressure, P, is normalized by $P_c=2\pi NR\Delta\gamma$: $P^*=P/P_c$. The amplitude of the oscillation in physical units is equal to 1 nm.

For the purpose of the present investigation, the most important feature displayed by Fig. 3 is the hysteresis with end-point memory displayed by the plot of the normalized pressure, P^*, versus the normalized relative approach, Δ^*. Such hysteretic behavior is caused by the forces of adhesion. This fact can be better understood by the analysis of the interfacial stiffness, K_N, as a function of the relative approach (Fig. 4). In a quasi-static experiment the stiffness, K_N, which in Fig. 4 is normalized with respect to $K'=F_c/\delta_c$, is evaluated mathematically by the first derivative of the static pressure with respect to the relative approach, $K_{N,o}=(\partial P/\partial\Delta)_{\Delta_o}$ at $\Delta=\Delta_0$, where Δ_0 is the value of the approach at equilibrium (marked by the point "E" in Fig. 4). In Fig. 4, the behavior of K_N as the interface reaches its equilibrium during the quasi-static loading is illustrated by the dotted line. When probed dynamically at the position of its current equilibrium by a periodic perturbation that produces a variation of the normalized approach not exceeding the normalized distance $\overline{\Delta}/2=\delta_c\delta/2$, the stiffness of the interface undergoes a discontinuous positive variation after reaching its maximum value at the end of the first compressive phase. After such an event, it varies continuously along the upper side of the path corresponding to the unloading cycle. In other words, the interface dynamic stiffness is that which the interface displays at the beginning of a full unloading cycle. In Fig. 4, the initial part of the unloading cycle is marked by the letters A and B. Thus, for small variations of the instantaneous relative approach from its value at equilibrium, the dynamic stiffness displayed by the interface during the following cycles is larger than the quasi-static stiffness, $K_{N,o}$, and can be approximated by a linear expansion in $\delta\Delta$: $K_N\cong K_o+K_1\delta\Delta$. The constant term K_o can still be evaluated as the first derivative of the pressure with respect to the relative approach, as in the case of $K_{N,o}$, $K_0=(\partial P/\partial\Delta)_{\Delta_0}$. However, care must be taken to perform this derivative during the unloading part of the cycle. Similarly, $K_1=(\partial^2 P/\partial\Delta^2)_{\Delta_0}$.

For variations of the instantaneous relative approach having amplitude $\delta\Delta$ such that $\delta\Delta\geqslant\overline{\Delta}/2$, at each turning point in a cycle and at a distance equal to Δ from them, the stiffness undergoes sudden discontinuous variations during all cycles following the first one. The positive jump at the beginning of the unloading phase is due to the stiffening reaction accompanying the stretching of the contacts last formed during the previous loading phase; that at the opposite end of the cycle is determined by the onset of the removal of the contacts under tension. The negative jump during unloading occurs when the first rupture of contacts takes place, while that during the loading cycle is determined by the completion of the removal of all the contacts under tension. In conclusion, during dynamic loading, the interface stiffness can be approximated by the following expression:

$$K_N(\Delta=\Delta_o+\delta\Delta)$$
$$=K_0+K_1\delta\Delta-K_2'H\left(\mathrm{sgn}\left(-\frac{\partial\Delta}{\partial t}\right)\right)H(\Delta_{\max}-\overline{\Delta}-\Delta)$$
$$-K_2''H\left(\mathrm{sgn}\left(\frac{\partial\Delta}{\partial t}\right)\right)H(\Delta-\Delta_{\min}-\overline{\Delta}),\qquad(8)$$

in which, in addition to the linear expansion already considered, two products of step functions have been introduced. In the first product, the first step function is nonzero only during the unloading phase of a cycle, while in the second one, the first step function is nonzero during the loading phase. The second step functions in each product describe the negative jumps of the stiffness occurring during a cycle. The coefficients K_2' and K_2'' measure the stiffness variations taking place during the jumps.

The behavior of the stiffness just discussed is another manifestation of the hysteresis with end-point memory affecting the pressure–approach relationship, and demonstrates the similarity between the dynamics of this type of interface and that of a fictitious HEU. Both can exist in two dynamic states, and the transitions between the latter are determined by threshold values of the independent variable, and by the history of the state. In addition, contacts between asperities are unquestionably material features with typical mesoscopic dimensions.

III. WAVE REFLECTION AND TRANSMISSION

In this section, the effective boundary conditions needed to describe the scattering of a longitudinal wave by an interface such as those described above are formulated. The propagation direction of the incident wave is assumed to be normal to the interface, and the boundary value problem is solved by means of a standard perturbation approach.

From Eq. (8), the behavior of the normal stiffness of the interface as a function of the interface opening displacement (IOD) oscillation, Δu is approximated by the following function:

$$K_N(\Delta u)=K_0-K_1\Delta u-K_2'H\left(\operatorname{sgn}\left(\frac{\partial \Delta u}{\partial t}\right)\right)\times H(\Delta u+\Delta u_{\max}-\overline{\Delta})-K_2''H\left(\operatorname{sgn}\left(-\frac{\partial \Delta u}{\partial t}\right)\right)\times H(\Delta u_{\max}-\overline{\Delta}-\Delta u), \tag{9}$$

where $\delta\Delta=-\Delta u$, $\Delta u_{\min}=\Delta_o-\Delta u_{\max}$, and the symbol $\Delta u_{\max}$ denotes the amplitude of the IOD's oscillation, Δu. Therefore, the first boundary condition to be enforced at the interface is

$$\sigma_{33}(0^+,t)=K_N(\Delta u)\Delta u, \tag{10}$$

where $K_N=K_N(\Delta u)$ is given by Eq. (9), $\Delta u(t)=u(0^+,t)-u(0^-,t)$, and $u(0^+,t)$ and $u(0^-,t)$ are the total displacement fields on the positive and negative sides of the interface, respectively. The second boundary condition requires the normal stress to be continuous across the interface,

$$\sigma_{33}(0^+,t)=\sigma_{33}(0^-,t). \tag{11}$$

The incident wave is assumed to propagate in the half-space $z<0$. The Lamé constants of the negative half-space are λ^- and μ^-, while λ^+ and μ^+ are those of the positive half-space. If A_{in} is the amplitude of the incident wave, the total fields in the negative and positive half-spaces can be written in terms of two new nondimensional functions $u^-(z,t)=A_{in}\xi^-(z,t)$, and $u^+(z,t)=A_{in}\xi^+(z,t)$, respectively. Introducing the following nondimensional variables, $\eta=k^-z$, where k^- is the longitudinal wave number in the negative half-space, and $\tau=\omega t$, where ω is the angular frequency of the incident wave, the boundary conditions for this problem can be cast in the following form:

$$k^-(\lambda^++2\mu^+)\frac{\partial \xi^+}{\partial \eta}=K_0\Delta\xi-K_1A_{in}\Delta\xi^2-K_2'\left[H\left(\operatorname{sgn}\left(\frac{\partial \Delta\xi}{\partial \tau}\right)\right)\times H(\Delta\xi+\Delta\xi_{\max}-\overline{\Delta})+\theta H\left(\operatorname{sgn}\left(-\frac{\partial \Delta\xi}{\partial \tau}\right)\right)\times H(\Delta\xi_{\max}-\overline{\Delta}-\Delta\xi)\right]\Delta\xi, \tag{12}$$

$$(\lambda^++2\mu^+)\frac{\partial \xi^+}{\partial \eta}=(\lambda^-+2\mu^-)\frac{\partial \xi^-}{\partial \eta}. \tag{13}$$

In Eq. (12), Δ has been redefined as $\overline{\Delta}=\overline{\Delta}/A_{in}$, and $\theta=K_2''/K_2'$. The solutions are sought in the form of a perturbation series in two small parameters, ε_1 and ε_2:

$$\xi^-(\eta,\tau)=\tfrac{1}{2}\{\exp[i(\tau-\eta)]-R\exp[i(\tau+\eta)]-\varepsilon_1U_1(\eta,\tau)-\varepsilon_2U_2(\eta,\tau)+\cdots+\text{C.C.}\}, \tag{14}$$

$$\xi^+(\eta,\tau)=\tfrac{1}{2}\{T\exp[i(\tau-\kappa\eta)]+\varepsilon_1V_1(\eta,\tau)+\varepsilon_2V_2(\eta,\tau)+\cdots+\text{C.C.}\}, \tag{15}$$

in which R and T are the linear reflection and transmission coefficients of the incident wave, $\varepsilon_1=K_1A_{in}/K_0$, $\varepsilon_2=K_2'/K_0$, $\kappa=(C_L^-/C_L^+)^2$, where the C_Ls are the longitudinal phase velocities in the two half-spaces, and C.C. represents the complex conjugate.

In the series expansions of Eqs. (14) and (15), the two perturbation parameters ε_1 and ε_2 are treated as independent of each other regardless of the fact that the physics of the interface is unique. Regardless of the nature of the connection between the two parameters, such an approach is justified in view of the additivity of the nonlinear corrections to the interfacial stiffness determined by them. In fact, accounting for the relationship between ε_1 and ε_2 leads to a system of boundary conditions for the first-order displacement fields in which the ''driving force'' is the linear superposition of the contributions due to hysteresis and to the term which is quadratic in the IOD. Therefore, the first-order correction of the displacement fields can be decomposed into two components, each of which is separately determined by the corresponding nonlinear term of the interfacial stiffness. In conclusion, for the purpose of finding the first-order correction to the displacement fields, ε_1 and ε_2 can be regarded as independent of each other.

Figure 5 illustrates an example of the dependence of the two nonlinear parameters ε_1 and ε_2 and of $\varepsilon_2'=\theta\varepsilon_2$ on the normalized stiffness $K_0/(Z^-\omega)$, where $Z^-=\rho^-C_L^-$, the product of the mass density of the medium and the phase velocity of the propagating wave, is the acoustic impedance

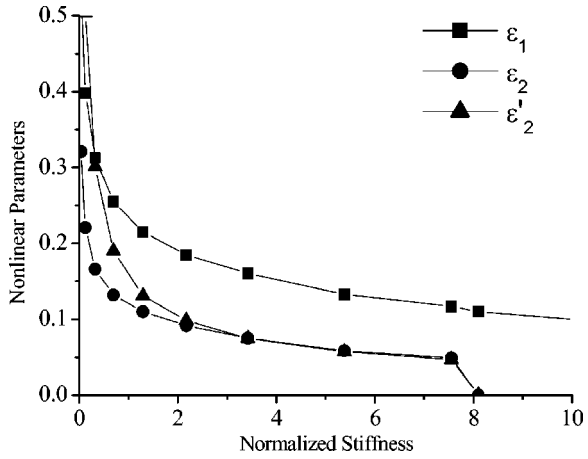


FIG. 5. Nonlinear perturbation parameters versus normalized stiffness. All three parameters diverge as the interface opens. Those controlling the nonlinearity due to adhesion suddenly go to zero around $K_0/Z^-\omega=8$, when the maximum variation of the interface opening displacement becomes smaller than $\overline{\Delta}/2$. The amplitude of the incident wave is $A_{in}=2$ nm.

of Perspex, and ω is the angular frequency of the incident wave. The amplitude of the latter is assumed to be $A_{in}=2$ nm. These data are evaluated for an interface between Perspex and steel with composite roughness $\sigma=\sqrt{\sigma_1^2+\sigma_2^2}=30$ nm, where σ_1 and σ_2 are the rms roughness of the two surfaces, and by a probability distribution density with a number of degrees of freedom $N=10$. All three parameters are shown to diverge as the interface opens. In addition, ε_2 and $\varepsilon_2'=\theta\varepsilon_2$ go suddenly to zero in the neighborhood of $K_0/Z^-\omega=8$, where, in this particular case, the amplitude of the IOD variation becomes smaller than $\overline{\Delta}$.

Introducing Eqs. (14) and (15) in the boundary conditions, and separating the terms according to their dependence on the perturbation parameters ε_1 and ε_2, the following boundary conditions for the zeroth- and first-order solutions are obtained:
zeroth-order system:

$$k^-(\lambda^++2\mu^+)(-i\kappa T)=K_0(T-1+R), \tag{16}$$

$$(\lambda^++2\mu^+)\kappa T=(\lambda^-+2\mu^-)(1+R); \tag{17}$$

first-order system in ε_1:

$$k^-(\lambda^++2\mu^+)\frac{\partial V_1}{\partial\eta}+\text{C.C.}$$
$$=K_0\left\{V_1+U_1+\text{C.C.}-\frac{1}{2}(T-1+R)^2\right.$$
$$\left.\times\exp[i2\tau]-|T-1+R|^2+\text{C.C.}\right\}, \tag{18}$$

$$(\lambda^++2\mu^+)\frac{\partial V_1}{\partial\eta}+\text{C.C.}=-(\lambda^-+2\mu^-)\frac{\partial U_1}{\partial\eta}-\text{C.C.}, \tag{19}$$

first order system in ε_2:

$$k^-\left(\lambda^++2\mu^+\right)\frac{\partial V_2}{\partial\eta}+\text{C.C.}$$
$$=K_0\left\{V_2+U_2+\text{C.C.}+\left[H\left(\text{sgn}\left(\frac{\partial\Delta\xi_0}{\partial\tau}\right)\right)\right.\right.$$
$$\times H\left(\Delta\xi_0+\overline{\Delta}_{\max}-\overline{\Delta}\right)+\theta H\left(\text{sgn}\left(-\frac{\partial\Delta\xi_0}{\partial\tau}\right)\right)$$
$$\left.\left.\times H\left(\overline{\Delta}_{\max}-\overline{\Delta}-\Delta\xi_0\right)\right]\left(\Delta\xi_0+\text{C.C.}\right\}\right), \tag{20}$$

$$(\lambda^++2\mu^+)\frac{\partial V_2}{\partial\eta}+\text{C.C.}=-(\lambda^-+2\mu^-)\frac{\partial U_2}{\partial\eta}-\text{C.C.}. \tag{21}$$

In Eq. (20), the arguments of the step functions have been approximated by using the solutions of the zeroth-order system, which is a reasonable approximation as long as the nonlinearity of the interface is small. The solutions of the zeroth-order system are

$$R=-\frac{i\omega Z^--K_0(1-Z^-/Z^+)}{i\omega Z^-+K_0(1+Z^-/Z^+)}\quad\text{and}\quad T=\frac{Z^-}{Z^+}R, \tag{22}$$

where $Z^\pm=(\rho C_L)^\pm$ are the longitudinal acoustic impedances of the positive and negative half-spaces, respectively. The symbol ρ represent the mass density of the medium. The solutions V_2 and U_2 of the first first-order system are found by employing R and T to evaluate the terms on the right-hand side of Eq. (20) which contain them. Such terms are further expanded in a Fourier series, so that Eq. (20) can be recast as follows:

$$k^-(\lambda^++2\mu^+)\frac{\partial V_2}{\partial\eta}+\text{C.C.}$$
$$=K_0\left\{V_2+U_2+\text{C.C.}+c_0+\sum_n c_n e^{in\tau}+\text{C.C.}\right\}. \tag{23}$$

This equation, together with Eq. (21), is solved by expanding V_2 and U_2 in Fourier series,

$$U_2(\eta,\tau)=\sum_n A_n e^{in(\tau+\eta)}\quad\text{and}\quad V_2(\eta,\tau)=\sum_n B_n e^{in(\tau-\kappa\eta)},$$

and by introducing such representations of the unknown solutions in the boundary conditions. Simple algebra leads to the following expressions for the coefficients of the series:

$$A_n=-\frac{K_0c_n}{in\omega Z^-+K_0(1+Z^-/Z^+)}\quad\text{and}\quad B_n=\frac{Z^-}{Z^+}A_n. \tag{24}$$

By using the harmonic balance methods, the solutions of the boundary value problem of Eqs. (18) and (19) are found to comprise terms that are constants and others that contain the second harmonic component. The amplitudes of the first ones are

$$U_{1,0}=\frac{|T-1+R|^2}{1+Z^-/Z^+}\quad\text{and}\quad V_{1,0}=\frac{Z^-}{Z^+}U_{1,0}, \tag{25}$$

while those of the second harmonic components are

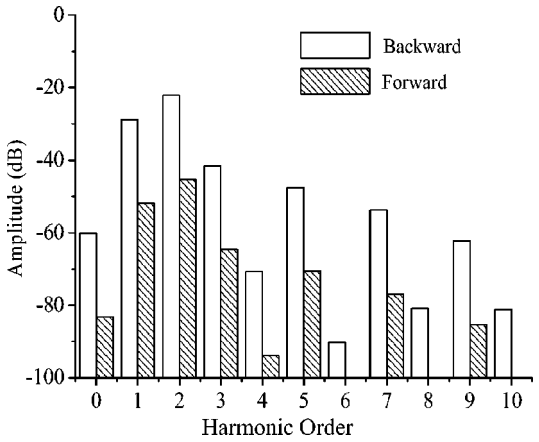


FIG. 6. Normalized spectrum of nonlinear waves that are scattered forward and backward by the interface considered in Figs. 3 and 4, and which is characterized by a normalized linear component of the interface stiffness $K_0/Z^-\omega=1.29$.

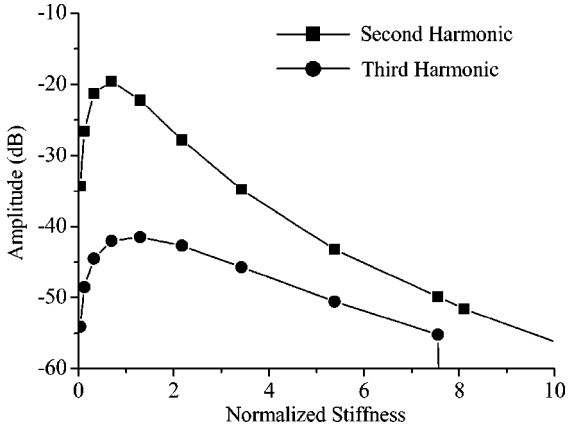


FIG. 7. Normalized amplitude of the second and third harmonic waves versus the normalized interfacial stiffness for the interface considered in the previous two figures.

$$U_{1,2}=\frac{K_0}{2}\frac{(T-1+R)^2}{i2\omega Z^-+K_0(1+Z^-/Z^+)} \quad \text{and} \quad V_{1,2}=\frac{Z^-}{Z^+}U_{1,2}. \tag{26}$$

Note the positive value of both $U_{1,0}$ and $V_{1,0}$. This result is consistent with the fact that a net effect of nonlinearity is the softening of the interface for increasing values of the amplitude of the incident wave.

Figure 6 illustrates an example of normalized spectrum of the higher harmonics generated upon scattering of an incident wave with amplitude $A_{in}=2$ nm, frequency $f=1$ MHz, and propagating in the Perspex half-space. The amplitude of the incident wave is used as normalization constant. The interface is that of Fig. 5 with a value of the normalized stiffness equal to 1.29. To the first-order approximation, and when hysteresis is activated—indeed, for values of $\Delta\xi_{\max}<\bar{\Delta}/2$ only the classical nonlinear corrections are present—the amplitude of all the higher harmonics is a function of A_{in}^2, as experimentally verified in damaged materials, at least for the third harmonic component.[5] In the same approximation, the amplitude of the even harmonics of an order higher than the second depends also on the departure of θ from unity, that is to say, on the loss of symmetry between the compressive and the tensile parts of each cycle.

It is important to point out the remarkable resemblance between the spectra in Fig. 6 and those predicted by Van den Abeele *et al.*[10] in their theoretical investigation on nonlinear propagation of acoustic waves in nonlinear media, in which the P-M model was used to characterize the type and degree of nonlinearity. Meegan *et al.*[3] and TenCate *et al.*[4] also reported measured spectra of waves propagating in sandstone and other geomaterials, which bear a strong resemblance to those of Fig. 6. In particular, the slow decay of the amplitude of the higher harmonics together with the dominant presence of the odd components seems constitute the acoustic signature of an hysteretic interface with end-point memory.

Finally, Fig. 7 illustrates the dependence of the second and third harmonics on the normalized interface stiffness, or, in other terms, on the interface opening. As the contacts begin to form, both components increase until they reach a maximum value near $K_0/Z^-\omega=1$. However, for very small values of K_0, the parameters ε_1 and ε_2 exceed the value of 0.3, rendering the approximation no longer accurate. The numerical estimates of the second and third harmonic amplitudes beyond the range of validity of the approximation are reported nonetheless to illustrate the trend the actual solutions are expected to follow. Note that the dependence of the amplitude of the third harmonic on the normalized stiffness is considerably weaker than that displayed in the case of surfaces in contact and subjected to an oscillating tangential load.[17] However, as already remarked in the discussion of Fig. 5, the amplitude of the third harmonic goes to zero when the amplitude of the IOD variation becomes smaller than Δ.

IV. CONCLUDING REMARKS

A few comments are still in order to complete the description of the present model. First of all, the mechanics of the contacts considered in this work does not account for any long-range interaction which may exist in real contacts, leading two asperities to form a contact when δ is still negative (the ''jump-on'' phenomenon[26,30]). In addition to the work done against the forces of adhesion during unloading, which, for a single contact, measures the area below the force-displacement curve for negative displacement values, some energy is lost due to the impact accompanying a jump-on event. A rigorous theoretical treatment should also account for this effect. However, as long as only a small percentage of the asperities is involved, as required by the Greenwood and Williamson approximation,[21] and the impact between asperities is characterized by a coefficient of restitution proximal to 1, such a mechanism is expected not to alter significantly the conclusions presented here since it involves a small fraction of the energy carried by the acoustic wave.

Equations (4), (6), and (7) relating the pressure, P, to the relative approach, Δ, have been obtained under the assumption that all the asperities have the same radius of curvature. This assumption is unlikely to be satisfied by normal, rough surfaces, and, therefore, the radius of curvature in the equations just cited has to be weighted by its own probability density function. Note also that contacts with different radii of curvature are characterized by different values of the Ta-

bor parameter, μ. Consequently, the stiffness of the interface is expected to undergo sudden transitions only at the endpoints of each cycle (see Fig. 4), the remaining two transitions becoming continuous, in view of the fact that contacts having different radii of curvature come apart at different values of the distance δ. The width of the regions where such continuous transitions occur closely follows that of the probability density function of the radius of curvature. Finally, under these conditions, the spectrum of a wave scattered by such an interface is expected to be characterized by a content of higher harmonics that is less pronounced than that shown in Fig. 5.

As the Tabor parameter, μ, goes to zero by maintaining a constant value of the surface energy, $\Delta\gamma$, the present model of the mechanics of two rough surfaces in contact is expected to approach that by Maugin,[31] which was developed by using the Derjaguin–Muller–Toporov force law of individual contacts. According to the latter model, the bodies in contact do not deform in the annular area, which surrounds the contact and where the force of adhesion acts. Furthermore, the maximum adhesive force is that predicted by Bradley. If the limit $\mu\rightarrow 0$ is achieved by letting the maximum adhesive stress, σ_0, become progressively smaller, the case described by Pecorari[17] of a nonlinear purely elastic interface is recovered.

The limitation of the present work notwithstanding, the similarity of the spectra of Fig. 6 and those experimentally observed in geomaterials, or predicted by phenomenological theories based on the P-M model, can hardly be considered coincidental. In fact, although it is reasonable to assume that more than one mechanism concurs to determine the nonlinear behavior of a complex medium with microstructure and/or damage, the presence of soft bonds and microcracks in rocks or other composite materials provides ample opportunity for contact nonlinearity to play a significant role. Furthermore, since neither Hertzian nonlinearity nor the hysteresis of the tangential load-displacement relationship of the individual contacts leads to a nonlinear response of a magnitude comparable to that investigated here,[17] adhesion becomes a serious candidate to be considered as a fundamental mechanism behind the nonclassical nonlinear physical behavior of hysteretic materials with end-point memory.

Finally, the micromechanics model presented in this work may also provide the key to interpreting the excess linearity observed in partially saturated rocks.[32] In fact, the mechanics of adhesion between soft bodies can be used to reasonably describe the interaction between asperities covered by very thin layers of water or other liquids. This conjecture seems to be supported also by a theoretical investigation by Goryacheva and Makhovskaya.[33] Should this be the case, then stress-corrosion cracks formed in pipes containing corrosive liquids at high pressure may also be expected to display nonlinear acoustic properties which resemble those described in this work.

ACKNOWLEDGMENTS

This work was performed as part of the project ''Nonlinear scattering from partially closed cracks and imperfect interfaces,'' No. 14.43-010902,01156, supported by the Swedish Inspectorate for Nuclear Power (SKI). Partial support of the European Science Foundation through the NATEMIS program, and of the Swedish Research Council through Contract No. 621-2003-4320, is also acknowledged.

APPENDIX: FORCE LAW FOR A SINGLE CONTACT[20]

The equations, which, by linking the six contact parameters μ., a^*, c^*, k, F^*, and δ^*, establish the force law for a contact between spheres interacting by means of both adhesion and elastic forces, are presented next. Of these equations, the first one gives the normalized outer radius, $c^*=c/\beta$, as a function of the inner radius, a^*, which is treated as an independent variable, and the Tabor parameter, μ. The latter plays the role of input parameter of the model. The normalization constant β is defined in the text:

$$\mu\left[(c^{*2}-a^{*2})^{1/2}+\frac{\pi}{2}\mu\right](c^*-a^*)(c^*+2a^*)=3(c^*+a^*). \tag{A1}$$

With the second equation, the parameter k is found in terms of a^*, c^*, and μ:

$$\mu=k\frac{2}{\pi}(c^{*2}-a^{*2})^{1/2}. \tag{A2}$$

Finally, the normalized force, F^*, and the normalized relative approach between the centers of the spheres in contacts, δ^*, are obtained by solving the following equations:

$$F^*=\frac{2}{3\pi}[a^{*3}-k(c^{*3}-a^{*3})] \tag{A3}$$

and

$$\delta^*=a^{*2}-k(c^{*2}-a^{*2}). \tag{A4}$$

[1] R. A. Guyer and P. A. Johnson, ''Nonlinear mesoscopic elasticity: evidence for a new class of materials,'' Phys. Today **4**, 30–36 (1999).
[2] L. A. Ostrovsky and P. A. Johnson, ''Dynamic nonlinear elasticity in geomaterials,'' Riv. Nuovo Cimento **24**(7), 1–46 (2001).
[3] D. G. Meegan, Jr., P. A. Johnson, R. A. Guyer, and K. R. McCall, ''Observations of nonlinear elastic wave behavior in sandstone,'' J. Acoust. Soc. Am. **94**, 3387–3391 (1993).
[4] J. A. TenCate, K. E.-A. Van Den Abeele, and T. J. Shankland, ''Laboratory study of linear and nonlinear elastic pulse propagation in sandstone,'' J. Acoust. Soc. Am. **100**, 1383–1391 (1996).
[5] K. E.-A. Van Den Abeele, P. A. Johnson, and A. Sutin, ''Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage, Part I: Nonlinear wave modulation spectroscopy,'' Res. Nondestruct. Eval. **12**, 17–30 (2000).
[6] R. A. Guyer, J. A. TenCate, and P. A. Johnson, ''Hysteresis and the dynamic elasticity of consolidated granular materials,'' Phys. Rev. Lett. **82**, 3280–3283 (1999).
[7] J. A. TenCate, E. Smith, and R. A. Guyer, ''Universal slow dynamics in granular materials,'' Phys. Rev. Lett. **85**, 1020–1023 (2000).
[8] V. Zaitsev, V. Gusev, and B. Castagnede, ''Thermoelastic mechanism for logarithmic slow dynamics and memory in elastic wave interactions in individual cracks,'' Phys. Rev. Lett. **90**, 075501 (2003).
[9] K. R. McCall and R. A. Guyer, ''Equation of state and wave propagation in hysteretic nonlinear elastic materials,'' J. Geophys. Res. **99**(B12), 23,887–23,897 (1994).
[10] K. E.-A. Van Den Abeele, P. A. Johnson, R. A. Guyer, and K. R. McCall, ''On the quasi-analytic treatment of hysteretic nonlinear response in elastic wave propagation,'' J. Acoust. Soc. Am. **101**, 1885–1898 (1997).
[11] B. Capogrosso-Sansone and R. A. Guyer, ''Dynamic model of hysteretic elastic systems,'' Phys. Rev. B **66**, 224101 (2002).
[12] P. P. Delsanto and M. Scalerandi, ''Modelling nonclassical linearity, con-

ditioning, and slow dynamics effects in mesoscopic elastic materials,'' Phys. Rev. B **68**, 064107 (2003).
[13] M. Scalerandi, P. P. Delsanto, and P. A. Johnson, ''Stress induced conditioning and thermal relaxation in the simulation of the quasi-static compression experiments,'' J. Phys. D **36**, 288–293 (2003).
[14] M. Scalerandi, V. Agostini, P. P. Delsanto, K. Van Den Abeele, and P. A. Johnson, ''Local interaction simulation approach to modelling nonclassical, nonlinear elastic behavior in solids,'' J. Acoust. Soc. Am. **113**, 3049–3059 (2003).
[15] J. A. TenCate, D. Pasqualini, S. Habib, K. Heitman, D. Higdon, and P. A. Johnson, ''Nonlinear and nonequilibrium dynamics in geomaterials,'' Phys. Rev. Lett. **93**(6), 065501 (2004).
[16] D. J. Holcomb and J. L. Stevens, ''The reversible Griffith crack: a viable model for dilatancy,'' J. Geophys. Res. **85**, 7101–7107 (1980).
[17] C. Pecorari, ''Nonlinear interaction of plane ultrasonic waves with an interface between rough surfaces in contact,'' J. Acoust. Soc. Am. **113**, 3065–3072 (2003).
[18] M. M. Sharma and A. N. Tutuncu, ''Grain contact adhesion hysteresis: a mechanism for attenuation of seismic waves,'' Geophys. Res. Lett. **21**, 2323–2326 (1994).
[19] N. Burnham and A. A. Kulik, ''Surface Forces and Adhesion,'' in *Handbook of Micro/nano Tribology*, in *CRC Series in The Mechanics and Material Science*, 2nd ed., edited by B. Bhushan (CRC Press, New York, 1999), Chap. 5.
[20] J. A. Greenwood and K. L. Johnson, ''An alternative to the Maugis model of adhesion between elastic spheres,'' J. Phys. D **31**, 3279–3290 (1998).
[21] J. A. Greenwood and J. B. P. Williamson, ''Contact of nominally rough surfaces,'' Proc. R. Soc. London, Ser. A **295**, 300–319 (1966).
[22] According to its original definition, μ is the ratio between the extension of the contact at which its rupture occurs under fixed load, $(R\Delta\gamma^2/E'^2)^{1/3}$, and the range of the interaction potential between the surfaces, ε: $\mu=(R\Delta\gamma^2/E'^2)^{1/3}/\varepsilon$. However, for two spheres interacting via a Lennard-Jones force law, $\varepsilon=1.026\Delta\gamma/\sigma_o$, which renders the definition of μ by Greenwood and Johnson[20] equivalent to that by Tabor for all practical purposes.
[23] The notation $f=f(x|\alpha)$ is used to describe the dependence of the dependent variable f on the independent variable x and on a system parameter α.
[24] R. S. Bradley, ''The cohesive force between solid surfaces and the surface energy of solids,'' Philos. Mag. **13**, 853–862 (1932).
[25] K. L. Johnson, K. Kendall, and A. D. Roberts, ''Surface energy and contact of elastic solids,'' Proc. R. Soc. London, Ser. A **324**, 301–313 (1971).
[26] J. A. Greenwood, ''Adhesion of elastic spheres,'' Proc. R. Soc. London, Ser. A **453**, 1277–1297 (1998).
[27] K. N. G. Fuller and D. Tabor, ''The effect of surface roughness on the adhesion of elastic solids,'' Proc. R. Soc. London, Ser. A **345**, 327–342 (1975).
[28] S. R. Brown and C. H. Scholz, ''Closure of elastic random surfaces in contact,'' J. Geophys. Res. **90**, 5531–5545 (1985).
[29] A. Baltazar, S. I. Rokhlin, and C. Pecorari, ''On the relationship between ultrasonic and micro-mechanic properties of contacting rough surfaces,'' J. Mech. Phys. Solids **50**, 1397–1416 (2002).
[30] K. Inagaki, O. Matsuda, and O. B. Wright, ''Hysteresis of the cantilever shift in ultrasonic force microscopy,'' Appl. Phys. Lett. **80**, 2386–2388 (2002).
[31] D. Maugis, ''On the contact and adhesion of rough surfaces,'' J. Adhes. Sci. Technol. **10**, 161–175 (1996).
[32] K. E.-A. Van Den Abeele, J. Carmeliet, P. A. Johnson, and B. Zinszner, ''Influence of water saturation on the nonlinear mesoscopic response of earth materials, and the implications to the mechanism of nonlinearity,'' J. Geophys. Res. **107**, 101029–101040 (2002).
[33] I. G. Goryacheva and Y. Y. Makhovskaya, ''Adhesive interaction of elastic bodies,'' J. Appl. Math. Mech. **65**, 273–282 (2001).

Nonlinear acoustic simulations using direct simulation Monte Carlo[a)]

Amanda L. Danforth[b)]
Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania 16802

Lyle N. Long
Department of Aerospace Engineering and Acoustics, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 4 March 2004; revised 23 June 2004; accepted 28 June 2004)

In the simulation of fluid dynamics, one can either treat the fluid as a continuum or as discrete particles. Although popular for acoustics, the continuum model is limited to small Knudsen numbers (the ratio of mean free path to a length scale). Particle methods are necessary for, but not limited to, problems with Knudsen numbers greater than 0.1, which can occur in shockwaves, microdevices, high frequency sound or rarefied gases. Some well known particle methods include Monte Carlo, cellular automata, discrete velocity, lattice Boltzmann, and molecular dynamics. The direct simulation Monte Carlo (DSMC) method describes gas flows through direct physical modeling of particle motions and collisions. DSMC can model problems for the entire range of Knudsen numbers. In particular, DSMC is capable of simulating nonlinear acoustics, as well as the details of viscous dissipation, dispersion, nonequilibrium effects, and other physical properties. A DSMC method has been implemented for one-dimensional nonlinear acoustics problems on parallel computers using object-oriented C++ and the message passing interface (MPI). DSMC results will be shown and compared with continuum theory and continuum simulations. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1785614]

PACS numbers: 43.25.Cb [RR] Pages: 1948–1955

I. INTRODUCTION

The primary concern of computational nonlinear acoustics is the development and application of numerical models that solve the equations of nonlinear acoustic propagation. In order to create a representative model of the nonlinear system, it is desirable, but often difficult, to incorporate all physical properties of interest. Typical numerical techniques that approximate nonlinear effects have produced encouraging results.[1] However, there is a desire for a more general approach to studying nonlinear acoustics to incorporate all physical effects. The direct simulation Monte Carlo is one possible approach which can include essentially all the physics of interest.

In the numerical study of fluid dynamics, there are two distinct approaches. The first is to consider the fluid as a continuum. In this way, the flow description is characterized by temporal and spatial evolution of the fluid through partial differential equations. Continuum methods use traditional quantities such as ''density'' and ''pressure'' as dependent variables. On the other hand, there is a more general microscopic approach, which investigates the fluid at the particle level. The flow is described by particle position and velocity for all particles at all times. Continuum methods traditionally use the Navier–Stokes or the Euler equations as a mathematical model, and provide a useful description for acoustic wave propagation for many problems. However, the continuum model has its limitations. The macroscopic model assumes deviations from thermal equilibrium are small, and it is the failure of the closure of the Navier–Stokes equations that limits the applications of this approach. Using the Knudsen number (Kn), which is defined as the ratio of mean-free-path to a length scale, as a scaling parameter the continuum regime breaks down for $\mathrm{Kn}>0.1$ approximately.[2]

Particle or nonequilibrium methods are necessary for (but are not limited to) problems with $\mathrm{Kn}>0.1$, which can occur in shockwaves, microdevices, high frequency sound, rarefied gases, and other flows. The mathematical model at this level is the Boltzmann equation:

$$\frac{\partial}{\partial t}(f)+\mathbf{c}\cdot\frac{\partial}{\partial \mathbf{r}}(f)+\mathbf{F}\cdot\frac{\partial}{\partial \mathbf{c}}(f) = \int\int\int_{c}\int_{e^*=0}^{4\pi}(f^*f_1^*-ff_1)g\alpha_1 de^* d\mathbf{c}, \tag{1}$$

where $f(\mathbf{c},\mathbf{r},t)$ is the velocity distribution function in phase space at time t, which specifies the number of molecules that have velocities in the neighborhood of $\mathbf{c}$ and positions in the neighborhood of $\mathbf{r}$ at a given time t. f_1 is a seven-dimensional function with $f_1=f(\mathbf{c_1},\mathbf{r},t)$, and g is the magnitude of relative velocities between particles $|\mathbf{c_1}-\mathbf{c}|$. The term ff_1 accounts for the molecules that leave the neighborhood of $\mathbf{c}$. Similarly, the term $f^*f_1^*$ accounts for the molecules that enter the neighborhood of $\mathbf{c}$. Therefore, the right-hand side of Eq. (1) represents the rate of change in the velocity-distribution function f due to intermolecular colli-

[a)]Portions of this work were presented in ''Acoustic propagation using the direct simulation Monte Carlo,'' 145th Meeting of the Acoustical Society of America, Austin, Texas, November 2003.

[b)]Electronic mail: ald227@psu.edu

sions. **F** represents an external force field, which may be absent in many applications. It can be shown that the Boltzmann equation will reduce to the Navier–Stokes equations for low Kn and the Euler equations as $\mathrm{Kn}\rightarrow 0$.[3]

The Knudsen number is large for sound propagation in very dilute gases, in microchannels, or at high frequencies, and thus requires a particle method (or kinetic theory) solution. There have been numerous attempts to study sound propagation for high Kn based on kinetic theory.[4–12] Most of these attempts use approximations that replace the Boltzmann collision integral. For example, Wang Chang and Uhlenbeck[5] studied sound propagation using the method of eigenfunctions. Numerical procedures applied to the eigenfunction method were performed by Pekeris *et al.*[6] In Sirovich and Thurber, sound propagation was studied using a kinetic model description based on the Bhatnagar–Gross–Krook (BGK) equation,[7,8] while Marques[9] replaced the collision integral with a relaxation-time term. All attempts report that the speed and absorption of sound depend heavily on Kn. Experimental results by Greenspan,[10] Meyer and Sessler,[11] and Schotter[12] show close agreement for the methods above for $\mathrm{Kn}<1$, but for larger Kn, the results are poor.[10–12]

Some other particle methods include cellular automata,[13–15] discrete velocity,[16,17] lattice Boltzmann,[18–22] and molecular dynamics.[23–25] The cellular automata method considers the evolution of idealized particles which move at unit speed from one node to another on a discretized temporal grid. The discrete velocity method allows molecules to move in continuous space, with velocities that are discrete, and nonuniformly distributed. The lattice Boltzmann method was developed from cellular automata and discrete velocity methods, and applies Boltzmann equation approximations to cellular automata models. Particle motion using the lattice Boltzmann method is restricted to a lattice, with simple collision rules to conserve mass and momentum. In molecular dynamics, particle interactions are calculated by force potentials. Molecular dynamics is usually required to simulate dense gases or liquids,[26–28] therefore the expensive intermolecular force calculations make it too expensive for dilute gases. Methods such as cellular automata, discrete velocity, and lattice Boltzmann are usually approximate simulations of fluid dynamics.

The direct simulation Monte Carlo (DSMC) method[2,29,30] is a particle method that was chosen for the current study. DSMC describes gas flows through direct physical modeling of particle motions and collisions, and has been shown to be equivalent to a numerical solution of the Boltzmann equation.[31] The following section discusses in detail the advantages of DSMC over conventional finite difference methods for nonlinear acoustic propagation problems.

A parallel, object oriented DSMC solver was developed for this problem. The code was written in C++ and is designed using the benefits of an object oriented approach.[32–34] The parallelization is achieved by performing parallel ensemble averaging, using the Message Passing Interface (MPI) library[35,36] for interprocessor communications. The reason for implementing a parallel model is to reduce the high computation time, and make the problem more efficient.

II. THE DIRECT SIMULATION MONTE CARLO METHOD

The direct simulation Monte Carlo (DSMC) method was introduced by Bird in the 1960s.[2] The terminology ''direct simulation'' is used because the method directly simulates the physics involved with particle interactions. It is also ''Monte Carlo'' in that particle collisions are determined stochastically. While DSMC does not solve the Boltzmann equation directly, the result of performing a DSMC simulation produces the same result as a Boltzmann equation solution.[31] A Monte Carlo technique, as used here, is a very effective way to solve a multi-dimensional problem.

The method is capable of applications for the entire Kn range, although is most effective at higher Kn. DSMC was originally designed for aerospace applications in dilute gases, and thus has been thoroughly tested for high Kn flows (>0.2). It has been applied to hypersonic and rarefied gas flows,[37–42] and has in fact become the principle method for high Kn simulations. DSMC has also been successfully applied to modeling chemical reactions such as explosions.[43,44] DSMC has also been used to study sound in dilute gases (high Kn).[45] However, there has been limited work in using DSMC for acoustics and low Kn problems (continuum regime),[46,47] and this is one motivation for this work.

DSMC is based on the kinetic theory of gas dynamics,[3] where representative particles are followed as they move and collide with other particles. The movement of particles is determined by their velocities, while the collisions are determined statistically, but are required to satisfy mass, momentum and energy conservation. This treatment of stochastic collisions is the fundamental difference between DSMC and molecular dynamics simulations.[23–25] In molecular dynamics, particle collisions are determined using force potentials such as the Lennard-Jones potential, and often involve many simultaneous interactions. However, DSMC is usually limited to binary collisions. In DSMC the particle positions and velocities are initialized randomly and the boundary conditions ultimately determine the final solution, as in continuum methods.

DSMC simulations become more and more exact as the time step and the cell size tend to zero. If the cell sizes are too large, then the macroscopic gradients are underpredicted and thus give incorrect results, but this method has been proven to converge to the Boltzmann equation in the limit of small time step and small cell size.[31] This is important in that the Boltzmann equation includes all physical properties of interest for sound propagation. Without modification to the method, properties such as viscous dissipation, attenuation, nonlinear effects, dispersion, nonequilibrium effects, entropy waves, and more are available. In fact, DSMC has been able to model more physical phenomena than is usually contained within the Boltzmann equation.

The basic algorithm is illustrated in Fig. 1. There are five main steps in the DSMC algorithm: initialize data, move particles, sort particles, collide particles, and sample cells. The last four steps are repeated in time and also, possibly, over ensembles.

The first step is to initialize all the variables. All ambient quantities are supplied as input parameters, and then all par-

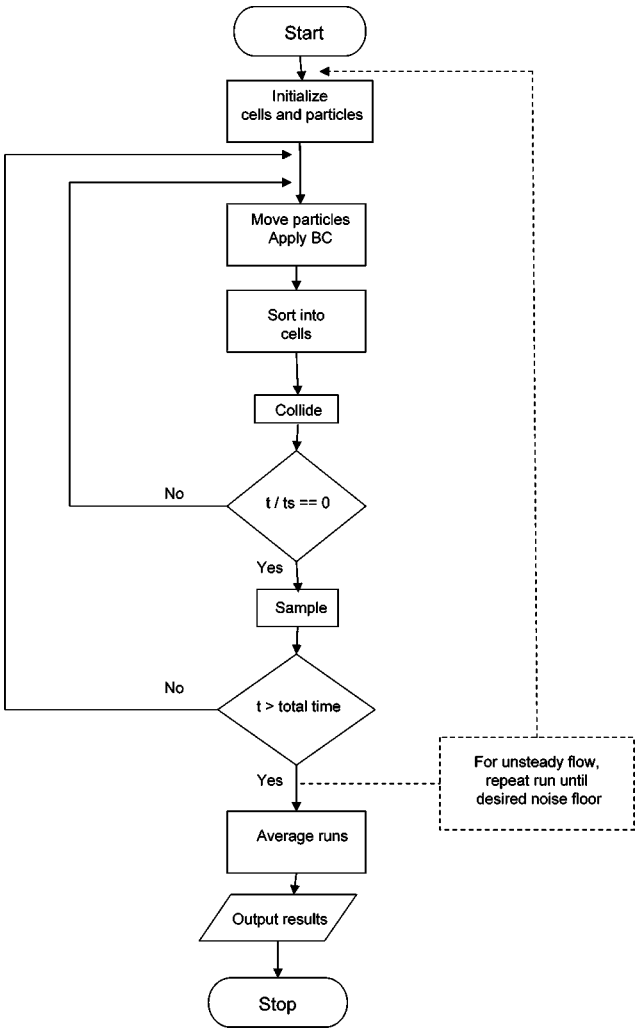


FIG. 1. A flowchart of a typical DSMC simulation.

ticles are given random positions and random Maxwellian velocities. Also, all cells are given particles, the amount depending on the initial conditions. The initial conditions for acoustic wave propagation are given by a density distribution, which is defined by the acoustic source of interest, with initial flow velocity equal to zero. Several different initial conditions have been considered here.

The next step in the algorithm is to move the particles based on their initial random (Maxwellian) velocities for a time Δt. It is assumed that there are no intermolecular collisions during this time, but collisions with boundaries are taken into account. Some boundary conditions that can be included are diffuse boundaries, specular boundaries, periodic boundaries, inflow/outflow boundaries, and moving boundaries.

The particles are then sorted to determine which cell they are in, which is quite easy using C++ pointers. This is a prerequisite step for modeling collisions and sampling.

The next step, and the most unique to DSMC, is the collision routine. The collision process is stochastic and sets DSMC apart from other particle methods. The collision routine starts by picking two particles that are in the same cell at random and then computing their relative velocity. The probability of these two randomly picked particles colliding is related to their relative velocity divided by the maximum velocity in the cell. If this ratio is larger than a randomly generated number, then the two particles collide (like hard spheres), but if the probability is smaller, then two new particles are chosen. Postcollision velocities are then computed assuming random scattering. This process of colliding particles continues until the appropriate number of collisions has been performed.

The sampling of the cells is the next step in the algorithm. This uses the particle mass and velocity components to compute macroscopic quantities such as flow velocity, temperature, and density. These quantities are calculated at the geometric center of each cell and are defined to be the average in that cell. This sampling is *not* performed every time step.

This procedure, of moving, sorting, colliding particles, and sampling cells, is decoupled during each time step. The system loops through all the time steps, while storing the sampled data, until the required amount of time has been completed. Since DSMC is an explicit time marching method, an unsteady flow solution can be produced. To reduce scatter in time-dependent cases, ensemble averaging is necessary, which means repeating the whole process until the noise floor is reduced to a desirable level.

III. COMPUTING ISSUES: OOP AND PARALLEL PROCESSING

An object oriented programming (OOP) approach, written in C++,[32–34] is used in the development of this DSMC study. Such an implementation allows the code to be more readable, manageable, reusable, and extendable. The program is also more efficient by using such C++ characteristics as memory handling and pointers. Using the object oriented C++ concepts of *class structure*, the code is easy to modify and understand. An object oriented technique is a very natural approach for this problem because the particles and cells are physical objects. In addition, OOP allows encapsulation, the grouping together of data and methods (i.e., functions). Using objects, the direct simulation Monte Carlo method flows in a logical sequence and is easy to understand. The responsibility of the detailed calculations is contained in the individual object.

Since the code runs independent simulations and then performs ensemble averages to reduce scatter, it lends itself very nicely to a parallel computing algorithm. Instead of having one processor do all the independent runs, it is easy to have different computers simultaneously run different ensembles, and then average them.[48]

To achieve this simple parallel processing algorithm, a master/slave routine is constructed. The master's main job is to schedule the processors for work, organize collected data, and do the final computing of needed flow properties. The slaves compute the ensemble runs and send sampled data back to the master for final processing. The programming here is made possible by using the Message Passing Interface (MPI)[36] to send messages back and forth between processors. Also, the programming for this type of algorithm is quite simple because there are very few MPI calls.

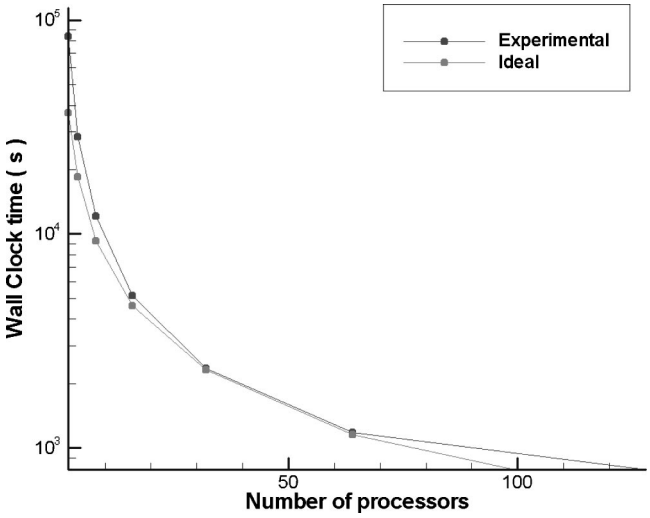


FIG. 2. CPU time on Mufasa with respect to number of processors.

The code was run on a 160-processor Beowulf cluster running Linux. It has a Dolphin network for fast communication and 2.8-GHz Athlon processors. It has a peak speed of 0.4 teraflops (floating point operations per second). In order to illustrate the computational speed of the machine, Fig. 2 shows how many seconds it took to run the program using different numbers of processors. This case uses 1000 cells and 150 particles per cell. Despite the efficiency of the parallel algorithm, the CPU time for these DSMC programs is still fairly large.

IV. CODE VALIDATION AND RESULTS

The parallel DSMC method can be compared to Euler or Navier–Stokes equation solutions, as well as nonlinear acoustic theory.[1,49] The initial conditions of interest in this study include different acoustic sources (such as Gaussian and sinusoidal) of varying width or frequency.

The linear one-dimensional Euler equations, Eqs. (2) and (3), can be combined to yield the wave equation, where a_0 is the speed of sound. Results from a DSMC simulation of a Gaussian pulse compared to a second order accurate in space and time finite difference simulation of the wave equation are shown in Fig. 3. The wave equation is solved using a second order accurate finite difference scheme on a uniform grid. The simulation is of argon at standard atmospheric conditions (ambient pressure $=111\,285$ Pa, temperature $=300$ K). The Gaussian pulse is 100 cells wide with each cell being $\frac{1}{2}$ of a mean-free-path wide. The mean-free-path in these conditions is approximately 50 nm. The domain is about 2000 cells long, which is equivalent to 50 μ. Each cell has about 150 particles. This means that roughly 300 000 particles are followed throughout the simulation. There were 8000 time steps, each of size 7.5 ps, meaning the elapsed time is about 60 ns. The acoustic density amplitude of the Gaussian source is 20% of the ambient density:

$$\frac{\partial \rho'}{\partial t} + \rho_0 \frac{\partial u'}{\partial x} = 0, \tag{2}$$

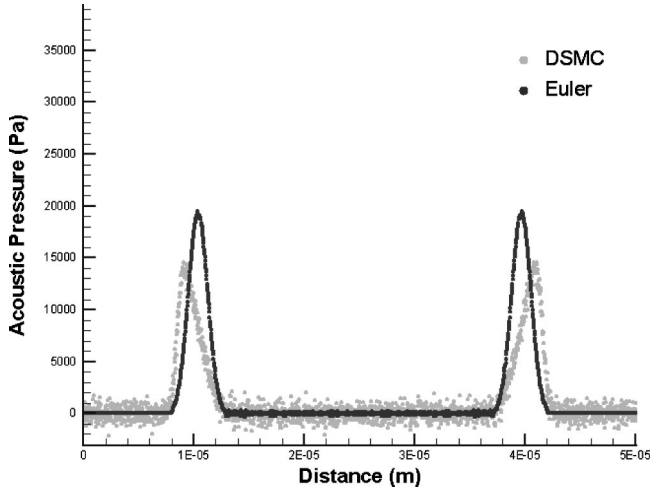


FIG. 3. Comparison of the DSMC results (light) against the linear Euler equations analytical solution (dark) for a Gaussian source at time $t = 2.1$ ns.

$$\rho_0 \frac{\partial u'}{\partial t} + \frac{\partial p'}{\partial x} = 0, \quad p' = a_0^2 \rho'. \tag{3}$$

It is clear in Fig. 3 that the DSMC solution exhibits nonlinear wave steepening, while the wave equation solution does not. Due to the large amplitudes, it is expected that the DSMC wave will steepen due to nonlinear effects. Since the wave equation or linear Euler equations are not capable of handling these effects, they will not predict the correct behavior of the system. Thus the differences between the solutions, shown in Fig. 3, are expected.

Solutions to the nonlinear Euler equations [Eq. (4)] were also computed using a finite difference scheme run in Matlab. The solutions were one-dimensional in space, and the algorithm used was the McCormack time accurate algorithm.[50] Second order artificial viscosity was included to reduce numerical oscillations. A comparison with DSMC and the nonlinear Euler equations given by Eq. (4) is shown in Fig. 4 with the same initial conditions. It is clear that the comparison in this case gives excellent results.

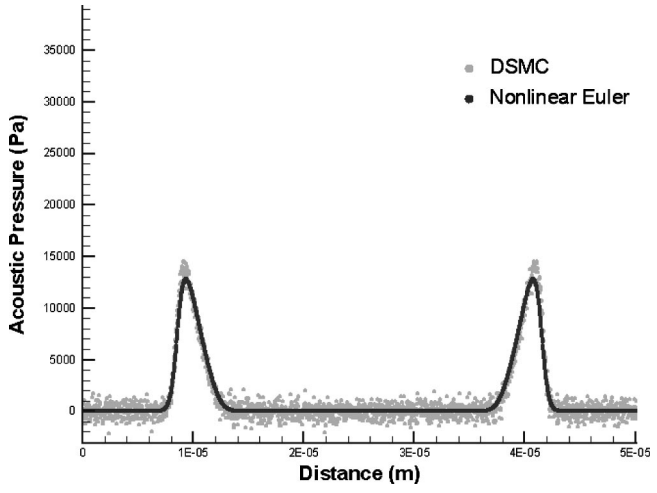


FIG. 4. Comparison of the DSMC results (light) against the nonlinear Euler equations analytical solution (dark) for a Gaussian source at time $t = 2.1$ ns.

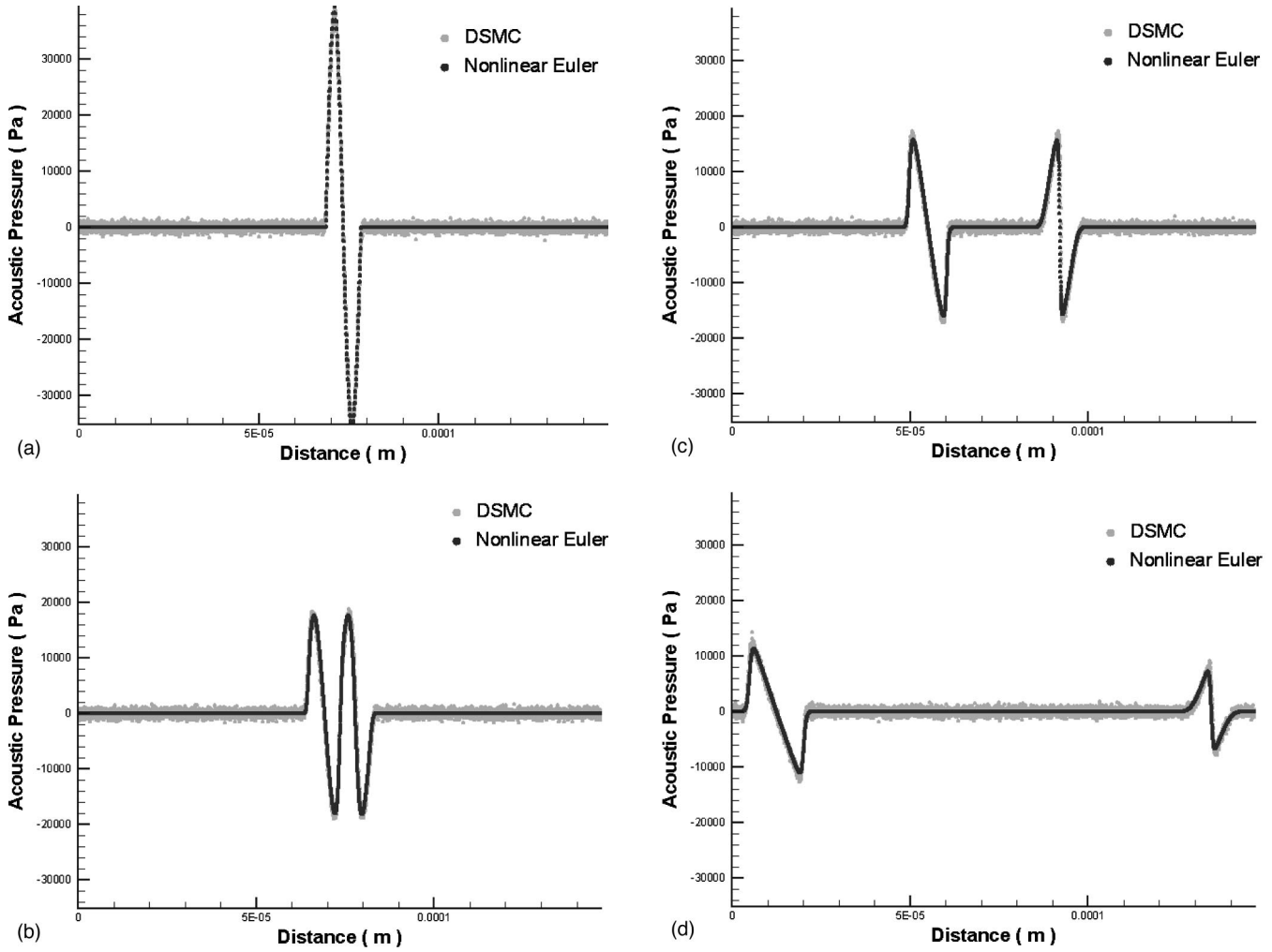


FIG. 5. Comparison of DSMC results (light) against the nonlinear Euler equations analytical solution (dark) for a sinusoidal source of frequency 30 MHz for various times. (a) Time $t=0$ ns. (b) Time $t=0.225$ ns. (c) Time $t=0.75$ ns. (d) Time $t=3$ ns.

$$\frac{\partial \rho}{\partial t}+\frac{\partial \rho u}{\partial x}=0, \quad \frac{\partial \rho u}{\partial t}+\frac{\partial \rho u^2}{\partial x}=-\frac{\partial p}{\partial x}, \tag{4}$$

$$\frac{\partial \rho E}{\partial t}+\frac{\partial \rho (E+p) u}{\partial x}=0, \quad p=\rho R T.$$

Sinusoidal sources of varying frequency were also considered. A sinusoidal density disturbance of one period was created in the middle of the domain, and allowed to propagate in both directions, with initial flow velocity equal to zero. Figures 5(a)–5(d) show the evolution of such a wave. In this case, there are 400 cells per wavelength, again with each cell being $\frac{1}{2}$ of a mean-free-path. This corresponds to a frequency of 30 MHz. The domain is 6000 cells long, which is equivalent to 150 μs, again with 150 particles per cell. There were 25 000 time steps, each of time step size of 7.5 ps, resulting in an elapsed time of 0.2 μs. The initial acoustic density amplitude is 20% of the ambient density. Due to the high amplitude and frequency, it is expected that nonlinear effects such as wave steepening and nonlinear absorption occur quite quickly in the simulation. The shock distance for a sinusoidal source is given by Eq. (5).[1,49]

$$\bar{x}=\frac{\rho_0 a_0^2}{\beta k P_0}, \tag{5}$$

where ρ_0 is the ambient density, a_0 is the speed of sound, β is the coefficient of nonlinearity, k is the acoustic wave number, and P_0 is the initial acoustic pressure amplitude. To nondimensionalize the spatial scale, define the shock inception distance:

$$\sigma=\frac{x}{\bar{x}}, \tag{6}$$

which implies that the wave initially shocks at the value $\sigma=1$. Because of the nature of the waveforms of the right and left going waves, the two evolve differently, as noticed in Fig. 5(c), with the appropriate old age effects noticed in Fig. 5(d). Excellent agreement between both the wave speed and the amplitude is shown in Fig. 5.

Another nonlinear acoustic effect of interest is the generation of harmonics due to the nonlinear wave steepening. To further validate the code, a harmonic generation analysis of this problem was also performed. Figure 6(a) shows the FFT result from the initial sinusoidal source, where Fig. 6(b) shows the FFT result when the sine wave has reached shock

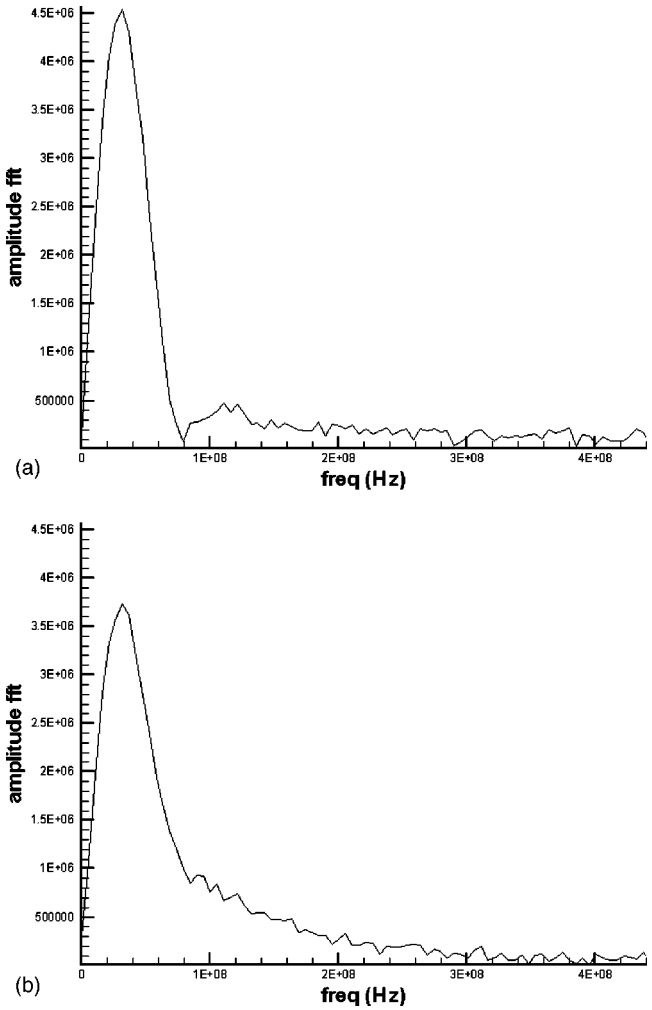


FIG. 6. The frequency shirt to higher harmonics. Part (a) initial source, part (b) at shock inception distance, sigma=1.

inception distance. As the wave steepens, the amplitude of the fundamental reduces while the amplitudes of the higher harmonics get larger, and thus the shift in the spectrum is evident.

To investigate the spectrum shift in further detail, the Fourier component amplitudes of the first four harmonics were compared to the analytical Blackstock Bridging function solution[1] as a function of propagation distance which is shown in Fig. 7. It is clear that DSMC results compare very well with the Blackstock Bridging function. The second, third, and fourth harmonic component amplitudes continue to increase until about $\sigma=1.5$, where the absorption becomes the dominant physical property, and all harmonics dissipate. This is in agreement with nonlinear acoustic theory where shock growth peaks at $\sigma=\pi/2$.[1]

Since the Euler equations are only valid as $\mathrm{Kn}\rightarrow 0$, and DSMC is able to simulate high Kn numbers, it is expected that there will be significant differences in the results between DSMC and the nonlinear Euler equation solution for finite Kn. Figure 8 shows the differences in the solutions for high Kn=0.16. Similar differences would be observed between DSMC and the Navier–Stokes equations, since the Navier–Stokes equations are also not valid at large Kn numbers.

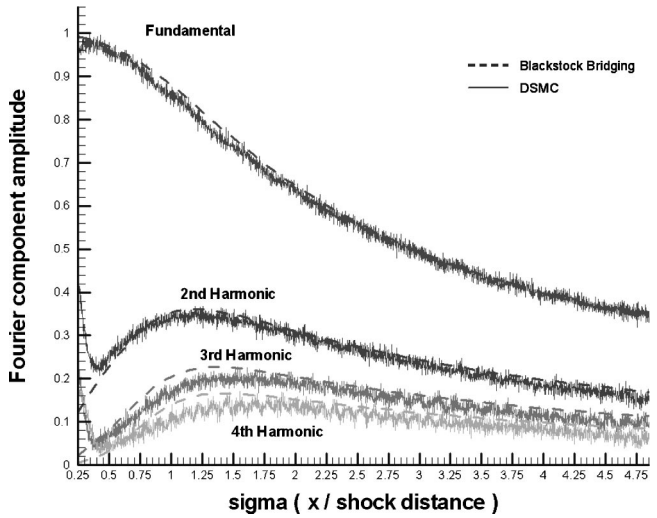


FIG. 7. Fourier component amplitude for $0\leqslant\sigma\leqslant 5$ with DSMC solution (solid) compared to the Blackstock bridging function (dashed) of the fundamental, second, third, and fourth harmonics.

Experimental and other kinetic theory results suggest that the absorption of sound for high Kn differs significantly than for low Kn.[4–7,9–12] The scaled absorption coefficient, α/k_0, where α is the absorption coefficient and k_0 is the acoustic wave number, for Kn ranging from 0.04 to 35 is shown in Fig. 9. DSMC simulation results are compared to a classical Navier–Stokes prediction, experimental work by Greenspan,[10] and previous kinetic theory results by Hadjiconstantinou and Garcia[45] whose work was based on standing waves. The variation in Kn was obtained by keeping the cells per wavelength constant at 100, but varying the cell size from $\frac{1}{2}$ of a mean-free-path to $\frac{3}{1000}$ of a mean-free-path. The absorption coefficient results were obtained by finding the maximum pressure at each time step and calculating a line of best fit for on a logarithmic scale.

For high Kn simulations, the wavelength of the signal becomes significantly shorter than a mean-free-path. Experimental results for $\mathrm{Kn}>5.5$ are thus difficult to obtain because of scattering and high absorption.[10] The simulations presented here have successfully captured the physics for

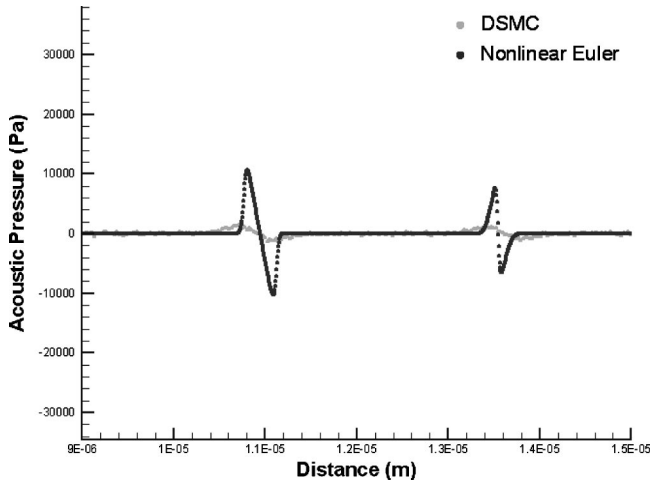


FIG. 8. Comparison of Euler and DSMC for high Kn=0.16.

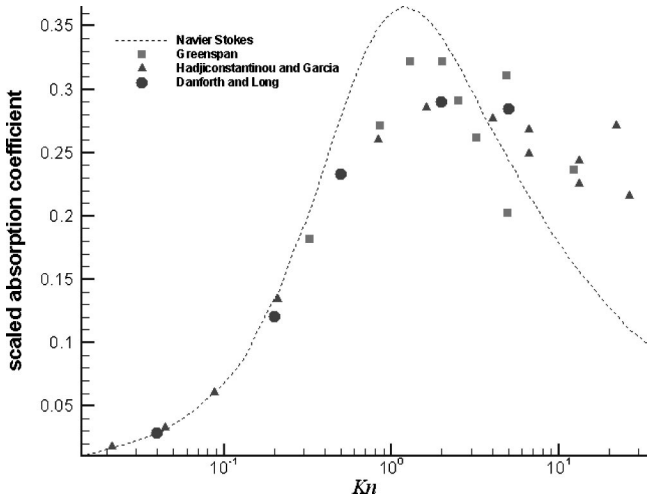


FIG. 9. Scaled absorption coefficient for varying Kn ranging from 0.04 to 35.

such cases by decreasing our cell size to $\frac{3}{1000}$ of a mean-free-path.

The calculations presented here confirm that the Navier–Stokes prediction for the absorption coefficient is only correct for low Kn, and a kinetic theory approach must be used for high Kn calculations. The absorption coefficient results are in good agreement with experimental results and previous work on standing waves.

V. CONCLUSIONS

The motivation for this study was twofold. First, it was desired to create a numerical method for nonlinear acoustics that included properties such as viscous dissipation, attenuation, nonlinear effects, dispersion, thermal nonequilibrium effects, and entropy waves. Second, the expansion of the applicability of the particle model, the direct simulation Monte Carlo (DSMC), to small Kn was desired. An introduction to the robustness of the DSMC method was presented by investigating its applications, ranging from rarefied gas simulations to explosions, that span the entire Knudsen number range.

The current model uses DSMC to study nonlinear acoustic propagation. This model was implemented using object oriented programming in C++ which makes the entire code easier to read, maintain, and expand. The code also uses a parallel processing algorithm to further improve the efficiency of the model. Using DSMC for the study of nonlinear acoustics was validated by comparisons to the nonlinear Euler equations and by studying nonlinear harmonic generation. Sound propagation for high Kn is discussed where the absorption of sound is higher than a Navier–Stokes or Euler prediction. Good agreement is found between experiment and previous work. Results indicate that DSMC is a suitable technique for modeling sound propagation for all values of Kn.[51]

Despite the efficiency of the parallel algorithm, the CPU time for using DSMC can be large compared to that of a continuum model. Also, since the size of our domain is on the order of microns, the simulations are limited to small scale simulations at atmospheric conditions. For larger scale simulations, the size of the problem may easily become larger than the available memory of a machine. Since the parallel algorithm is designed to simulate different ensembles on different processors, memory constraints for using this algorithm could become an issue. Thus, an alternative to the parallel algorithm described here might need to be sought such as that described by Sharma and Long.[47]

In the near future it is planned to perform nonlinear acoustics simulations in multiple dimensions as well as implementing a diatomic collision routine, both of which are straightforward to implement.

ACKNOWLEDGMENTS

We would like to thank the National Science Foundation for funding the Consortium for Education in Many Body Applications, CEMBA (http://www.cemba.psu.edu), Grant No. NSF-DGE-9987589.

[1] M. F. Hamilton and D. T. Blackstock, *Nonlinear Acoustics* (Academic, San Diego, 1998).
[2] G. A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*, 2nd ed (Clarendon, Oxford, 1994).
[3] S. Chapman and T. G. Cowling, *Mathematical Theory of Non-Uniform Gases* (Cambridge U.P., Cambridge, 1939).
[4] M. Greenspan, ''Transmission of sound waves in gases at very low pressures,'' in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1965), Vol. II, Pt. A, Chap. 1, pp. 1–45.
[5] C. S. Wang Chang and G. E. Uhlenbeck, in *Studies in Statistical Mechanics*, edited by J. de Boer and G. E. Uhlenbeck (North-Holland, Amsterdam, 1970), Vol. V, pp. 43–75.
[6] C. L. Pekeris, Z. Alterman, L. Finkelstein, and K. Frankowski, ''Sound propagation in a gas of rigid spheres,'' Phys. Fluids **5**, 1608–1616 (1962).
[7] L. Sirovich and J. K. Thurber, ''Propagation of forced sound waves in rarefied gasdynamics,'' J. Acoust. Soc. Am. **37**, 329 (1965).
[8] P. L. Bhatnagar, E. P. Gross, and M. Krook, ''A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems,'' Phys. Rev. **94**, 511–525 (1954).
[9] W. Marques, ''Dispersion and absorption of sound in monatomic gases: An extended kinetic description,'' J. Acoust. Soc. Am. **106**, 3282–3288 (1999).
[10] M. Greenspan, ''Propagation of sound in five monoatomic gases,'' J. Acoust. Soc. Am. **28**, 644 (1956).
[11] E. Meyer and G. Sessler, Z. Phys. **149**, 15–39 (1957).
[12] R. Schotter, ''Rarefied gas acoustics in the noble gases,'' Phys. Fluids **17**, 1163–1168 (1974).
[13] W. Aspray, *John von Neumann and the Origins of Modern Computing* (MIT, Cambridge, 1990).
[14] L. N. Long, R. E. Coopersmith, and B. G. McLachlan; ''Cellular automatons applied to gas dynamic problems,'' in *Proceedings of the AIAA 19th Fluid Dynamics, Plasma Dynamics, and Lasers Conference*, AIAA Paper No. 87-1384, Honolulu, HI (1987).
[15] U. Frisch, B. Hasslacher, and Y. Pomeau, ''A lattice gas automaton for the Navier-Stokes equations,'' Los Alamos Report No. LA-UF-85-3503, September 1985.
[16] J. E. Broadwell, ''Study of rarefied shear flow by the discrete velocity method,'' J. Fluid Mech. **19**, 401–414 (1964).
[17] J. E. Broadwell, ''Shock structure in a simple discrete velocity flow,'' Phys. Fluids **7**, 1243 (1964).
[18] S. Chen and G. D. Doolen, ''Lattice Boltzmann method for fluid flows,'' Annu. Rev. Fluid Mech. **30**, 329 (1998).
[19] R. Benzi, S. Succi, and M. Vergassola, ''The lattice Boltzmann equation: theory and applications,'' Phys. Rep. **222**, 146 (1992).
[20] J. M. Buick, C. L. Buckley, C. A. Greated, and J. Gilbert, ''Lattice Boltzmann BGK simulation of nonlinear sound waves: the development of a shock front,'' J. Phys. A **33**, 3917–3928 (2000).
[21] Y. Sudo and V. W. Sparrow, ''A new lattice gas model for 1-D sand propagation,'' J. Comput. Acoust. **1**, 423 (1994).
[22] Y. Sudo and V. W. Sparrow, ''A new lattice gas model for 1-D sound propagation,'' J. Comput. Acoust. **3**, 69 (1995).

[23] D. C. Rapaport, *The Art of Molecular Dynamics Simulation* (Cambridge U.P., Cambridge, 1995).
[24] M. P. Allen, and D. J. Tilldesley, *Computer Simulation of Liquids* (Oxford U.P., New York, 1987).
[25] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, San Diego, 1996).
[26] T. L. Kaltz, L. N. Long, and M. M. Micci, ''Supercritical vaporization of liquid oxygen droplets using Molecular Dynamics,'' Combust. Sci. Technol. **136**, 279–301 (1998).
[27] C. Nwobi, L. N. Long, and M. M. Micci, ''Molecular dynamics studies of properties of supercritical fluids,'' J. Thermophys. Heat Transfer **12**, 351–354 (1998).
[28] Kantor, L. N. Long, and M. M. Micci, ''Molecular Dynamics simulation of dissociation kinetics,'' J. Thermophys. Heat Transfer **15**, (2001).
[29] F. J. Alexander and A. L. Garcia, ''The direct simulation Monte Carlo method,'' Comput. Phys. **11**, 588 (1997).
[30] E. S. Oran, C. K. Oh, and B. Z. Cybyk, ''Direct simulation Monte Carlo: Recent advances and applications,'' Annu. Rev. Fluid Mech. **30**, 403–440 (1998).
[31] W. Wagner, ''A convergence proof for Bird's direct simulation Monte Carlo method for the Boltzmann equation,'' J. Stat. Phys. **66**, 1011 (1992).
[32] V. Shtern, *Core C++* (Prentice Hall, Upper Saddle River, NJ, 2000).
[33] D. Yang, *C++ and Object-oriented Numeric Computing* (Springer, New York, 2001).
[34] P. S. Wang, *Standard C++ with Object-oriented Programming* (Brooks/Cole, Pacific Grove, 2001).
[35] W. Gropp, E. Lusk, and A. Skjellum, *Using MPI: Portable Parallel Programming with Message Passing Interface*, 2nd ed. (MIT, Cambridge, 1999).
[36] P. S. Pacheco, *Parallel Programming with MPI* (Morgan Kaufmann, San Francisco, 1997).
[37] G. A. Bird, ''Monte Carlo simulation of gas flows,'' Annu. Rev. Fluid Mech. **10**, 11–31 (1978).
[38] G. A. Bird, ''Direct simulation of the Boltzmann equation,'' Phys. Fluids **13**, 2676 (1970).
[39] E. P. Muntz, ''Rarefied gas dynamics,'' Annu. Rev. Fluid Mech. **21**, 387–417 (1989).
[40] E. Salomons and M. Mareschal, ''Usefulness of the burnett description of strong shock waves,'' Phys. Rev. Lett. **69**, 269–272 (1992).
[41] D. I. Pullin, J. Davis, and J. K. Harvey, ''Monte Carlo calculations of the rarefied transition flow past a bluff faced cylinder,'' *10th International Symposium on Rarefied Gas Dynamics*, Aspen, CO (1976).
[42] L. N. Long, ''Navier-Stokes and Monte Carlo results for hypersonic flows,'' AIAA J. **29**, (1991).
[43] L. N. Long and J. B. Anderson, ''The simulation of detonations using a Monte Carlo Method,'' *22rd International Symposium on Rarefied Gas Dynamics*, Sydney, Australia (2000).
[44] J. B. Anderson and L. N. Long, ''Direct Monte Carlo simulation of chemical reaction systems: Prediction of ultra fast detonations,'' J. Chem. Phys. **118**, 3102–3110 (2003).
[45] N. G. Hadjiconstantinou and A. L. Garcia, ''Molecular simulations of sound wave propagation in simple gases,'' Phys. Fluids **13**, 1040–1046 (2001).
[46] D. I. Pullin, ''Direct simulation methods for compressible inviscid idea-gas flow,'' J. Comput. Phys. **34**, 231–244 (1980).
[47] A. Sharma and L. N. Long, ''Numerical simulation of the blast impact problem using the direct simulation Monte Carlo (DSMC) method,'' to appear in J. Comput. Phys. (2004).
[48] L. N. Long and K. S. Brentner, ''Self-scheduling parallel methods for multiple serial codes with application to WOPWOP,'' in *Proceedings of the 38th Aerospace Sciences and Meeting Exhibit*, AIAA Paper 2000-0346, Reno, NV (2000).
[49] A. D. Pierce, *Acoustics, an Introduction to Its Physical Principles and Applications* (McGraw–Hill, New York, 1981).
[50] J. C. Tannehill, D. A. Anderson, and R. H. Pletcher, *Computational Fluid Mechanics and Heat Transfer* (Taylor and Francis, Philadelphia, 1997).
[51] A. L. Danforth, ''Nonlinear acoustic simulations using an object-oriented, parallel direct simulation Monte Carlo method,'' M.S. thesis, the Pennsylvania State University, August 2004.

Microparticle concentration in short path length ultrasonic resonators: Roles of radiation pressure and acoustic streaming

Larisa A. Kuznetsova[a] and W. Terence Coakley
School of Biosciences, Cardiff University, Park Place, Cardiff CF10 3TL, United Kingdom

(Received 12 February 2004; revised 5 July 2004; accepted 7 July 2004)

Acoustic streaming in ultrasonic (1.4–3.0 MHz) circular and rectangular resonators of path length approximately one-half or one quarter wavelength (λ) has been characterized by particle image velocimetry (PIV) using fluorescent 1 μm diam latex markers. Particles of all diameters examined (1, 24, 80 μm) moved into pressure node planes within 4 s of initiation of sonication. The larger particles then moved within that plane to one or more preferred positions. 1 μm particles in a λ/2 cylindrical resonator with a single nodal concentration region for larger particles were convected by Rayleigh-type streaming from the center of the node plane to its edge. In contrast, particles concentrated at many loci in two planes of a second cylindrical and a rectangular chamber. Small scale wall-associated Rayleigh-type vortices occurred in a λ/4 chamber. More unexpectedly, wall-independent bulk suspension vortices, with circulation planes parallel to the transducer radiating surface, were recorded in both resonators. Tracer particles experienced radial forces that drove them towards or away from the center of the vortices to be concentrated at its center or entrained in a vortex perimeter ring. These different outcomes are discussed in terms of lateral radiation force distribution in the node planes.
[DOI: 10.1121/1.1785831]

PACS numbers: 43.25.Nm, 43.25.Gf, 43.25.Qp, 43.35.Bf [YHB] Pages: 1956–1966

I. INTRODUCTION

Several forces act on microparticles suspended in ultrasonic standing waves (USWs). The strongest is the axial component of the direct acoustic radiation force (DRF_a),[1] which drives particles towards, and holds them at, the pressure node or anti-node planes (depending on the acoustic properties of the suspension), where the potential energy of the particles is a minimum. The force is given by

$$\mathrm{DRF}_a = -\pi P_0^2 V_p \beta_0 / (2\lambda) \cdot \Phi(\beta, \rho) \cdot \sin(4\pi z/\lambda). \quad (1)$$

In Eq. (1) P_0 is the acoustic pressure amplitude, β_0 is the compressibility of fluid, λ is the wavelength of ultrasound in the suspending phase, V_p is the particle volume, and variable z is distance from a pressure node.

The acoustic contrast factor, Φ, of the suspended particles is

$$\Phi = \frac{5\rho_p - 2\rho_0}{2\rho_p + \rho_0} - \frac{\beta_p}{\beta_0}, \quad (2)$$

where ρ_p and ρ_0 are the densities of the particles and fluid, respectively, and β_p is the compressibility of the particles. The lateral components of direct radiation force, which are reported to be about two orders of magnitude smaller,[2,3] act within the planes and contribute to the lateral concentration into clumps.

Ultrasonic resonators may be fabricated so that the depth of the chamber (acoustic path length in the resonator) is equal to either λ/2 or λ/4 with associated reflector thicknesses of λ/4 or λ/2, respectively. The acoustic pressure amplitude at the glass–air surface is in each case zero (pressure release condition).[4,5] The only acoustic pressure node in the fluid phase of each resonator occurs one half wavelength away from the glass–air surface, i.e., in the center of the chamber or at the surface of the reflector, for the λ/2 and λ/4 fluid phases, respectively. In idealized cases, when it is assumed that the sound source radiates a plane longitudinal wave, suspended particles in a still medium are expected to be driven to those nodal positions, and to concentrate in a monolayer, as it is shown in the final details of Figs. 1(a) and (b). We note that z in Eq. (1) is defined as being zero at the fluid phase pressure node. This distance is λ/4 from the coupling steel layer surface in both the half wavelength and quarter wavelength bulk phase resonators (Fig. 1).

A number of models have been developed for the calculation of one-dimensional acoustic pressure profiles in a standing wave chamber.[6–9] In practice however there may be nonuniformity in the pressure distribution in a chamber and as a result, the distribution of particles may present a more complex pattern.

In addition to the direct radiation force, particles in a chamber can be influenced by several types of acoustically driven flows—streaming—which differ in their origin and spatial scale. Acoustic streaming is nonoscillatory steady fluid motion originating from the spatial nonuniformity of the sound field or from energy dissipation at the liquid–solid interfaces of the container. The following classification of streaming proves useful:[10]

(i) Eckart streaming, which results from the attenuation of a traveling wave in the bulk fluid;
(ii) Schlichting micro-streaming which is a vortex flow inside the thin boundary layer at a solid–liquid interface;

[a]Electronic mail: kuznetsoval@cardiff.ac.uk

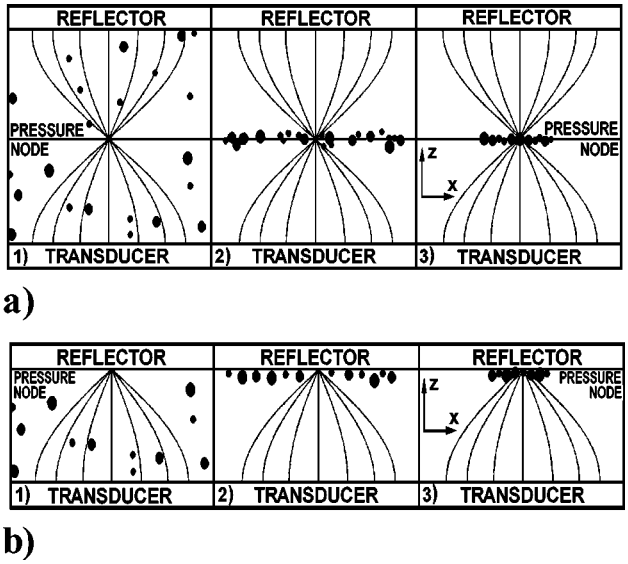


FIG. 1. Effect of the direct acoustic radiation force (DRF) on suspended particles in a resonator with (a) λ/2 water path length depth (λ/4 reflector thickness), (b) λ/4 water path length depth (λ/2 reflector thickness): (1) Distribution of particles before ultrasound application, (2) axial DRF effect, (3) lateral DRF effect; *x*—lateral direction, *z*—direction of sound propagation.

(iii) Rayleigh streaming which is a vortex flow outside the boundary layer.

Rayleigh carried out the first theoretical study of streaming when he treated the case of standing waves between parallel walls.[11] Since then other comprehensive reviews on acoustically induced flow have been published.[10,12–16] Extensive experimental work on acoustic streaming has become possible due to the development of optical techniques of flow visualization—particle image velocimetry (PIV)[17] and Laser Doppler Anemometry (LDA).[18] Eckart- and Rayleigh-type streamings induced in air and water mediums in relatively big containers (1–50 cm) have been studied.[19–22] The PIV technique has also been applied to the study of Rayleigh streaming in a chamber with sub-millimetre acoustic pathlength.[23]

Interest in acoustic streaming has increased over the last decade as its applications in various processes have been reported. These include acoustic mixing,[24] enhancement of heat and mass transfer[25,26] and metal electrodeposition.[27] Streaming can also be utilized in medical diagnostics.[28,29] Induction of acoustic streaming among other approaches has been explored to noninvasively effect rapid convective movement in microfluidic devices. Microfluidics has emerged as an important research topic in fluid mechanics in recent years.[30] Microfluidic devices (dimensions ranging from millimetres to micrometres with the capability to handle volumes of fluids in the range of nano- to micro-liters[31]) usually include a sensing surface in contact with a sample flowing at low Reynolds number through a test channel. The acoustic devices included a 3.5 MHz bulk acoustic standing wave for particle and molecular mixing,[32] a 50 MHz[33] or a loosely focused 240 MHz longitudinal travelling wave to induce Eckart streaming.[34] Liu *et al.*[35] employed microstreaming induced by trapped bubbles to mix samples and enhance DNA-surface kinetics on a polymerase chain reaction biochip. In additional to longitudinal wave flexural plate waves (FPW) devices have been employed to enhance biosensor antibody-antigen detection.[36] The FPW-induced micro-streaming has been measured, alone among the different systems developed, by PIV analysis.[37]

Short path length ultrasonic resonators are also of interest due to their ability to manipulate micro-particles and cells in small volumes for analytical purposes. A situation has been described for a sub-wavelength chamber where Rayleigh-type streaming is dominant for small (1 μm) particles whereas DRF determines the movement of 25 μm particles.[23] The chamber employed had a single particle potential energy well. Chambers with different field geometry, rectangular and cylindrical, may have multiple regions of particle aggregation. The present work sets out to investigate the influence of radiation pressure and induced streaming on particle behavior in such short path length systems. It is shown that both acoustic streaming and the direct radiation force play important roles in particle movements and concentration.

II. EXPERIMENT

A. Chamber construction and drive

The chambers—cylindrical and rectangular—employed in this work are based on three main features: An ultrasonic transducer, a spacer which defines the depth of the channel and a glass reflector. A coupling stainless steel layer (1.5 mm thick for the rectangular chamber and 1 and 3 mm thick for the fixed and variable depth cylindrical chambers, respectively) to which the transducer is glued separates the latter from the water layer, as shown in Fig. 2(a).

The thickness of the 12 mm diam transducer disc in the cylindrical resonators was 1.33 mm so that the transducer's nominal resonance frequency was about 1.5 MHz. The transducer's back electrode was etched[38] to produce an active radiation area of 8 mm in diameter. The fixed channel depth of one cylindrical chamber was 0.5 mm, which is equivalent to λ/2 in water at 1.5 MHz [Fig. 2(b)]. The screw-top and elastic O-ring gasket of the second cylindrical chamber [Fig. 2(c)] allowed continuous variation of the channel depth from 0.5 (λ/2) to 0.25 mm (λ/4); 1 or 2 mm thick glass reflectors were used for the two extreme cases, respectively.[7,39] The channel depth was measured in air by focusing the microscope objective on the inner surface of the reflector and then on the base of the chamber.

The transducer of a rectangular resonator [Fig. 2(d)] had a nominal thickness resonance frequency of 3 MHz. The active radiation area was 20×10 mm^2. The area of the channel was 51×10 mm^2. Brass spacers of 125 and 250 μm thick enabled the channel depth to be varied. Standard microscope glass slides (1 mm thickness) were used as reflectors. Details of the construction of the chamber have previously been described.[40]

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 scans the voltage-

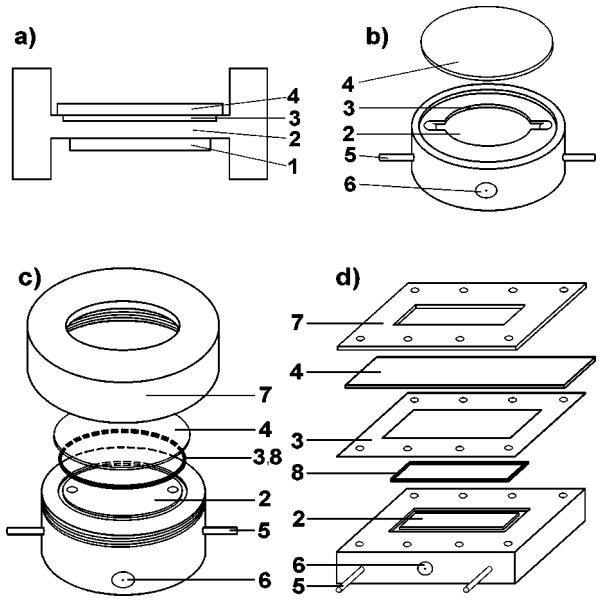


FIG. 2. Design of resonators: (a)—general multilayer structure, (b)—cylindrical resonator with fixed water layer thickness, (c)—cylindrical resonator with variable water layer thickness, (d)—rectangular resonator; 1—piezoelectric transducer, 2—coupling steel layer, 3—spacer for water layer, 4—glass reflector (1 mm thick for $\lambda/2$ path length and 2 mm thick for $\lambda/4$ path length cylindrical resonators), 5—water inlet (outlet), 6—electrical connection, 7—screw-top (c) or brass plate (d), 8—O-ring (c) or elastic gasket (d).

frequency spectra from a voltmeter and allows the frequency to be held at voltage maximum.[41] A manually controlled function generator (HM 8130, HAMEG Instruments) has also been used for the generation of ultrasound. It allowed either the maintenance of fixed ultrasonic conditions or the gradual variation of the parameters of the sound field during the experiment.

B. Suspensions

Suspensions of polystyrene plain 24 and 80 μm diam and fluorescent 1 μm latex beads in deionized water (Polysciences) were diluted with deionized water to the concentrations required.

C. Flow analysis

The observation of particle movement was carried out with an Olympus BX41M epifluorescent microscope. A Fujitsu video camera was connected via a 0.5 microscope adaptor and the images were recorded onto a standard video tape. The recorded video sequences were transferred to a PC video card in digital format with hardware MJPEG data compression (Pinnacle Miro Video DC30+) and then further processed with PIV software.

Two software packages were used for the estimation of velocity fields: PIV SLEUTH[42] and FLOWMANAGER developed by Dantec Dynamics (Denmark).

D. Acoustic pressure amplitude measurement

The acoustic pressure amplitude was estimated from the balance of the direct radiation and gravitational forces acting on a particle in suspension.[43] The levitation of a single 24 micron latex sphere was observed with a microscope. The transducer's voltage, which is linearly proportional to acoustic pressure, was gradually decreased to a threshold value after which the buoyancy-corrected gravitational force was no longer balanced by the axial direct radiation force [Eq. (1)], and the particle sedimented. Linear extrapolation of the pressure at the threshold voltage to the voltage applied in an experiment gave the acoustic pressure amplitude at the different experimental conditions.

E. Admittance measurements

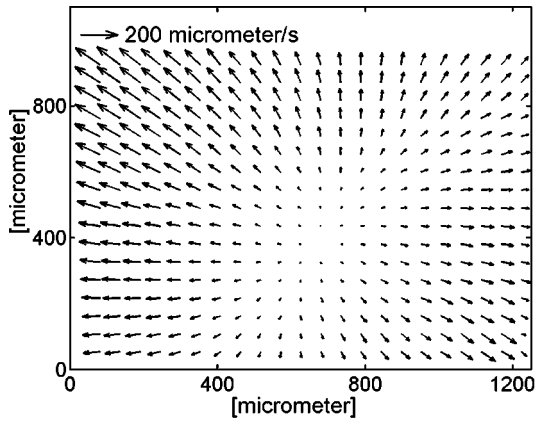
Electrical admittance of the chambers was measured using an Agilent Technologies Network Analyzer (HP 8753E).

III. RESULTS

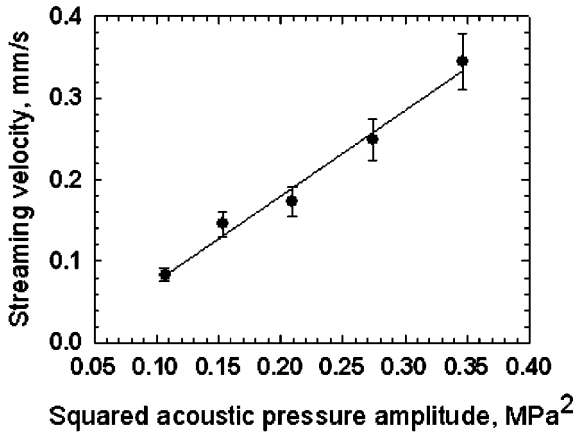
A. Cylindrical 1.5 MHz chamber with fixed channel depth

At the channel depth of 0.5 mm ($\lambda/2$) with a quarter wavelength reflector, the cylindrical resonator used in this work [Fig. 2(b)] had one central pressure node plane as illustrated schematically in Fig. 1(a). Frequency-voltage spectra showed a resonance at ~1.38 MHz. The Hameg wave generator provided this drive frequency in further experiments. When ultrasound was applied, 24 μm latex particles, driven by the axial and lateral direct radiation forces, concentrated in a single clump near the center of the chamber in the only pressure node [Fig. 1(a) and part 3].

The behavior of small (1 μm) latex beads was different and was essentially as described by Spengler *et al.*[23] for a rectangular transducer with a circular back electrode. The difference arises because the smaller beads in the acoustic field are affected by streaming to a greater degree than bigger particles.[23] When ultrasound was applied to the cylindrical resonator the small particles, driven by the axial direct radiation force, came to the pressure node plane at which the microscope was pre-focused within 200 ms. They then moved radially from the central point towards the periphery of the chamber. After a few seconds the particles concentrated at the edge of the active radiation area resulting in the formation of a particle-free circle of about 6 mm in diameter. The particles did not re-enter this space. The behavior is consistent with the suggestion[23] that, at first following the streaming in the pressure node plane, the particles reached the area where, according to the Rayleigh theory, the velocity of streaming in lateral direction was negligible, and were held there in the nodal plane by the axial direct radiation force. Measurements of streaming velocity were made at different acoustic pressures. The velocity of the particles in the pressure node plane increased with the increase of acoustic pressure in the system. Figure 3(a) displays the velocity field in the pressure node plane about the center after the beginning of sonication for the acoustic pressure amplitude of 520 kPa. In Fig. 3(b) streaming velocity is presented as a function of acoustic pressure amplitude squared. These Rayleigh type vortices (whose planes of circulation are perpendicular to the transducer plane) and their scales of circulation will be

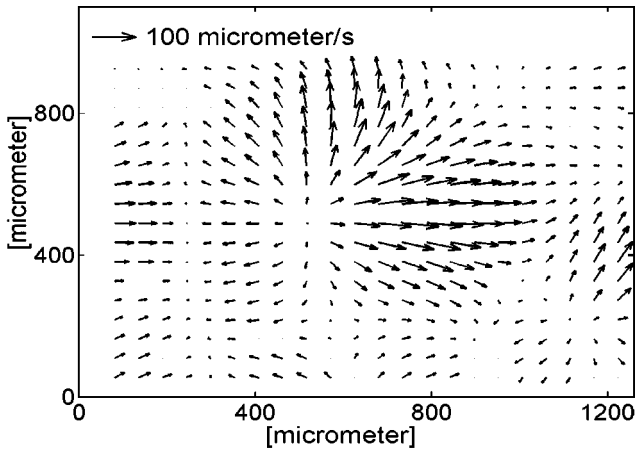


a)

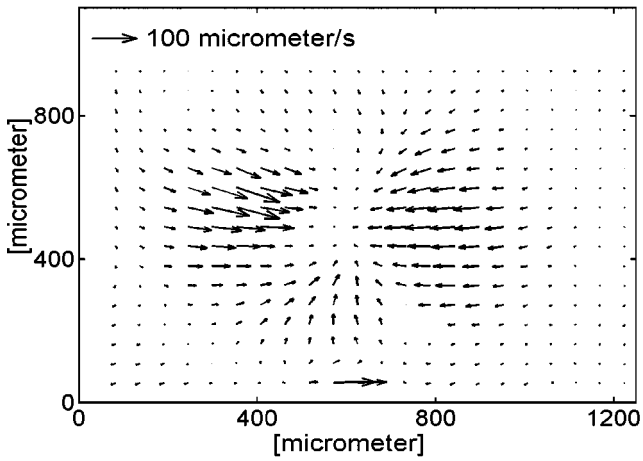


b)

FIG. 3. Streaming field in the fixed path length λ/2 cylindrical resonator at 1.38 MHz: (a)—Streaming velocity distribution in the nodal plane, pressure amplitude 520 kPa; (b)—average streaming velocity vs acoustic pressure amplitude squared, 0.9 mm from the axial source point of the vortex ($\mathbf{R_L}$).



a)



b)

FIG. 4. Velocity fields in the variable path length cylindrical chamber with 0.25 mm channel depth (λ/4): Streaming (a) from the central point of one and (b) towards the central point of another of a number of off axis vortices ($\mathbf{R_S}$).

referred to as: $\mathbf{R_L}$ for large container-scale Rayleigh streaming and $\mathbf{R_S}$ for the small scale near-wall streaming described below.

B. Cylindrical 1.5 MHz chamber with variable depth

Electrical admittance measurements and frequency-voltage scans showed resonances at 1.43 and 1.47 MHz when the channel depths were 0.5 mm (λ/2) and 0.25 mm (λ/4), respectively. These values are close to the transducer's thickness resonance frequency of 1.5 MHz and they were maintained as drive frequencies during the experiments.

1. λ/4 path length

The majority of the 1 μm particles concentrated at the inner surface of the glass reflector in the 0.25 mm (λ/4) path length chamber, where, because of the half wavelength thick reflector in this case, a pressure node is predicted. However, immediately after ultrasound application and before the particles concentrated, numerous small vortices in vertical planes were observed near the reflector. In some areas the particles came to the pressure node plane and moved radially from the center of the vortex to its edges producing, for these near-wall vortices, a pattern similar to the Rayleigh-type streaming shown in Fig. 3(a). In the adjacent regions the particles in the vortices moved from the periphery of the area to its center, then towards the base of the chamber, returned to the periphery and back to the surface of the reflector. Figure 4 shows the velocity fields of the two areas described above.

In the bulk fluid a few vortices circulated in the planes parallel to the transducer plane (i.e., *x*-*y* planes). They were 100–150 μm in diameter and the particles in the vortices circulated continuously without redistribution during the time of experiments. The density of particles in these vortices was much lower than in the nodal plane. The symbol $\mathbf{Vor_{X\text{-}Y}}$ will be used for this kind of vortex whose circulation is parallel to the transducer plane.

2. λ/2 path length

At the channel depth of 0.5 mm (λ/2) with a quarter-wavelength reflector, the distribution of particles differed significantly from the one described for the fixed channel (λ/2) depth chamber in Sec. III A. The contrasting behavior is at-

tributed here to differences in the chambers construction (e.g., different thickness of a coupling steel layer, rubber side walls of the variable depth chamber) and hence, differences in the sound field profiles and boundary layer at the periphery.

After ultrasound application, 24 μm diam latex beads aggregated in many clumps in the central plane of the chamber, where a pressure node was expected to occur. The 1 μm particles concentrated mainly in that plane as well, without a regular pattern. Vortices were also observed in the chamber at the distance of 180–220 μm from the base of the chamber. These vortices, which circulated parallel to the transducer plane ($\mathbf{Vor_{X\text{-}Y}}$), included toroids and counter-rotating vortex pairs of scale ranging from typically sub-millimetre to as high as 5 millimetres.

The following series of experiments was carried out to establish the $\mathbf{Vor_{X\text{-}Y}}$ velocity profile for planes at different distances from the base of the chamber. Before each experiment, the microscope was focussed at the base or in a plane at a distance from it that increased in 25 μm increments to finish at the reflector surface. The behavior of the 1 μm particles in the planes was examined during the first few seconds after ultrasound application. It was observed that within 40–50 μm distances from the base and 100–120 μm from the reflector, the particles moved towards the node plane and had left the planes of focus, within 1 s of the initiation of sonication. There was no evidence of vortex flow at any stage in those planes. In the planes close to the pressure node, it took 2–4 s for the particles to leave the planes and they started moving in circular trajectories as they entered the vortices near the node plane. The experimental evidence suggests that these vortices have no hydrodynamic connection with the walls. Experiments repeated at different applied voltages showed that the velocity of particles in the vortices increased with an increase in acoustic pressure amplitude.

Similar features were also observed in the rectangular chamber.

C. Rectangular 3 MHz chamber

The electrical admittance measurements and frequency-voltage spectra showed a resonance in the chamber at 2.9 MHz, which is close to the transducer's resonance frequency. The channel depths required for $\lambda/2$ and $\lambda/4$ path lengths were 0.250 and 0.125 mm, respectively, whereas practically the depths were 0.270–0.280 and 0.130–0.140 mm, respectively, due to the limitations in gasket-forming techniques. The reflector was a 1 mm thick ordinary microscope glass slide.

In both sound fields 24 μm latex spheres concentrated in a number of clumps in two layers parallel to the transducer plane. The distances between the clumps and the walls and between the layers varied depending on the depth of the chamber. The top layer for the 0.130 mm system was about 10–15 μm from the reflector. Concentration of the particles occurred within a few seconds on ultrasound application. 1 μm latex beads also concentrated in many small clumps in two layers, for both path lengths.

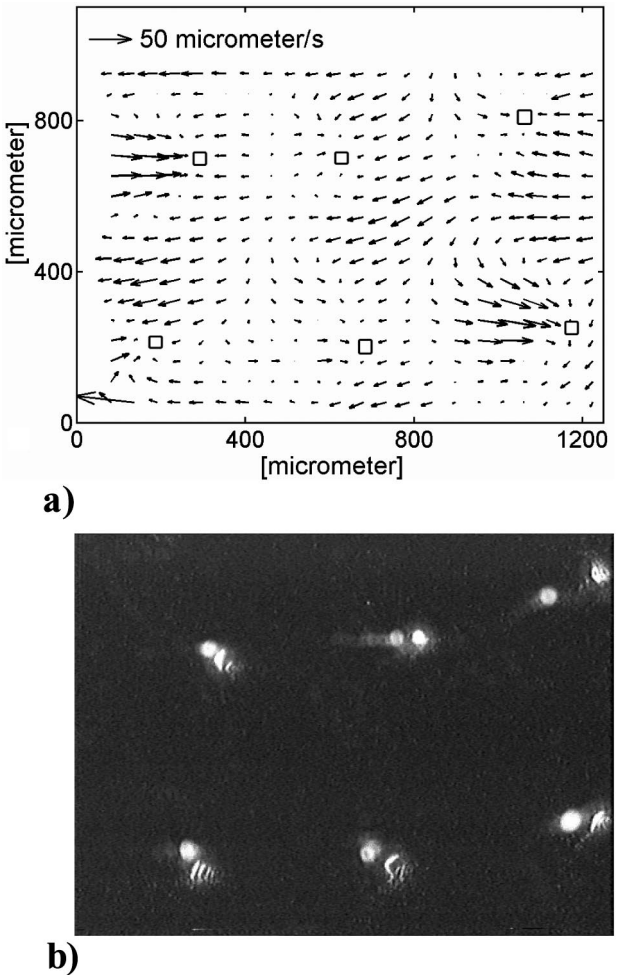


FIG. 5. Clump formation in the rectangular chamber (2.89 MHz, 0.130 mm path length ($\lambda/4$); 1 μm latex) (a)—velocity field, 1 s of sonication (empty blocks show the sites where clumps started to form), (b)—microscopic images of the formed clumps.

In each plane there were two main regions of clump formation within the active radiation area (i.e., the electrode areas). These regions were centerd on the long horizontal axis of the sound field and were symmetrical about the midpoint of this axis with an edge–edge separation of 5 to 6 mm. The typical distances between the small clumps in horizontal planes were 0.3–0.6 mm. The particle distribution in the chamber's channel beyond the boundaries of the etched part of the transducer (20×10 mm^2) had no regular pattern and was least affected by ultrasound.

1. $\lambda/4$ path length

Figure 5 shows the microscopic image of the clumps together with the PIV results of the initial moment of particle aggregation in the plane near the glass reflector, at which the microscope was pre-focused. The clumps were completely formed within 20 s of sound application. As it is seen from the vector map [Fig. 5(a)], the particles moved from different directions to the aggregate formation sites. The second layer of clumps can also be seen as blurred images in Fig. 5(b). No vortices were observed in the region of the clumps in Fig. 5. A few vortices of about 100 μm in diameter circulated in the x-y plane at a distance 40–50 μm from the reflector.

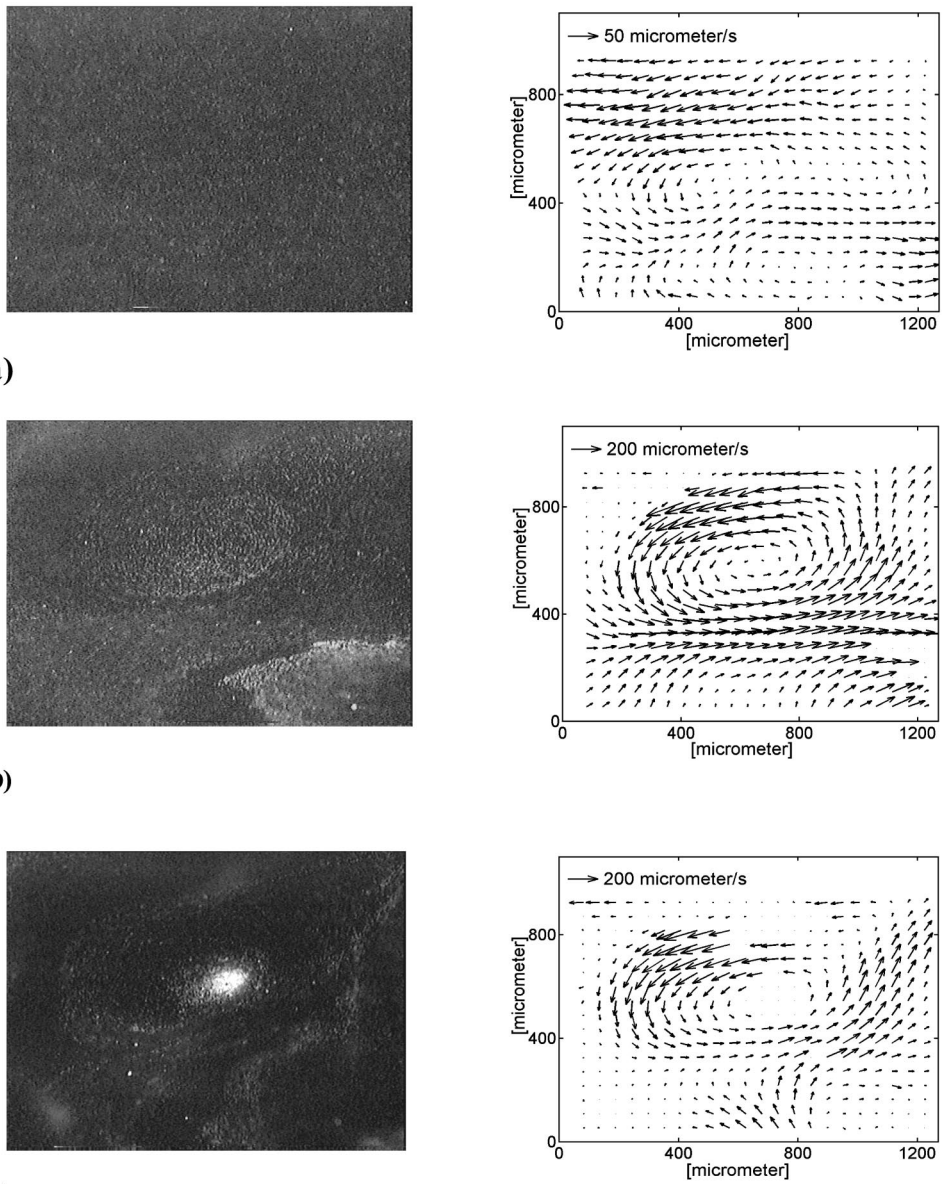


FIG. 6. Microscopic images of vortex trapping ($\mathbf{Vor_{X\text{-}Y}}$) and PIV results of the process (2.89 MHz, 600 kPa, λ/2 path length): (a)—1 s, (b)—16 s, (c)—38 s of sonication.

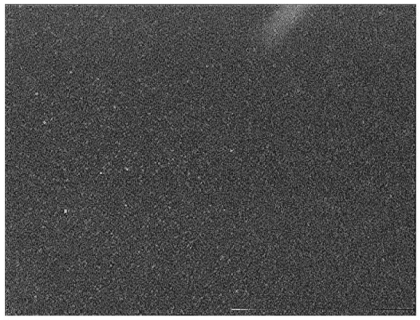
2. λ/2 path length

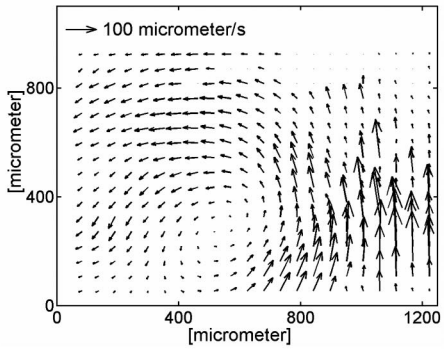
In the 0.270 mm resonator vortices were observed at a distance of 1/3 to 1/2 of the channel's depth from the base of the chamber. The 1 μm latex beads following the streaming in the vortices produced several types of pattern, mainly spirals, orbital toroids and counter-rotating vortex pairs. The circulation of particles occurred in the planes parallel or at small angles to the transducer plane. It is not possible to select typical vortex dimensions as the scale of the streaming ranged from 0.2 mm to 3 and 4 mm.

The 1 μm particles in these vortices exhibited two main types of behavior. The first case will be referred to as ''vortex trapping,'' when the particles moved across the streamlines of the vortex and concentrated in the center of rotation forming a clump, as shown in Fig. 6. In the second case, on the contrary, the particles moved away from the center and formed an orbital toroid, which gradually increased in diameter (Fig. 7).

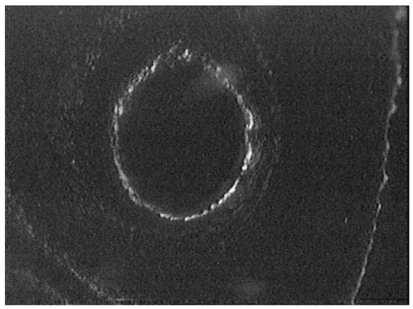
D. Particle distributions in bimodal sized suspensions

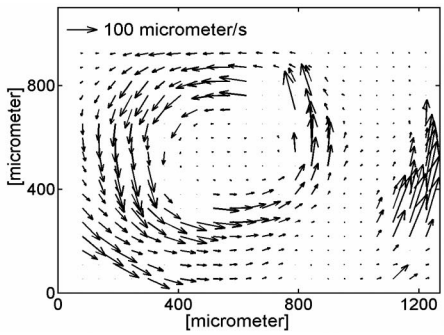
The following experiments were carried out to investigate the extent of relationships between vortex flow and particle concentration regions. Mixtures of 1 μm fluorescent beads and larger (24 or 80 μm) plain latex beads were introduced into the variable depth 1.5 MHz cylindrical chamber when the path length was λ/2. When ultrasound was applied vortices ($\mathbf{Vor_{X\text{-}Y}}$) were immediately generated in the chamber. The large particles mainly concentrated in a number of clumps. The redistribution of the 1 μm particles in the vortices proceeded in the way described above: As a rule, either clumps at the center or rotating rings were formed. In some of the vortices, the 24 μm particles together with the smaller particles formed a mix-sized clump at the center of the vortex. However, the preferred positions of the clusters of large particles were between counter-rotating vortices as shown for 80 μm particles in Fig. 8. In the vortices with relatively



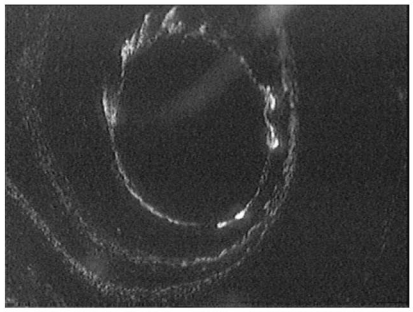


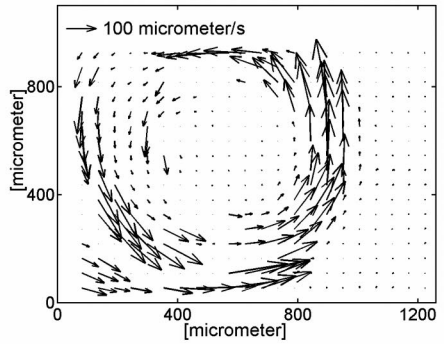
a)





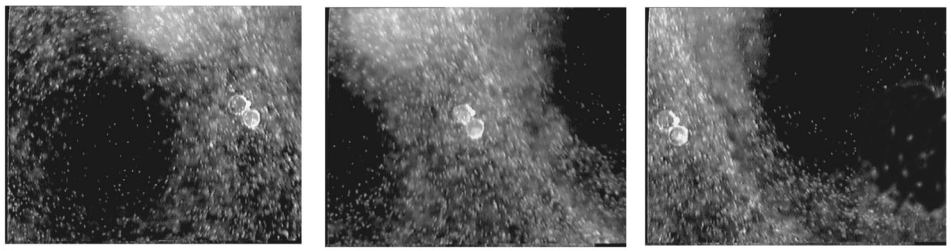
b)



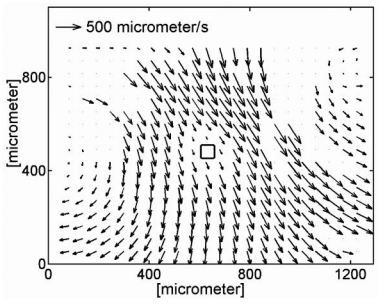


c)

FIG. 7. Microscopic images and PIV results of ring formation (**Vor**$_{\mathbf{X\text{-}Y}}$, 2.89 MHz, 310 kPa, $\lambda/2$ path length): (a)—1 s, (b)—39 s, (c)—58 s of sonication.



a)



b)

FIG. 8. Particle distribution in vortices (**Vor**$_{\mathbf{X\text{-}Y}}$): (a)—microscopic images of left, central, and right parts of a counter-rotating vortex pair with 80 μm beads between the vortices, (b)—velocity vector map of the central part (empty block shows the position of the 80 μm beads).

TABLE I. Summary of particle aggregate and vortex distributions in the range of fields examined.

Transducer, nominal thickness frequency (MHz)	Drive frequency (MHz); path length (Z)	24 μm aggregate in node plane	Rayleigh type circulation	1 μm latex, regular aggregates[a] (Fig. 5)	x-y plane vortex, $\mathbf{Vor_{X\text{-}Y}}$	1 μm latex, vortex center aggregate (Fig. 6), orbital concentration (Fig. 7)	Counter rotating vortices
Disc 1 1.5	1.38; (λ/2)	Single, central	Field width scale ($\mathbf{R_L}$)	No	No	No	No
Disc 2 1.5	1.43; (λ/2)	Multi	No	No	Yes	Both	Yes
Disc 2 1.5	1.47; (λ/4)	Multi	1 mm scale ($\mathbf{R_S}$)	No	Yes	No	No
Rect. Plate 3.0	2.89 (λ/2)	Multi	No	Yes	Yes	Both	Yes
Rect. Plate 3.0	2.89 (λ/4)	Multi	No	Yes	Yes	No	No

[a]Regular aggregates are formed under the effect of lateral DRF in the nonstreaming context.

high flow velocity (about 1 mm/s) a small number of the 24 μm beads participated in the streaming forming, together with 1 μm particles, an orbital toroid. Finally, there were occasional counter-rotating pairs in which particles in one of the vortices moved outwards to form a rotating ring of particles whereas a clump was formed in the center of the other.

IV. DISCUSSION

A. Particle aggregate and vortex distributions in different fields

The main features of the different chambers employed and the associated characteristic streaming patterns are summarized in Table I. Row 1 is the fixed path length cylindrical chamber with Rayleigh streaming ($\mathbf{R_L}$, Fig. 3). It formed a single central aggregate of relatively large (24 μm) particles and showed no evidence of smaller vortices in the x-y plane. The four remaining rows identify the systems that formed a number of 24 μm-particle clumps distributed across the area of the field. This occurrence of multiple aggregates was invariably accompanied by the existence of regions containing vortices in the x-y plane ($\mathbf{Vor_{X\text{-}Y}}$).

The existence of two planes of particle concentration in the rectangular chamber indicates that the layers of which the resonator is comprised do not ideally match the requirements for a single resonance in the water layer. This should be particularly the case for the λ/2 water layer arrangement with 1 mm thick (λ/2 rather than λ/4) glass reflector. This apparently led to the situation where in addition to a node at which the particles concentrated there was an anti-node in the axial pressure profile within the water layer.[7] The DRF_a drove particles in opposite directions away from the anti-node so that some concentrated near the wall.

B. Rayleigh-type streaming in the cylindrical field

In order to estimate how the experimental results in the cylindrical chamber with fixed depth correlate with a Rayleigh micro-streaming pattern, the velocities obtained in the experiment were compared to the average streaming velocity in a nodal plane calculated with the Rayleigh model. While the streaming pattern in the node plane observed in the experiments was symmetrical to the central point and matched the acoustic field of a cylinder, the theory of Schuster and Matz,[44] where the radius of the cylinder is small compared to a wavelength, could not be precisely applicable to the system. The values derived for the radial velocity components from the streaming between parallel walls theory[11,13] and streaming in a cylinder theory[11,44] do not differ much for similar geometric dimensions.[23,45]

As it was reported,[11,23] velocity increased gradually and reached its average value at approximately 30% of the channel's half-width. The increase in this region was essentially linear. The experimental velocity estimated from PIV results in this work also increased linearly as a function of distance from the central point of the vortex. The velocity values at the border of the field of view were linearly extrapolated to the distance, where, according to the Rayleigh theory, the average streaming velocity is reached (0.9 mm from the central point corresponds to 30% of the channel's half-width in the present work). As Table II shows, the experimental velocity is quantitatively of the same order of magnitude as the calculated average Rayleigh micro-streaming velocity.

In Fig. 3(b) streaming velocity at the distances of 0.9 mm from the central point (average value) is plotted against squared acoustic pressure amplitude. The experimental dependence is well described by a straight line with a slope of 1.05 $\mathrm{mm\,s^{-1}\,MPa^{-2}}$. The average streaming velocity is essentially proportional to the squared acoustic particle velocity, which is in its turn linearly proportional to the acoustic pressure amplitude and the voltage applied to a transducer.[11,43]

The streaming pattern observed and its radial and pressure dependence are consistent with results recently reported for a rectangular field λ/2 chamber designed to give a single central large particle aggregate.[23] One can conclude that even

TABLE II. Streaming velocity obtained in experiments and calculated with the Rayleigh model.

Acoustic pressure amplitude, kPa	Calculated average Rayleigh streaming velocity, $\mu\mathrm{m\,s^{-1}}$	Measured average streaming velocity, $\mu\mathrm{m\,s^{-1}}$
380	35	86
450	50	116
520	66	166

TABLE III. Estimates of the Stokes drag and the centrifugal force operating on a particle with tangential ν_T and radial ν_R velocity orbiting in the x-y plane vortices.

Pressure amplitude (kPa)	ν_T (μm s^{-1})	ν_R (μm s^{-1})	Annular radius r (μm)	Stokes drag, F_s (pN)	Centrifugal force (pN)
310	100	6	270	5.1×10^{-2}	9.7×10^{-10}
600	590	20	300	17.1×10^{-2}	3.1×10^{-8}

though the assumptions made by Rayleigh for his model were not fully met in these experiments, it is still reasonable to use the model as a guide.

C. General particle movement within hydrodynamic vortices

The 1.0 μm diam latex beads were employed here as tracers to detect flow velocities by the PIV method. However, it is clear, since these particles become redistributed within vortices in, or at a small angle to, the x-y plane as sonication continues ($\mathbf{Vor_{X\text{-}Y}}$, Figs. 6 and 7), that they are experiencing significant forces within the vortices.

It is known in the nonacoustic context that particle movement within a vortex can arise through centrifugal force driving the circulating particles outwards from the vortex center or through an inward centripetal force when particles have a density greater or less than that of the suspending phase, respectively.[46] The tangential (ν_T) or radial components (ν_R) of the microparticle velocity during development of the orbital toroid in Fig. 7 are given in Table III. The Stokes drag F_s on the particles that move across the streamlines with a velocity ν_R is given by

$$F_s = 6\pi\eta a \nu_R, \tag{3}$$

where η is the viscosity of the liquid and a is the radius of the sphere. F_s is shown for two measured velocities in Table III. The centrifugal force on a particle of buoyant mass m_B at a distance r from the center of curvature of the path of a particle with tangential velocity ν_T is given by

$$F_c = m_B \nu_T^2 / r. \tag{4}$$

The path is approximated here as a circle and r becomes the distance from the center of the vortex. It can be seen from Table III that the centrifugal force operating on the particle is many orders of magnitude too small to account for the movement of particles to the path of high particle concentration (Fig. 7). Neither, for the present situation where particle and fluid densities are essentially constant, does centrifugal force accommodate the option of radially inwards or outwards movements of particles.

It is reasonable to assume that the mechanism of particle re-distribution in the vortices originates from the nonuniformities in the lateral acoustic pressure profile. The narrow annular concentration in the plane of the larger cross-sectional area of the oblate spheroid developed from particle movement outwards from the center (Fig. 7). The order of magnitude of the acoustic pressure amplitude required to drive the individual 1 μm particles radially outwards was estimated as follows. The radiation force on 1 μm particle was taken as that for a harmonic plane standing wave, as in Eq. (1). When the pressure at the particle ring was taken as zero, and the frequency as 2.9 MHz the value of P_0 works out as 690 kPa. This estimate is of the same order of magnitude as the pressure measured in the node plane by the levitation technique (600 kPa, methods above).

D. Vortex development in or at a small angle to the x-y plane ($^{\|T}$Vor) and particle redistribution within those vortices

The presence of Rayleigh-type vortices is associated with energy dissipation at the liquid–solid interfaces. However, the stable vortex streaming patterns shown in Figs. 6 and 7 were not hydrodynamically linked to the walls of the resonator and, therefore, do not have their origins in Rayleigh or Schlicting streaming. The energy absorption required to maintain $\mathbf{Vor_{X\text{-}Y}}$ circulation could not come from a pure standing wave, therefore, the field necessarily has a travelling wave components either in one or both directions parallel or perpendicular to the transducer plane. It has been shown elsewhere that conventional ''quartz wind'' (Eckart) streaming is suppressed in short path length ultrasonic resonators in a 3.3 MHz sound field.[47] The present work suggests that cavitation can be ruled out as the cause of the described vortices. At the beginning of each experiment no air bubbles were observed in the chambers, nor did they appear during the experimental runs when the streaming vortices were well established. This is consistent with the results obtained from monitoring the acoustic field with a hydrophone in a similarly designed chamber and in the same range of frequencies and acoustic pressure amplitudes.[48] It was shown that no cavitation activity occurred in that chamber unless contrast agent bubbles were purposely introduced into the system.

Time independent flows in sonicated liquids have been extensively studied.[14,15,49] Nyborg[14,49] pointed out many years ago that in general a torque is exerted on any volume of a sonicated liquid even when the latter is homogeneous. The time-averaged torque L on a spherical element about its center is given by

$$L = (I/2\rho_0)\nabla \times F, \tag{5}$$

where I is the moment of inertia of the sphere about its center and F is a force per unit volume. The magnitude of F and, therefore, of the torque depends only on the first order velocity u and tends to be large where u is nonuniform, e.g., near the edge of a beam or an edge of an obstacle in the field. While two- or three-dimensional (2D or 3D) pressure and velocity distributions in real operational nominally ''plane'' standing waves have not yet been determined for short path length chambers, the origin of radiation torque distributions which drive the vortices is not known. In the following we discuss the redistribution of particles that occurs when these vortices are present.

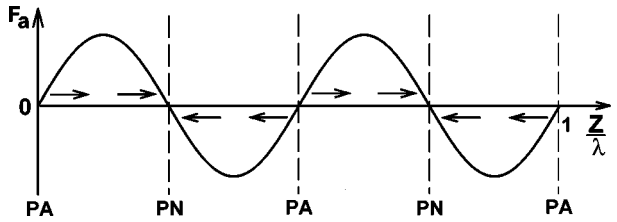


FIG. 9. Direct radiation force variation along sound beam axis for a plane standing wave; arrows show the directions of micro-particle migration towards pressure nodes.

The axial direct radiation force that acts on suspended particles in an ideal plane standing wave field is described by Eq. (1). Its variation along the sound beam axis is shown in Fig. 9. Although the force is zero near both pressure nodes and anti-nodes, a stable equilibrium occurs only near pressure nodes.[3,45] Experimental studies[45] have shown that the acoustic pressure in a real anti-nodal planes is not constant, but has a complex distribution with minima and maxima. Since the chambers employed in the present work are of the order of 10 wavelengths in the transverse dimension, there is likely to be some variation in radiation across the face of the transducers. The experimental results of the present work then may be explained by a lateral radiation force variation having the same spatially periodic dependence as that shown in Fig. 9. The lateral radiation force drives particles towards the stable pressure node positions and away from pressure anti-nodes. The redistribution of particles in the vortices ($\mathbf{Vor_{X\text{-}Y}}$, Figs. 6 and 7) is then consistent with a vortex having a center at a local pressure minimum when particles clump at the center (Fig. 6) and with a local pressure maximum when they move to an orbital path (Fig. 7). The implication that orbital toroids are at pressure minima is supported by the observation that the largest particle studied (80 μm) for which the effect of radiation force (particle volume dependent) prevails many times over the streaming drag (particle radius dependent),[23] usually took up a stable position at the overlap of two counter-rotating vortices (Fig. 8). The experimental observations also showed that counter-rotating vortices may present as if both centers are at a pressure minimum (Fig. 8) or at pressure maximum or as if one center is a node and other and anti-node.

The results presented in this work for the first time clearly showed that acoustic streaming vortices of different kind developed in single- and multi-well short path length USSW resonators. Since chambers similar to the ones described above are employed for cell and particle manipulation[40,50] and cell sonoporation[48] studies, the knowledge of physical environment in which the processes occur proves very useful. In the context of the microfluidic application given in the Introduction, the results of the present work suggest that the smaller Rayleigh vortices shown in the $\lambda/4$ resonator may have the capability to enhance molecular diffusion to a sensor surface by reducing the boundary layer thickness at that surface. Finally our results confirm the conclusion of Spengler *et al.*[23] that in a $\lambda/2$ resonator showing container-wide Rayleigh vortices particles of the order of 25 μm diam. are driven by radiation force to the center of the pressure node plane while the volume dependent [Eq. (1)] force on 1 μm particles cannot overcome convective streaming and they are swept from the axial region. However, we go on to show that where smaller scale vortices $\mathbf{R_S}$ are produced the radiation force on 1 μm particles determines the position that the particles occupy in those vortices.

V. CONCLUSIONS

Radially symmetric flow from the center of the node plane ($\mathbf{R_L}$) in the $\lambda/2$ cylindrical chamber was analyzed by PIV. The streaming velocities as well as their dependence on distance and squared acoustic pressure amplitude, were consistent with the Rayleigh model. Rayleigh-type vortices in which the 1 μm particles streamed radially towards or away from the centres of small areas ($\mathbf{R_S}$) were observed in the pressure node plane of $\lambda/4$ second cylindrical resonator. At $\lambda/2$ channel depth of this and of a rectangular chamber vortices developed away from the walls in planes parallel to the transducer plane ($\mathbf{Vor_{X\text{-}Y}}$). Particles in these vortices became redistributed and formed either a clump at the center or a ring rotating about the vortex center. The results suggest that the lateral acoustic pressure gradients determined the redistribution of particles in the vortices.

It follows from the above that, for the $\mathbf{Vor_{X\text{-}Y}}$ situation of rows 4 and 5 of Table I, concentration of 1 μm particles arises from radiation pressure and does not per se depend on streaming. The vortices are often associated with local pressure patterns capable of concentrating particles but particle concentration can occur in the absence of streaming (Fig. 5) and $\mathbf{Vor_{X\text{-}Y}}$ can occur without particle concentration (Sec. III B 2).

ACKNOWLEDGMENTS

The authors are happy to acknowledge useful discussion with Prof. W. L. Nyborg. They are grateful to Sian Armstrong, School of Engineering, Cardiff University, for the admittance measurements and to Dr. Colin Bates, School of Engineering, Cardiff University, for Dantec Dynamics software setup. L. A. Kuznetsova was supported by the BBSRC (Research Grant No. 72/E17416).

[1] L. P. Gor'kov, "On the forces acting on a small particle in acoustical field in an ideal fluid," Sov. Phys. Dokl. **6**, 773–775 (1962).
[2] T. Tuzuiti, T. Kozuka, and H. Mitome, "Measurement of distribution of acoustic radiation force perpendicular to sound beam axis," Jpn. J. Appl. Phys., Part 1 **38**(5B), 3297–3301 (1999).
[3] S. M. Woodside, B. D. Bowen, and J. M. Piret, "Measurement of ultrasonic forces for particle-liquid separation," AIChE J. **43**, 1727–1736 (1997).
[4] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
[5] J. J. Hawkes, W. T. Coakley, M. Gröschl, E. Benes, S. Armstrong, P. J. Tasker, and H. Nowotny, "Single half-wavelength ultrasonic particle filter: Predictions of the transfer matrix multilayer resonator model and experimental filtration results," J. Acoust. Soc. Am. **111**, 1259–1266 (2002).
[6] M. Hill, Y. Shen, and J. J. Hawkes, "Modelling of layered resonators for ultrasonic separation," Ultrasonics **40**, 385–392 (2002).
[7] M. Hill, "The selection of layer thicknesses to control acoustic radiation force profiles in layered resonators," J. Acoust. Soc. Am. **114**, 2654–2661 (2003).
[8] E. Benes, M. Gröschl, H. Nowotny, F. Trampler, T. Keijzer, H. Böhm, S. Radel, L. Gherardini, J. J. Hawkes, R. König, and Ch. Delouvroy, "Ultrasonic separation of suspended particles," in *Proceedings of IEEE Int. Ultrasonic Symposium* (Atlanta, Georgia, USA, 2001), pp. 649–659.

[9] D. L. Miller, ''Instrument for microscopical observation of the biophysical effects of ultrasound,'' J. Acoust. Soc. Am. **60**, 1203–1212 (1976).
[10] L. K. Zarembo, ''Acoustic streaming,'' in *High-Intensity Ultrasonic Fields*, edited by L. D. Rozenberg (Plenum, New York, 1971), Vol. 85, pp. 137–199.
[11] J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. 2.
[12] H. Mitome, ''The mechanism of generation of acoustic streaming,'' Electron. Commun. Jpn., Part 2: Electron. **81**(10), 1614–1620 (1997).
[13] W. L. Nyborg, ''Acoustic streaming,'' in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, London, 1965), Vol. 2, P. B, pp. 265–331.
[14] W. L. Nyborg, ''Acoustic streaming equations: Laws of rotational motion for fluid elements,'' J. Acoust. Soc. Am. **25**, 938–944 (1953).
[15] N. Riley, ''Steady streaming,'' Annu. Rev. Fluid Mech. **33**, 43–65 (2001).
[16] W. L. Nyborg, ''Acoustic streaming,'' in *Nonlinear acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, Boston, 1998), pp. 207–228.
[17] M. Raffel, C. Willert, and J. Kompenhans, *Particle image velocimetry: a practical guide* (Springer, Heidelberg, 1998).
[18] M. Campbell, J. A. Cosgrove, C. A. Greated, S. Jack, and D. Rockliff, ''Review of LDA and PIV applied to the measurement of sound and acoustic streaming,'' Opt. Laser Technol. **32**, 629–639 (2000).
[19] H. Mitome, T. Kozuka, and T. Tizuiti, ''Measurement of the establishment process of acoustic streaming using laser Doppler velocimetry,'' Ultrasonics **34**, 527–530 (1996).
[20] S. Nomura, K. Murakami, and Y. Sasaki, ''Streaming induced by ultrasonic vibration in a water vessel,'' Jpn. J. Appl. Phys., Part 1 **39**, 3636–3640 (2000).
[21] J. P. Share, C. A. Greated, C. Gray, and D. M. Campbell, ''The measurement of acoustic streaming using particle image velocimetry,'' Acustica **68**, 168–172 (1989).
[22] J. A. Cosgrove, J. M. Buick, S. D. Pye, and C. A. Greated, ''PIV applied to Eckart streaming produced by a medical ultrasound transducer,'' Ultrasonics **39**, 461–464 (2001).
[23] J. F. Spengler, W. T. Coakley, and K. T. Christensen, ''Microstreaming effects on particle concentration in an ultrasonic standing wave,'' AIChE J. **49**, 2773–2782 (2003).
[24] C. Suri, K. Takenaka, K. Kojima, and K. J. Koyama, ''Experimental study of a new liquid mixing method using acoustic streaming,'' Chem. Eng. J. **35**, 497–502 (2002).
[25] E. H. Trinh, 1998. ''Acoustic streaming in microgravity: flow stability and heat transfer enhancement,'' [WWW]⟨http://ncmr04610.cwru.edu/events/fluids1998/abstracts/348.pdf⟩ [Accessed 5 February, 2004].
[26] E. H. Trinh and J. L. Robey, ''Experimental study of streaming flow associated with ultrasonic levitation,'' Phys. Fluids **6**, 3567–3579 (1994).
[27] J. A. D. Jensen, P. Pocwiadowski, P. O. A. Perrson, L. Hultman, and P. Møller, ''Acoustic streaming enhanced electrodeposition of nickel,'' Chem. Phys. Lett. **368**, 732–737 (2003).
[28] X. Shi, R. W. Martin, S. Vaezy, and L. A. Crum, ''Quantitative investigation of acoustic streaming in blood,'' J. Acoust. Soc. Am. **111**, 1110–1121 (2002).
[29] K. R. Nightingale, P. J. Kornguth, and G. E. Trahey, ''The use of acoustic streaming in breast lesion diagnosis: A clinical study,'' Ultrasound Med. Biol. **25**, 75–87 (1999).
[30] M. A. Stremler, F. R. Haselton, and H. Aref, ''Designing for chaos: applications of chaotic advection at the microscale,'' Philos. Trans. R. Soc. London, Ser. A **362**, 1019–1036 (2004).
[31] J. M. Ottino and S. Wiggins, ''Introduction: mixing in microfluidics,'' Philos. Trans. R. Soc. London, Ser. A **362**, 923–935 (2004).
[32] K. Yasuda, ''Non-destructive, non-contact handling method for biomaterials in micro-chambers by ultrasound,'' Sens. Actuators B **64**, 128–135 (2000).
[33] J. C. Rife, M. I. Bell, J. S. Horwitz, M. N. Kabler, R. C. Y. Auyeng, and W. J. Kim, ''Miniature valveless ultrasonic pumps and mixers,'' Sens. Actuators, A **86**, 135–140 (2000).
[34] X. Zhu and E. S. Kim, ''Microfluidic motion generation with acoustic waves,'' Sens. Actuators, A **66**, 355–360 (1998).
[35] R. Liu, R. Lenigk, R. L. Druyor-Sanchez, J. Yang, and P. Grodzinski, ''Hybridization enhancement using cavitation microstreaming,'' Anal. Chem. **75**, 1911–1917 (2003).
[36] A. M. Wang, R. Kiwan, M. White, and R. L. Ceriani, ''A silicon-based ultrasonic immunoassay for detection of breast cancer antigens,'' Sens. Actuators B **49**, 13–21 (1998).
[37] 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**, 1463–1471 (2000).
[38] J. J. Hawkes and W. T. Coakley, ''Force field particle filter, combining ultrasound standing waves and laminar flow,'' Sens. Actuators B **75**, 213–222 (2001).
[39] J. J. Hawkes, M. Gröschl, E. Benes, H. Nowotny, and W. T. Coakley, ''Positioning particles within liquids using ultrasound force fields,'' on the CD-ROM: *Sevilla, September 16–20, Proceedings, Forum Acusticum 2002* (ISBN 84-87985-07-6), Revista De Acustica 33 no 3-4, Acoustics of disperse particulate matter paper SS-PHA-01-007-IP.
[40] J. J. Hawkes, M. J. Long, W. T. Coakley, and M. McDonnell, ''Ultrasonic deposition of cells on a surface,'' Biosens. Bioelectron. **19**, 1021–1028 (2004).
[41] J. J. Hawkes and W. T. Coakley, ''A continuous flow ultrasonic cell-filtering method,'' Enzyme Microb. Technol. **19**, 57–62 (1996).
[42] K. T. Christensen, S. M. Soloff, and R. J. Adrian, ''Integrated particle image velocimetry interrogation/validation software,'' Technical report No 943, Theor. Appl. Mechan., University of Illinois, Urbana-Champaign (2000).
[43] J. F. Spengler, M. Jekel, K. Christensen, R. Adrian, J. J. Hawkes, and W. T. Coakley, ''Observation of yeast cell movement and aggregation in a small scale MHz-ultrasonic standing wave field,'' Bioseparation **9**, 329–341 (2001).
[44] K. Schuster and W. Matz, ''Uber stationare stromungen in kundtschen rohr,'' Akust. Z. **5**, 349–352 (1940).
[45] R. K. Gould and W. T. Coakley, ''The effect of acoustic forces on small particles in suspension,'' in *Proceedings of the 1973 Symposium on Finite Amplitude Wave Effects in Fluids* (Pergamon, Guildford, UK, 1974), pp. 252–257.
[46] P. Nielsen, ''Turbulence effects on the settling of suspended particles,'' J. Sediment. Petrol. **63**, 835–838 (1993).
[47] J. Spengler and M. Jekel, ''Ultrasound conditioning of suspensions—studies of streaming influence on particle aggregation on a lab- and pilot-plant scale,'' Ultrasonics **38**, 624–628 (2000).
[48] S. Khanna, N. N. Amso, S. J. Paynter, and W. T. Coakley, ''Contrast agent bubble and erythrocyte behavior in a 1.5 MHz standing ultrasound wave,'' Ultrasound Med. Biol. **29**, 1463–1470 (2003).
[49] W. L. Nyborg, ''Physical principles of ultrasound,'' in *Ultrasound: its application in medicine and biology*, edited by F. J. Fry (Elsevier, Amsterdam, 1978), Vol. 1, pp. 1–75.
[50] W. T. Coakley, D. Bazou, J. Morgan, G. A. Foster, C. W. Archer, K. Powell, K. A. J. Borthwick, C. Twomey, and J. Bishop, ''Cell–cell contact and membrane spreading in an ultrasound trap,'' Colloids Surf., B **34**, 221–230 (2004).

Acoustic concentration of particles in piezoelectric tubes: Theoretical modeling of the effect of cavity shape and symmetry breaking

Shulim Kogan, Gregory Kaduchak,[a] and Dipen N. Sinha
Los Alamos National Laboratory, MS D429, Los Alamos, New Mexico 87545

(Received 4 February 2004; revised 21 June 2004; accepted 4 July 2004)

A new class of simple, highly efficient, cylindrical acoustic concentration devices has been developed based upon cylindrical (or near cylindrical) geometries [Kaduchak *et al.*, Rev. Sci. Instrum. **73**, 1332–1336 (2002)] for aerosol concentration applications. The concentrators are constructed from single PZT tubes driven at or near the breathing mode resonance. Acoustic concentration of aerosols is performed within the tube cavity. It has been found that slight modifications to the cylindrical cavity geometry can significantly increase the collection efficiency and assist in precise particle positioning. This paper analyzes the theoretical framework for the acoustic concentration of particles in these devices for various geometrical perturbations. The cavity geometries studied are (1) hollow cylindrical piezoelectric tube, (2) hollow piezoelectric tube with an inner concentric solid cylinder insert, (3) a hollow piezoelectric tube with a concentric elliptic insert which breaks the circular-cylindrical symmetry, and (4) a hollow elliptic cylindrical piezoelectric tube. It is shown that breaking the circular symmetry within the cavity localizes the particles in small spatial regions within the cavity. This localization of particles may be very useful in applications requiring aerosol collection or particle stream positioning. [DOI: 10.1121/1.1785613]

PACS numbers: 43.25.Uv, 43.25.Qp, 43.25.Gf [MFH] Pages: 1967–1974

I. INTRODUCTION

Detection, identification, and separation of aerosols have applications in a wide variety of fields. Particle size, size distribution, and composition are quantities that are important to many researchers, from those studying the environmental impact of industrial smog to military specialists detecting the presence of chemical or biological warfare agents in the air. Most aerosol detection devices (particle sizers, optical classifiers, etc.) yield higher detection sensitivities if they are given higher aerosol concentrations at the detector. In addition, particle laden exhaust streams may be ''cleaned'' by separating and collecting particles that are considered harmful to the environment. The present research investigates a new class of aerosol concentrators that are based on cylindrical, acoustic resonators.[1–4] It is anticipated that this type of device can be implemented to increase the concentration of aerosol particles at collection points for increased sensitivity of in-line aerosol detectors and also for efficient separation of aerosols from flow streams (see Fig. 1).

Previous experimental work on low-power cylindrical acoustic concentrators demonstrated the concentration efficiency in a near perfect cylindrical cavity.[1–3] This work showed that extremely efficient, low power concentration is possible by implementing a PZT tube to serve as both the cylindrical cavity as well as the acoustic driving source. It was followed by findings that showed the advantages of breaking the circular symmetry of the resonant cavity.[2,4] Symmetry breaking showed tremendous value in decreasing the spatial distribution of concentrated particles, thereby substantially increasing the concentration efficiency and resulting in a significant simplification of the particle collection apparatus. In this paper, a theoretical basis for the cylindrical acoustic levitator and the effects of symmetry breaking are given. It addresses the acoustic wave field in the resonant cavity and describes the resultant force fields for both cavities of cylindrical symmetry and cavities of lower order symmetry. More specifically, details of acoustic concentration of particles in devices of various geometries are presented: (1) circular cylinder, (2) circular cylinder with concentric cylindrical insert, (3) circular cylinder with concentric elliptic insert, and (4) elliptic cylindrical cavity.

II. THE CONCENTRATION FORCE

It is well known that, in a standing acoustical wave, an acoustic radiation force acts on a spherical particle.[5–9] As was shown by Gor'kov[8] and others, this force results from the interference of the incident standing wave and the wave scattered by the particle. Following Gor'kov,[8] Barmatz and Collas[10] developed a general theory to derive acoustical modes for various geometries of acoustical cavities. This theory[5–10] assumes that the size R_p of the particles to be concentrated is many times smaller than the wavelength λ of the acoustic field in a fluid medium. It also assumes that the amplitude of the acoustic wave is small. In particular, the velocity amplitude of the particle is much smaller than the sound speed in the fluid. Moreover, a low concentration of the particles is considered and multiple scattering of the acoustic waves by the particles is negligible. We also con-

[a]Author to whom correspondence should be addressed. Electronic mail: kaduchak@lanl.gov

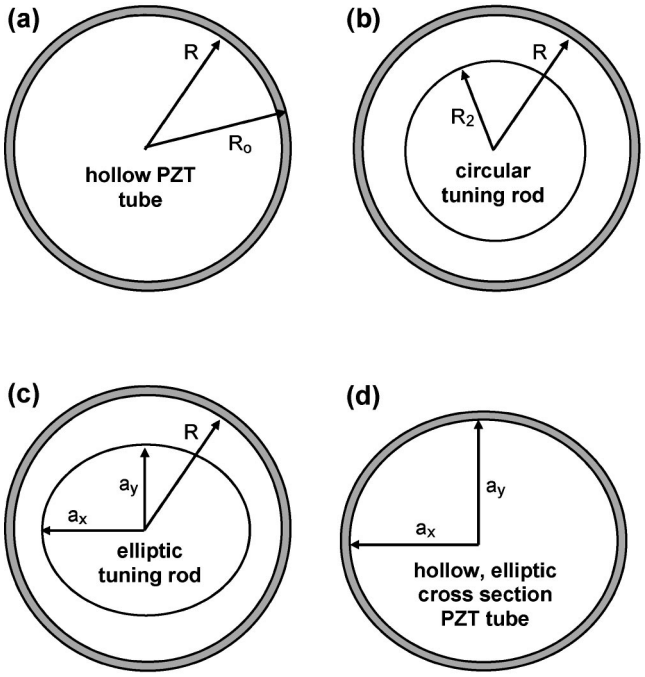


FIG. 1. The schematic design of the concentrating device. (a) Circular cylinder. (b) Circular cylinder with an inner circular cylindrical insertion. (c) Circular cylindrical tube with a coaxial elliptic insertion inside. (d) Hollow elliptic cylindrical tube.

sider below only spherical particles and do not analyze the effects of the deviation of the particles' geometry from a spherical one.

When the amplitude of the fluid oscillation is smaller than the dimension of the particle, the motion of the fluid around the particle is potential[11] (the velocity $\mathbf{v}$ is curlless, **curl** $\mathbf{v}=0$). Then the velocity $\mathbf{v}(\mathbf{r},t)=\mathbf{grad}\Phi(\mathbf{r},t)$ where $\Phi(\mathbf{r},t)$ is the velocity potential. This allows for solution of the wave equation for Φ and to express all quantities in terms of this potential. The solution to the wave equation for Φ can be expressed as a sum of the wave in the absence of particles (subscript 0) and the scattered wave (subscript sc):

$$\Phi(\mathbf{r},t)=\Phi_0(\mathbf{r},t)+\Phi_{\mathrm{sc}}(\mathbf{r},t). \tag{1}$$

The goal of the theory is to express the scattered wave in terms of the incident standing wave and, finally, to express the force in terms of the standing wave potential $\Phi_0(\mathbf{r},t)$ and acoustic velocity $\mathbf{v}(\mathbf{r},t)$ in the absence of particles.

The components F_i of the force $\mathbf{F}$ acting on a spherical particle are given by[8]

$$F_i=-\frac{1}{2}\rho V_p\nabla_i\left[\frac{1}{s^2}\overline{\dot{\Phi}_0^2}-\frac{3}{2}\overline{v^2}\right]. \tag{2}$$

Here ρ is the mass density of the fluid in which the particles are suspended, s is the sound speed in the fluid, $V_p=4\pi R_p^3/3$ is the volume of the particle of radius R_p, upper dot stands for time derivative, and the upper bar means averaging over time. In air ρ and s are much smaller than the corresponding quantities in the particles (ρ_p and s_p). Therefore the factors in Ref. 8 depending on the ratios ρ/ρ_p and s/s_p are close to unity and are omitted in Eq. (2).

As follows from Eq. (2) the force $\mathbf{F}$ is potential, $\mathbf{F}=-\mathbf{grad}U(\mathbf{r})$, and the potential energy of a particle is

$$U(\mathbf{r})=U_0\left[\frac{2}{3s^2v_0^2}\overline{\dot{\Phi}_0^2}-\overline{v^2/v_0^2}\right], \tag{3}$$

where $U_0=\pi\rho R_p^3v_0^2$ is the characteristic potential energy of a particle and v_0 is the amplitude of the transducer wall velocity.

On the rhs of Eq. (3) the first term is positive and the second is negative or zero. Therefore the minima of the potential energy are located at points where the acoustic velocity v is high. No minimum of $U(\mathbf{r})$ can exist at a point where, by symmetry, $\mathbf{v}=0$ (like the center of a circular cylinder when driven in an axisymmetric mode).

Considering the fluid ideal is a common approximation in the theory of the force acting on the particle, including the theory in Ref. 8 (the effect of the fluid viscosity on this force was studied by Doinikov[12,13]). The dropping of viscosity and other irreversible losses results in the infinite growth of the particle potential well depth as the frequency approaches the exact resonance frequency. However, in fluids with low viscosity, e.g., air, and at frequencies used in our experiments, its effect cannot significantly alter the geometry of the potential relief $U(\mathbf{r})$ and the qualitative features of the particles' concentration.

III. PARTICLE CONCENTRATION IN A CIRCULAR-CYLINDRICAL TUBE

We consider in this section the concentration of particles in a hollow piezoelectric cylinder. It is included into this presentation primarily for completeness. A diagram of this type of concentrator is shown in Fig. 1(a). R and R_o are the inner and outer radius, respectively. We assume the properties of the tube's wall to be uniform. The inner and outer surfaces are uniformly metallized to form electrodes. For this situation, the eigenmodes generated in the fluid are purely radial vibrations: acoustic (fluid) velocity, $\mathbf{v}$, and the velocity potential, Φ, depend on the radial variable r only. Consequently, the solution of the wave equation for the Fourier amplitude $\Phi(r,\omega)$ is proportional to the Bessel function $J_0(kr)$, where $k=\omega/s$, and the only component of the velocity is $v_r=d\Phi/dr\propto J_1(kr)$. The only component of the force, $\mathbf{F}$, acting on the particle is directed along the radius, $F=F_r(r)$. This force concentrates the particles in rings corresponding to the minima of the potential energy $U(r)$ [Eq. (3)].

The radial vibrations of the fluid in the cylindrical cavity are determined by the velocity amplitude of the oscillating transducer wall, v_0, and the proximity of the wall oscillation frequency f to the transducer resonance frequency f_{res}. At the inner wall (radius R) the velocities of the fluid and of the transducer surface are equal. Therefore, in the fluid, the potential and the velocity can be expressed as

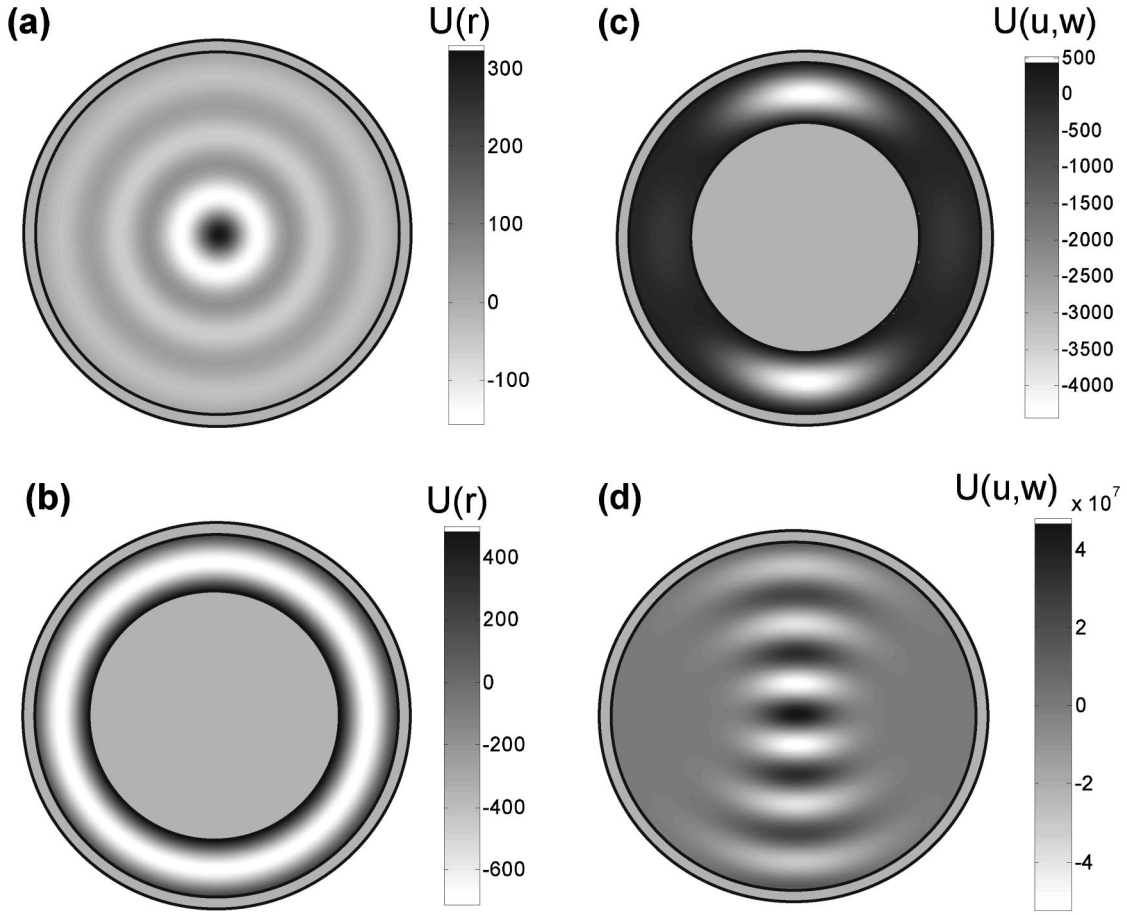


FIG. 2. Calculated potential energy of a particle in air under acoustical vibration of the device wall. (a) Circular cylinder, $R=0.845$ cm, $f=66.7$ kHz. (b) Circular cylinder ($R=1.96$ cm) with an inner circular cylindrical insertion ($R_2=0.61$ cm), $f=23.2$ kHz. (c) Circular cylindrical tube ($R=1.96$ cm) with a coaxial elliptic insertion inside ($a_x=1.265$ cm, $a_y=1.2$ cm), $f=23.2$ kHz. The number of cylindrical harmonics retained, $n_M=10$. (d) Hollow elliptic cylindrical tube ($a_x=0.9063$ cm, $a_y=0.8642$ cm, $a_y/a_x=0.954$), $f=63.06$ kHz (close to resonance).

$$\Phi(r,\omega)=-\frac{v_0}{k}\frac{J_0(kr)}{J_1(kR)},\quad v_r(r,\omega)=v_0\frac{J_1(kr)}{J_1(kR)}. \tag{4}$$

The potential energy of a spherical particle of radius R_p [Eq. (3)] equals

$$U(r)=U_0\frac{\frac{2}{3}J_0^2(kr)-J_1^2(kr)}{J_1^2(kR)}. \tag{5}$$

The depth of the potential wells where the particles are concentrated depends on the proximity to the resonance frequency of the piezoelectric cylinder and on the acoustic losses in the device (e.g., losses in the transducer, radiation into the air outside the transducer, losses in the air inside the transducer, losses due to open ends of the cylinder, etc.). Since we have not taken into account these losses, the numerical results for $U(r)$ show the geometry (relief) of the potential reasonably well but the depth of the potential wells only qualitatively. Figure 2(a) illustrates the calculated particle potential energy in a circular cylinder-transducer.

Figure 3(a) displays a PZT tube with the wavefield described by Fig. 2(a). Water vapor from a commercial ultrasonic humidifier is pumped into the tube's inner cavity and it is immediately transported to three nodal positions that form three concentric rings. The rings correspond to the potential energy minima positions described by Fig. 2(a). The tube has an inner radius of $R=0.845$ cm and is operating near its breathing mode resonance of 66.7 kHz. It is the same tube used in Figs. 3 and 4 of Ref. 1.

IV. PARTICLE CONCENTRATION IN A CYLINDRICAL TUBE WITH AN INNER COAXIAL TUNING ROD

It is instructive to note that it is possible to "tune" the internal geometry of the tube to a desired number of pressure nodes. This can be done by including a rigid, coaxial rod insert inside the cavity of the tube. For the present concentrator geometry, the fluid with particles flows in the annular space between the hollow piezoelectric cylinder's inner wall (radius R) and a coaxial, non-piezoelectric solid cylinder (rod) of radius $R_2<R$ as shown in Fig. 1(b). The radius of the inner rod may be machined to adjust the resonance frequency of the fluid (e.g., air) in the annular space to match the transducer resonance frequency f_{res}. In this device, the solutions to the equation for the Fourier amplitude of the function $\Phi(r,t)$ is a sum of two Bessel functions:

$$\Phi(r,\omega)=a_J J_0(kr)+a_Y Y_0(kr),\quad v_r=d\Phi(r)/dr. \tag{6}$$

At the surface of the inner cylinder, $r=R_2$, the radial velocity $v_r=0$ and at $r=R$ it equals $v_r=v_0$. From these boundary conditions we obtain the coefficients a_J and a_Y.

The amplitudes of the velocity potential and the velocity in the fluid between the two cylinders equal

$$\Phi(r,\omega)=\frac{v_0}{k}\frac{-Y_1(kR_2)J_0(kr)+J_1(kR_2)Y_0(kr)}{J_1(kR)Y_1(kR_2)-Y_1(kR)J_1(kR_2)},$$

$$v_r(r,\omega)=v_0\frac{Y_1(kR_2)J_1(kr)-J_1(kR_2)Y_1(kr)}{J_1(kR)Y_1(kR_2)-Y_1(kR)J_1(kR_2)}. \quad (7)$$

The dimensionless potential energy of the particles follows from Eqs. (3) and (7):

$$\frac{U(r)}{U_0}=\frac{1}{2}\frac{\frac{2}{3}[J_1(kR_2)Y_0(kr)-Y_1(kR_2)J_0(kr)]^2-[Y_1(kR_2)J_1(kr)-J_1(kR_2)Y_1(kr)]^2}{[J_1(kR)Y_1(kR_2)-Y_1(kR)J_1(kR_2)]^2}. \quad (8)$$

The resonance condition in the fluid between the two cylinders is satisfied when the denominator in Eqs. (7) and (8) is zero. For a transducer with inner radius $R=1.96$ cm operating at a breathing mode resonance of $f_{\text{res}}=23.2$ kHz, the resonance in air occurs at $R_2=1.224$ cm. The difference $R-R_2=0.7345$ cm is close to half wavelength in air at f_{res}. Figure 2(b) shows the calculated potential where one potential well between the inner wall of the tube and the coaxial insert exists. Experimental data for this configuration is given in Sec. IV of Ref. 1.

V. SYMMETRY BREAKING: ELLIPTICAL COAXIAL INSERT

For good spatial resolution of the region of concentrated aerosol, the process of symmetry breaking the cavity proves quite powerful. Up to this point, the piezoelectric tube has been driven in an axisymmetric structural mode coupled to an axisymmetric gas mode in the tube's air-filled cavity. The overall result has been that the particles are forced into a cross-sectional geometry defined by concentric rings. By symmetry breaking the geometry of the cavity it is possible to collapse the dimension of the collection ''rings'' down to localized geometries that are more representative of collection ''points.''

Consider a cylindrical PZT tube-transducer with a coaxial rigid body (insertion) inside. Let the cross section of the insertion be elliptic as shown in Fig. 1(c). The ellipse is determined by the equation

$$\frac{x^2}{a_x^2}+\frac{y^2}{a_y^2}=1, \quad (9)$$

where a_x and a_y are the half-axes of the ellipse ($a_x>a_y$). The problem we address here is how the particles, immersed in the fluid (e.g., air), are concentrated under vibrations of the cylindrical transducer wall in the presence of the elliptical insertion.

Since all quantities are independent of the coordinate z along the cylinder axis the problem is two-dimensional in the XY plane. The potential of the velocity $\Phi(x,y,\omega)$ in the fluid satisfies the 2D wave equation $\Delta\Phi+k^2\Phi=0$, where $k=\omega/s$. The boundary conditions are (1) at the transducer wall ($r=R$) the normal component of the velocity $v_r=\nabla_r\Phi$ is v_0 and is independent of the polar angle ϕ, and (2) the component of the velocity, $v_n=\nabla_n\Phi$, normal to the elliptical surface of the insertion is zero at its surface. The rigid body is acting as an acoustic mirror. The problem is not trivial due to different symmetry of the two boundaries: circular and elliptical. It is convenient to use, for the boundary condition at the surface of the insertion, the elliptical coordinates, u and w, which are connected with the polar coordinates, r and ϕ, by equations

$$r=a\sqrt{\cosh^2 u-\sin^2 w}, \quad \tan\phi=\tanh u\tan w. \quad (10)$$

The half-axes equal $a_x=a\cosh(u_0)$ and $a_y=a\sinh(u_0)$, where u_0 is the value of coordinate u that corresponds to the surface of the elliptical insertion, and $a=\sqrt{a_x^2-a_y^2}$ is the distance between the center of the ellipse and any of its two foci.

It is convenient to seek the velocity potential as an expansion in the products of Bessel functions, $J_M(kr)$ and $Y_M(kr)$, and functions $e^{iM\phi}$:

$$\Phi(r,\phi,\omega)=\sum_{M'=-\infty}^{+\infty}[a_{M'}J_{M'}(kr)+b_{M'}Y_{M'}(kr)]e^{iM'\phi}. \quad (11)$$

Multiplying the radial derivative of $\Phi(r,\phi,\omega)$ by $e^{-iM\phi}$ and integrating it over the polar angle ϕ from $\phi=0$ up to 2π yields the first boundary condition on radial velocity at $r=R$:

$$J'_M(kR)a_M+Y'_M(kR)b_M=(v_0/k)\delta_{M,0}. \quad (12)$$

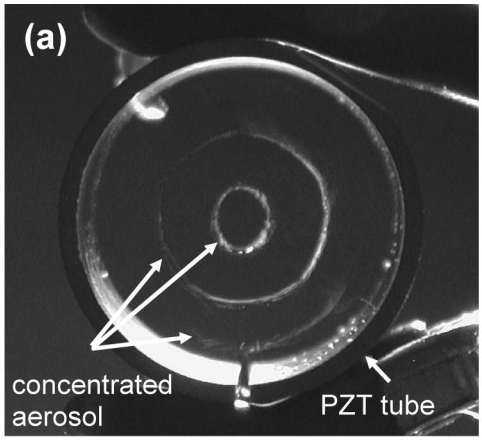
Here prime denotes differentiation with respect to the argument kr of the function. Using these equations we can eliminate the coefficients b_M:

$$b_M=\frac{-J'_M(kR)a_M+(v_0/k)\delta_{M,0}}{Y'_M(kR)}. \quad (13)$$

Substitution of this equation into Eq. (11) yields

$$\Phi(r,\phi,\omega)=\frac{v_0}{k}\left\{\frac{Y_0(kr)}{Y'_0(kR)}+\sum_{M=-\infty}^{+\infty}\frac{a_M}{v_0/k}\times\left[J_M(kr)-\frac{J'_M(kR)}{Y'_M(kR)}Y_M(kr)\right]e^{iM\phi}\right\}. \quad (14)$$

It is convenient, for numerical solution, to write the sum in Eq. (14) as a sum over non-negative numbers M only:



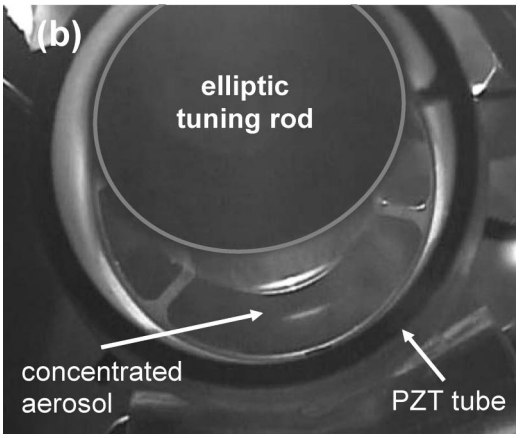


FIG. 3. (a) Water vapor is pumped into a PZT tube corresponding to the cavity design in Fig. 1(a). The aerosol is concentrated into three rings. (b) Water vapor is pumped into a PZT tube corresponding to the cavity design in Fig. 1(c). The aerosol is concentrated near a single nodal position along the minor axis of the elliptical insert.

$$\Phi(r,\phi,\omega)=\frac{v_0}{k}\left\{\frac{Y_0(kr)}{Y_0'(kR)}+\frac{1}{2}c_0\left[J_0(kr)-\frac{J_0'(kR)}{Y_0'(kR)}Y_0(kr)\right]+\sum_{M=1}^{\infty}\left[c_M\cos(M\phi)+s_M\sin(M\phi)\right]\times\left[J_M(kr)-\frac{J_M'(kR)}{Y_M'(kR)}Y_M(kr)\right]\right\}. \tag{15}$$

Here

$$c_M=\frac{a_M+(-1)^M a_{-M}}{v_0/k},\quad s_M=i\,\frac{a_M-(-1)^M a_{-M}}{v_0/k}. \tag{16}$$

In deriving Eqs. (15) and (16) we used the relationships $Z_{-M}=(-1)^M Z_M$, where Z_M is any Bessel function. The coefficient $s_0=0$.

The second boundary condition (at $u=u_0$) can be written as follows:

$$\mathrm{grad}_u\,\Phi=\left(\frac{1}{\sqrt{g_{uu}}}\frac{\partial\Phi}{\partial u}\right)_{u=u_0}=\frac{1}{\sqrt{g_{uu}(u_0)}}\left(\frac{\partial\Phi}{\partial r}\frac{\partial r}{\partial u}+\frac{\partial\Phi}{\partial\phi}\frac{\partial\phi}{\partial u}\right)_{u=u_0}=0. \tag{17}$$

Here, $g_{uu}=a^2(\sinh^2 u+\sin^2 w)$ is the component of the metric tensor. The derivatives $\partial r/\partial u$ and $\partial\phi/\partial u$ are easily calculated:

$$\left(\frac{\partial r}{\partial u}\right)_{u=u_0}=a\sqrt{\sinh^2(u_0)+\sin^2\phi},\qquad \left(\frac{\partial\phi}{\partial u}\right)_{u=u_0}=\frac{\sin(2\phi)}{\sinh(2u_0)}. \tag{18}$$

Substitution of these equations into Eq. (17) yields

$$a\sqrt{\sinh^2(u_0)+\sin^2\phi}\,\frac{\partial\Phi}{\partial r}+\frac{\sin(2\phi)}{\sinh(2u_0)}\frac{\partial\Phi}{\partial\phi}=0. \tag{19}$$

The condition $u=u_0$ couples the variables r and ϕ:

$$kr(\phi)=\frac{ka\cosh(u_0)}{\sqrt{1+\sin^2\phi/\sinh^2(u_0)}}. \tag{20}$$

After substitution of Eq. (15) and $r=r(\phi)$ the second boundary condition takes the form

$$\frac{1}{2}g_0(\phi)c_0+\sum_{M=1}^{\infty}\left[g_M(\phi)\cos(M\phi)-Mh_M(\phi)\times\sin(M\phi)\right]c_M+\left[g_M(\phi)\sin(M\phi)+Mh_M(\phi)\cos(M\phi)\right]s_M=f(\phi), \tag{21}$$

where

$$g_M(\phi)=ka\sqrt{\sinh^2 u_0+\sin^2\phi}\times\left[J_M'[kr(\phi)]-\frac{J_M'(kR)}{Y_M'(kR)}Y_M'[kr(\phi)]\right],$$

$$h_M(\phi)=\frac{\sin(2\phi)}{\sinh(2u_0)}\times\left[J_M[kr(\phi)]-\frac{J_M'(kR)}{Y_M'(kR)}Y_M[kr(\phi)]\right], \tag{22}$$

$$f(\phi)=-\sqrt{\sinh^2(u_0)+\sin^2\phi}\,\frac{Y_0'[kr(\phi)]}{Y_0'(kR)}.$$

In order to replace Eq. (21) by an algebraic equation we have to multiply Eq. (21) first by $\cos(M\phi)$ (with $M=0,...$), then by $\sin(M\phi)$ (with $M=1,...$) and integrate each of these two equations from $\phi=-\pi$ to $\phi=\pi$. The coefficients with c_M and s_M are, respectively, even and odd functions of ϕ. The function $f(\phi)$ is an even function of ϕ. We obtain two uncoupled system of equations, for c_M and s_M. The first is

$$\sum_{M'=0}^{\infty} d_{M,M'} c_{M'} = f_M \quad (M=0,1,...), \tag{23}$$

where

$$d_{M,0} = \frac{1}{\pi}\int_0^{\pi} d\phi g_0(\phi)\cos(M\phi),$$

$$d_{M,M'} = \frac{2}{\pi}\int_0^{\pi} d\phi [g_{M'}(\phi)\cos(M'\phi) - M' h_{M'}(\phi)\sin(M'\phi)]\cos(M\phi), \tag{24}$$

$$f_M = \frac{2}{\pi}\int_0^{\pi} d\phi f(\phi)\cos(M\phi).$$

It is an inhomogeneous linear system. The equations for s_M comprise a homogeneous system which has nonzero solutions only when the determinant of its matrix is zero. Therefore we may take $s_M = 0$.

In the numerical calculation only $n_M \gg 1$ numbers M are retained $(M, M' = 0,1,...,n_M - 1)$ in Eq. (23) for c_M. The found c_M should be substituted into Eq. (15):

$$\Phi(r,\phi,\omega) = \frac{v_0}{k}\left\{\frac{Y_0(kr)}{Y_0'(kR)} + \frac{1}{2}c_0\left[J_0(kr) - \frac{J_0'(kR)}{Y_0'(kR)}Y_0(kr)\right] + \sum_{M=1}^{\infty} c_M \cos(M\phi) \times \left[J_M(kr) - \frac{J_M'(kR)}{Y_M'(kR)}Y_M(kr)\right]\right\}. \tag{25}$$

The velocity components of the fluid equal

$$v_r = v_0\left\{-\frac{Y_1(kr)}{Y_0'(kR)} - \frac{1}{2}\left[J_1(kr) - \frac{J_0'(kR)}{Y_0'(kR)}Y_1(kr)\right]c_0 + \sum_{M=1}^{\infty}\cos(M\phi)\left[J_M'(kr) - \frac{J_M'(kR)}{Y_M'(kR)}Y_M'(kr)\right]c_M\right\}, \tag{26}$$

$$v_\phi = -v_0\frac{1}{r}\sum_{M=1}^{\infty} M c_M \sin(M\phi) \times \left[J_M(kr) - \frac{J_M'(kR)}{Y_M'(kR)}Y_M(kr)\right].$$

Using these equations and Eq. (3) one can find the potential energy of a particle at any point between the transducer wall and the elliptic insertion.

The results are presented in Fig. 2(c). Due to broken circular symmetry, the particles are concentrated in two narrow wells. The deepest parts of these wells have the form of symmetric arcs, each of which corresponds to a rather small polar angle. Remarkably, such drastic change of the potential energy geometry results from a small difference between the ellipse axes: $(a_x - a_y)/a_x = 0.051$. Figure 3(b) displays experimental results for the concentrator with elliptic insertion. The tube has the same properties as the tube described in the calculations leading up to Fig. 2(c).

VI. ELLIPTIC CAVITY CONCENTRATOR

Consider a hollow elliptic cylinder of piezoelectric ceramics. Let a_x and a_y be the half-axes of the ellipse $(a_x > a_y)$ as shown in Fig. 1(d). The parameter $a = \sqrt{a_x^2 - a_y^2}$. In elliptic coordinates $u\,(0 < u < \infty)$ and $w\,(0 < w < 2\pi)$, $a_x = a\cosh(u_0)$, $a_y = a\sinh(u_0)$, i.e., $u = u_0$ corresponds to the ellipse boundary. The wave equation for the velocity potential, $\Delta\Phi + k^2\Phi = 0$, takes the following form in the coordinates of an elliptic cylinder:

$$\frac{\partial^2\Phi(u,w,\omega)}{\partial u^2} + \frac{\partial^2\Phi(u,w,\omega)}{\partial w^2} + 4q[\sinh^2 u + \sin^2 w]\Phi(u,w,\omega) = 0, \tag{27}$$

where $q = (ka)^2/4$. The variables can be easily separated. The solutions to Eq. (27) are products of two Mathieu functions, one depending on w, the second on u.[14,15] In the system under consideration the velocity of the vibrating piezoelectric wall (amplitude v_0) is, at any point, perpendicular to the surface $u = u_0$, and the vibrations of all points are synchronous (common phase). Therefore, the dependence of the solution on w must be periodic. We must choose between four types of periodical Mathieu functions of w which differ by the parity and the period, π or 2π. The two components of the velocity are (the frequency argument is omitted)

$$v_w(u,w) = \nabla_w\Phi(u,w) = \frac{1}{a\sqrt{\sinh^2 u + \sin^2 w}}\frac{\partial\Phi(u,w)}{\partial w},$$
$$v_u(u,w) = \nabla_u\Phi(u,w) = \frac{1}{a\sqrt{\sinh^2 u + \sin^2 w}}\frac{\partial\Phi(u,w)}{\partial u}. \tag{28}$$

The velocity $v_u(u,w)$ must be the same at the opposite points of the ellipse. It means that the period of $v_u(u,w)$ in w is π and the dependence of $\Phi(u,w)$ must be even and periodic in w with period π. The solutions to Eq. (27) that satisfy these symmetry requirements are the products of Mathieu functions: $ce_{2m}(w,q)Ce_{2m}(u,q)$, where $2m = 0,2,4,...$. They are solutions to the equations

$$\frac{d^2 ce_{2m}}{dw^2} + [h_{2m} - 2q\cos(2w)]ce_{2m}(w,q) = 0,$$
$$\frac{d^2 Ce_{2m}}{du^2} - [h_{2m} - 2q\cosh(2u)]Ce_{2m}(u,q) = 0. \tag{29}$$

The functions $ce_{2m}(w)$ are even and periodic, with period π, only at definite characteristic values, h_{2m}, of the separation parameter. The corresponding modified Mathieu function, $Ce_{2m}(u,q)$, is obtained by replacing the argument w in $ce_{2m}(w,q)$ with iu where i is the imaginary unity: $Ce_{2m}(u,q) = ce_{2m}(iu,q)$.

The solution which satisfies the boundary condition, $v_u(u_0,w) = v_0$, is sought as a series:

$$\Phi(u,w) = v_0 a\sum_{m=0}^{\infty} a_m ce_{2m}(w,q)Ce_{2m}(u,q), \tag{30}$$

where a_m are dimensionless coefficients. The velocity normal to the transducer elliptic wall is given by the equation:

$$\nabla_u \Phi(u,w) = \pm \frac{1}{a\sqrt{\sinh^2 u + \sin^2 w}} \frac{\partial \Phi(u,w)}{\partial u}. \tag{31}$$

The boundary condition is therefore

$$\sum_{m=0}^{\infty} a_m \frac{ce_{2m}(w,q)}{\sqrt{\sinh^2(u_0)+\sin^2 w}} \frac{dCe_{2m}(u,q)}{du}\bigg|_{u=u_0} = 1. \tag{32}$$

Let us multiply both parts of this equation by $ce_{2n}(w,q)\sqrt{\sinh^2(u_0)+\sin^2(w)}$ and integrate over the period from $w=0$ to $w=\pi$. Due to the orthogonality of the functions $ce_{2m}(w,q)$ and $ce_{2n}(w)$ with $n \neq m$ (in the interval $[0,\pi]$)[14] and to their normalization to π in the interval $[0,2\pi]$[15] we obtain

$$a_m = \frac{2}{\pi} \frac{\int_0^{\pi} dw ce_{2m}(w,q)\sqrt{\sinh^2(u_0)+\sin^2 w}}{[dCe_{2m}(u,q)/du]_{u_0}}. \tag{33}$$

The components of the fluid velocity at any point (u,w) are

$$\begin{aligned} v_w(u,w) &= v_0 \frac{\Sigma_{m=0}^{\infty} a_m ce'_{2m}(w,q) Ce_{2m}(u,q)}{\sqrt{\sinh^2 u + \sin^2 w}}, \\ v_u(u,w) &= v_0 \frac{\Sigma_{m=0}^{\infty} a_m ce_{2m}(w,q) Ce'_{2m}(u,q)}{\sqrt{\sinh^2 u + \sin^2 w}}. \end{aligned} \tag{34}$$

The primes stand for differentiation with respect to the argument.

The potential energy of a spherical particle inside the elliptical transducer-container equals [see Eq. (3)]:

$$U(u,w) = U_0 \left\{ \frac{8q}{3} \left(\frac{\Phi(u,w)}{v_0 a} \right)^2 - \frac{v_u^2(u,w)+v_w^2(u,w)}{v_0^2} \right\}. \tag{35}$$

Here $U_0 = \pi \rho R_p^3 v_0^2$ (see Sec. II).

At small $w \to 0$ and $u \to 0$ the Mathieu functions are, respectively,

$$\begin{aligned} ce_{2m}(w,q) &\approx ce_{2m}(0,q)[1 - \tfrac{1}{2}(h_{2m} - 2q)w^2], \\ Ce_{2m}(u,q) &\approx Ce_{2m}(0,q)[1 + \tfrac{1}{2}(h_{2m} - 2q)u^2]. \end{aligned} \tag{36}$$

Since $Ce_{2m}(0,q) = ce_{2m}(0,q)$ we have also $Ce'_{2m}(0,q) = 0$. The parameter h_{2m} and the function $ce_{2m}(w)$ can be found numerically. As h_{2m} is known, the function $Ce_{2m}(u)$ is easily found by integration of the second Eq. (29).

No divergence occurs in Eq. (34) at $u=0$, $\sin w = 0$ (these points correspond to the foci of the ellipse: $x = \pm a$, $y=0$). Using the equation $ce_{2m}(0,q) = Ce_{2m}(0,q)$ one obtains

$$U(0,0) = U_0 \sum_{m=0}^{\infty} a_m [ce_{2m}(0,q)]^4 [8q/3 - (h_{2m} - 2q)^2]. \tag{37}$$

We now discuss some results of numerical calculations shown in Fig. 2(d). The circular symmetry of the valleys (rings) characteristic for circular cylinders [Sec. III and Figs. 2(a) and 3(a)] is significantly altered even when the ratio $a_y/a_x = 0.95$, i.e., close to unity. This slight alteration from circular symmetry collapses the equipotential rings observed in Fig. 2(a) into symmetrically located wells along the OY axis. The particles are thus concentrated within small areas around each of the symmetrical minima on the OY axis ($w = \pi/2$). At small frequencies when the wavelength $\lambda = s/f$ is of the same order as the dimensions of the ellipse, there is only one potential valley in each direction. Its absolute minimum is always at $w = \pi/2$ (OY axis). One can see the strong effect of the ellipse eccentricity (circular symmetry breaking) and its ability for localization of the potential wells within the cavity.

VII. SUMMARY AND CONCLUSIONS

The introduction of hollow piezoelectric transducer-containers opens up new vistas in the technique of particle concentration in fluids by acoustic vibrations. By solving the problems of concentration in a cylindrical transducer with an elliptic insertion (Sec. V) and in an elliptic transducer (Sec. VI) we have shown that the relief of the particle's energy in the acoustic field is extremely sensitive to the symmetry breaking of the device. The circular energy valleys (wells) and rings of concentrated particles in devices with circular symmetry are reduced effectively to localized symmetric valleys occupying small areas in the case of elliptic symmetry even if the ellipse eccentricity is only several percent (Secs. V and VI). This favors the concentration of particles in much smaller volumes of the fluid, increasing the efficiency of the device.

The calculations presented above show that the number of potential energy minima grows with frequency. The most convenient work frequency for concentration of particles is the lowest resonance when the sound wavelength is of the same order as the radius of the piezoelectric transducer-container. The calculations are in satisfactory qualitative agreement with experiments (more detailed experimental data will be published elsewhere). This agreement demonstrates, in addition, the reliability of the basic theory of particle concentration by acoustic vibrations.

ACKNOWLEDGMENTS

This work was partially financed by the U.S. Department of Defense and Laboratory Directed Research and Development Funds at Los Alamos National Laboratory.

[1] G. Kaduchak, D. N. Sinha, and D. C. Lizon, "Novel cylindrical, air-coupled levitation/concentration device," Rev. Sci. Instrum. **73**, 1332–1336 (2002).
[2] G. Kaduchak and D. N. Sinha, "Low Power Acoustic Harvesting of Aerosols," IEEE Ultrasonics Symposium, (2001), pp. 607–610.
[3] G. Kaduchak and D. N. Sinha, "Cylindrical Acoustic Levitator/Concentrator," U.S. Patent No. 6,467,350 (2001).
[4] G. Kaduchak and D. N. Sinha, "Cylindrical Acoustic Levitator/Concentrator Having Non-Circular Cross-Section," U.S. Patent No. 6,644,118 (2003).
[5] L. V. King, "On the acoustic radiation pressure on spheres," Proc. R. Soc. London, Ser. A **147**, 212 (1934).
[6] K. Yosioka and Y. Kawasima, "Acoustic Radiation Pressure on a Compressible Sphere," Acustica **5**, 167 (1955).

[7] T. F. W. Embleton, ''Mean force on a sphere in a spherical sound field,'' J. Acoust. Soc. Am. **26**, 40 (1954).
[8] L. P. Gor'kov, ''On the forces acting on a small particle in an acoustical field in an ideal fluid,'' Sov. Phys. Dokl. **6**, 773 (1962).
[9] W. L. Nyborg, ''Radiation Pressure on a Small Rigid Sphere,'' J. Acoust. Soc. Am. **42**, 947–952 (1967).
[10] M. Barmatz and P. Collas, ''Acoustic radiation potential on a sphere in plane, cylindrical, and spherical standing wave fields,'' J. Acoust. Soc. Am. **77**, 928 (1985).
[11] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon, Oxford, 1987).
[12] A. A. Doinikov, ''Acoustic radiation force on a spherical particle in a viscous heat-conducting fluid. I. General formula,'' J. Acoust. Soc. Am. **101**, 713–721 (1997).
[13] A. A. Doinikov, ''Acoustic radiation force on a spherical particle in a viscous heat-conducting fluid. II. Force on a rigid sphere,'' J. Acoust. Soc. Am. **101**, 722–730 (1997).
[14] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi. *Higher Transcendental Functions* (McGraw–Hill, New York, 1955), Vol. 3.
[15] N. W. McLachlan, *Theory and Applications of Mathieu Functions* (Clarendon, Oxford, 1947).

Examination of coherent surface reflection coefficient (CSRC) approximations in shallow water propagation

Kevin L. Williams,[a] Eric I. Thorsos, and W. T. Elam
Applied Physics Laboratory, College of Ocean and Fishery Sciences, University of Washington, 1013 N. E. 40th Street, Seattle, Washington 98105

(Received 30 December 2003; revised 23 June 2004; accepted 1 July 2004)

The parabolic wave equation (PE) code of Rosenberg [J. Acoust. Soc. Am. **105**, 144–153 (1999)] is used as a benchmark to study acoustic propagation in an ocean waveguide with a rough air/water interface. The PE results allow a close examination of the ability of a ray code [i.e., Gaussian RAy Bundle (GRAB)] to accurately estimate coherent field propagation using a coherent reflection coefficient derived from scattering theory. Comparison with PE implies that the Beckmann–Spizzichino model, as given within the GRAB software package, does not give accurate predictions of the coherent field at long ranges. Three other coherent reflection coefficient approximations are tested: the perturbation, the small slope, and the Kirchhoff approximations. The small slope approximation is the most accurate of the models tested. However, the Kirchhoff approximation is perhaps accurate enough for some purposes and would be simpler to implement as a module within GRAB. [DOI: 10.1121/1.1785617]

PACS numbers: 43.30.Cq, 43.30.Dr, 43.30.Hw [SLB] Pages: 1975–1984

I. INTRODUCTION

Ray-based methods are a common choice for modeling acoustic propagation in the ocean when processing speed is essential. Often it is the *coherent* field that is of interest and losses in the coherent field as it interacts with the rough ocean surface or bottom must be addressed. One strategy is to use scattering theory to determine coherent reflection coefficients that are then incorporated into the ray calculations. Accurate determination of these coherent reflection coefficients for shallow grazing angles is essential if one attempts to determine long range, shallow water propagation via ray codes.

Here we examine acoustic propagation in a shallow water waveguide with a rough air/water interface and flat water/sediment interface. The 1-D rough surface parabolic wave equation (PE) code of Rosenberg[1] is used as a benchmark to test the use of Coherent Surface Reflection Coefficients (CSRCs) in the Gaussian RAy Bundle (GRAB) code.[2] CSRCs are determined using three different scattering approximations: the small slope, the perturbation, and the Kirchhoff approximations. In addition, the ''Beckmann–Spizzichino'' model for CSRC, currently part of the GRAB software package, is also used. (It should be noted that the Beckmann–Spizzichino model as defined within GRAB is actually a combination of a theoretical model and empirical results.) The goal is to determine which, if any, allow a ray-based propagation model to accurately determine the coherent field within the waveguide for acoustic frequencies in the 1–10 kHz range (the so-called mid-frequency range).

The rough surface PE code and the small slope and perturbation scattering approximations require knowledge of the surface wave spectrum. We use a simplified version of an isotropic 2-D Pierson–Moskowitz spectrum which leads to a convenient analytic form for a 1-D roughness spectrum for use with the PE propagation model. Individual surface wave realizations based on this 1-D spectrum are used to determine PE results. For consistency the same 1-D spectrum is used with the scattering approximations in GRAB. The PE results for the propagated field are coherently averaged before comparison with the inherently ensemble averaged results generated by GRAB when a CSRC is used.

The particular form of a surface wave spectrum used in this work has been chosen for convenience in making PE–GRAB comparisons. The spectrum gives a rough approximation to fully developed sea surface wave spectra, and thus should be an adequate model for our purposes. However, since 1-D roughness spectra are necessary for these comparisons, our spectrum model should not be taken to generally apply to ocean conditions. It must be emphasized that when modeling propagation in the ocean, a 2-D roughness spectrum should be used to determine the perturbation or small slope approximations for the CSRC.

Section II describes some of the salient points of the codes being used. Section III compares PE and GRAB results for a 50-m-deep shallow water waveguide out to ranges of 20 km when both the air/water and water/sediment interfaces are flat. These results form a background against which to compare the results for rough air/water interfaces. Section IV gives details on the surface wave spectrum used and Sec. V gives details on the CSRC models. Sections VI and VII compare results derived from PE and GRAB in order to determine which of the CSRCs should be used in ray-based codes. Section VIII summarizes.

II. PROPAGATION CODES

The two propagation codes used in the comparison here are representative of separate classes of propagation models, either wave or ray based. The PE code[1] is a wave propagation model that includes the interaction of the wave field

[a]Electronic mail: williams@apl.washington.edu

with a single realization of a rough ocean surface. GRAB[2] is a ray-based code that produces results that must be compared with an average of PE results over an ensemble of rough surface realizations. Below, the salient points of each model relevant to the present effort are summarized. See Refs. 1 and 2 for further details.

A. Gaussian ray bundle (GRAB)

GRAB associates a "Gaussian ray bundle" of the form

$$\Psi = \frac{\beta_0 \Gamma^2}{\sqrt{2\pi}\sigma p_r r} \exp\left\{ -0.5\left[\frac{(z - z_\nu)}{\sigma} \right]^2 \right\} \tag{1}$$

with each ray traced. In Eq. (1) Ψ denotes the pressure field associated with a ray, β_0 depends on the source, z_ν is the depth (at the horizontal range r being examined) of the νth ray traced, p_r is the horizontal slowness, σ is a half-beamwidth and, most important here, Γ includes losses due to volume attenuation and boundary reflections. The half-beamwidth σ is given by

$$\sigma = \tfrac{1}{2}\max(\Delta z, 4\pi\lambda), \tag{2}$$

where Δz is the change in ray depth at a constant range due to a change in source angle $\Delta\theta_o$, where $\Delta\theta_o$ is the increment in source angle used, and where λ is the wavelength at the field point. The choice implied in Eq. (2) is used to prevent infinite field values near caustics, and was developed empirically.[2] The factor β_0 in Eq. (1) is chosen such that the energy within a Gaussian ray bundle equals the energy within a geometric-acoustic ray tube. One way to implement geometric acoustic calculations is to identify the test ray that is closest to the field point as an eigenray and use it (along with adjacent rays to determine spreading loss) to calculate received pressure. In the Gaussian ray bundle approach all bundles (there is a bundle associated with each ray traced) contribute at all depths but their amplitudes are weighted by the Gaussian function in Eq. (1). At a given range, all ray bundles (note again that there is a ray bundle associated with each ray traced) that have experienced the same number of bottom and surface reflections are combined and interpreted as an acoustic eigenray. Test cases shown in Ref. 2 indicate that this method can be quite accurate.

In the case of scattering from the rough ocean surface the goal is to determine, if possible, a coherent reflection coefficient to incorporate into Γ that allows the acoustic propagation of the coherent field to be accurately approximated. In what follows, accuracy is tested via comparisons with ensemble-averaged results of the complex wave field from the rough surface PE of Rosenberg. The comparison is facilitated by the fact that GRAB allows user-defined tables of the reflection coefficient (in dB) as a function of angle. Therefore, different models can be used to generate these tables and the results inserted without alteration of the GRAB software package.

B. Rough surface PE

The rough surface PE code of Rosenberg[1] is an extension of the (flat surface) wide angle PE code RAM developed by Collins.[3] In order to account for the effects of a rough sea surface, the vertical grid used to represent the field at a given range is divided into two sections: a fine grid that extends beyond the maximum range of surface height variation, and a coarser grid elsewhere. The rough surface realization obtained from the 1-D surface spectrum is then converted to a stair-stepped surface by reassigning the surface location to be the nearest grid point on the fine grid. The acoustic boundary condition on the sea surface is enforced by setting the pressure to zero on the grid points that define the rough surface. Comparisons with integral equation results show that this rough surface PE method yields accurate results for forward scattering from typical sea surface roughness.[1]

For the PE results shown, the coarse grid spacing was 0.025 m and the fine grid extended from 5 m above the mean surface to 10 m below (into the water) and had a grid spacing smaller by a factor of 2. The range step was 1 m. These increments were determined by making depth grid spacing and range step successively smaller until no significant difference in the pressure field was found between runs.

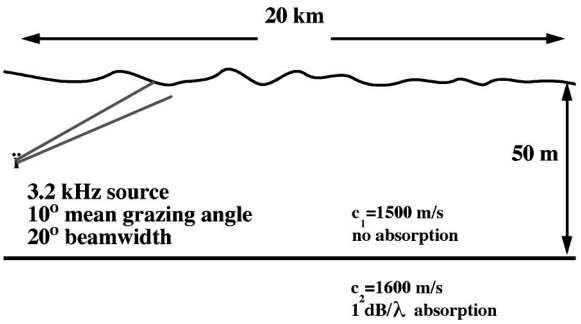


FIG. 1. Waveguide geometry, environmental parameters, and source characteristics used for this study. The source has a Gaussian beampattern with a 20° beamwidth. The center of the beam is aimed upward 10° above horizontal. All results except those of Sec. VII use a 3.2-kHz source. In Sec. VII a 6-kHz source is used.

III. RAY MODEL/PE COMPARISONS FOR FLAT SURFACES

Before proceeding to GRAB/PE comparisons for rough ocean surfaces it is necessary to develop a baseline using flat interfaces. The problem to be addressed is shown in Fig. 1. For the present case the ocean surface and bottom are assumed to have flat interfaces. An acoustic source operating at 3.2 kHz is located mid-water-column in a 50-m waveguide. The source has a Gaussian beampattern with a 20° beamwidth and the center of the beam is aimed upward 10° above horizontal. The sound speed in the water is 1500 m/s and it is assumed that there is no attenuation. The sound speed in the sediment is 1600 m/s, the sand/water density ratio is 2, and the sediment absorption is 1 dB/wavelength.

A primary requirement is the production of equivalent starting fields in the PE and GRAB calculations. It was found that, in order to eliminate diffraction effects occurring in the initial PE propagation of the Gaussian source, it sufficed to use the field produced by ray theory at a horizontal range of 40 m from the source as the starting PE field. Figure 2 shows a comparison of the intensity profile at 500 m for GRAB [Fig. 2(a)], PE [Fig. 2(b)], and GRAB and PE overlaid on

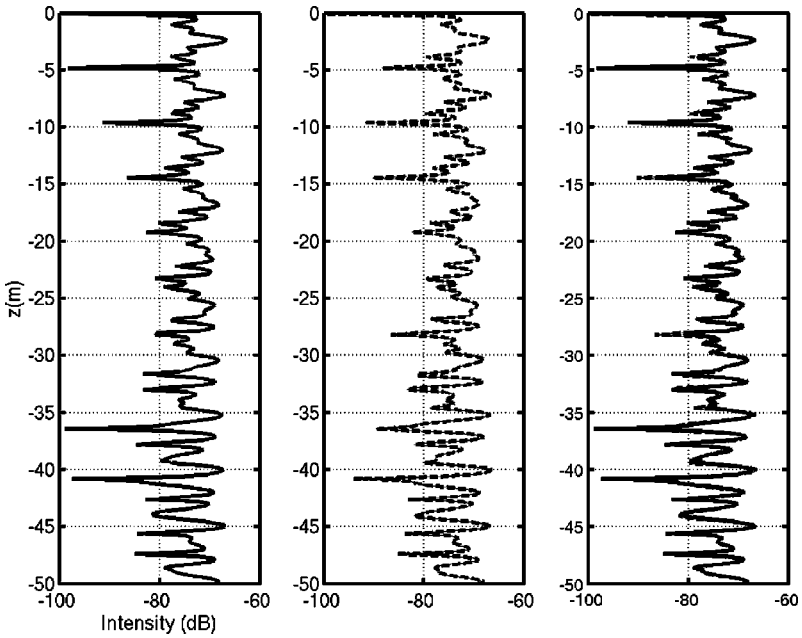


FIG. 2. The intensity profile at 500-m range for the waveguide of Fig. 1, assuming both the ocean bottom and surface are flat: (a) GRAB, (b) PE, and (c) overlay of GRAB and PE. The agreement between models is very good with the main differences being in the regions of the deep nulls.

each other [Fig. 2(c)]. In both cases the source level was set to 1 μPa rms (0 dB). The agreement is very good with the main differences being in the regions of the deep nulls.

Figure 3 shows intensity profiles for several more ranges in increments of 2.5 km out to 20 km. The comparisons shown do not change significantly if smaller angular increments or more stringent ''eigenray tolerances'' (that is, inclusion of lower amplitude paths) are used in GRAB. The general trend is for the GRAB/PE agreement to deteriorate with range. The remaining question of which code is correct was answered by use of a mode equation solution[4] for the same problem. The mode equation and PE results are indistinguishable at 20 km. Thus, PE equation results are taken as ground truth for the remainder of this paper.

A different type of comparison is appropriate to examine how well GRAB determines average transmission loss in the waveguide. Figure 4(a) shows $10\log_{10}$ of the mean intensity (i.e., the average intensity computed over a vertical slice) as a function of range. Again PE and GRAB agreement deteriorates as the range increases but the maximum difference [Fig. 4(b)] is less than 1 dB for the range plotted. The difference in Fig. 4(b) must be taken into account in determining which CSRC model is the most accurate when the ocean surface is rough.

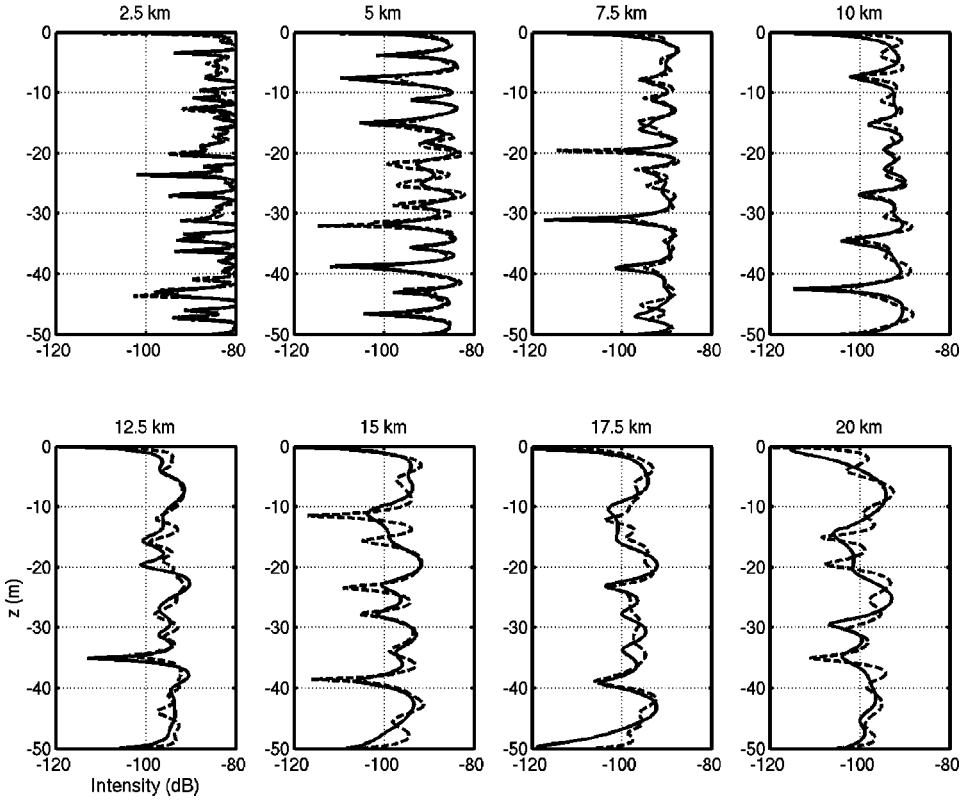


FIG. 3. Comparison of GRAB (solid lines) and PE (dashed lines) intensity profiles for ranges from 2.5 to 20 km. Interfaces are assumed to be flat. The general trend is for the GRAB/PE agreement to deteriorate with range. Comparisons with mode equation results imply that PE results can be considered ground truth.

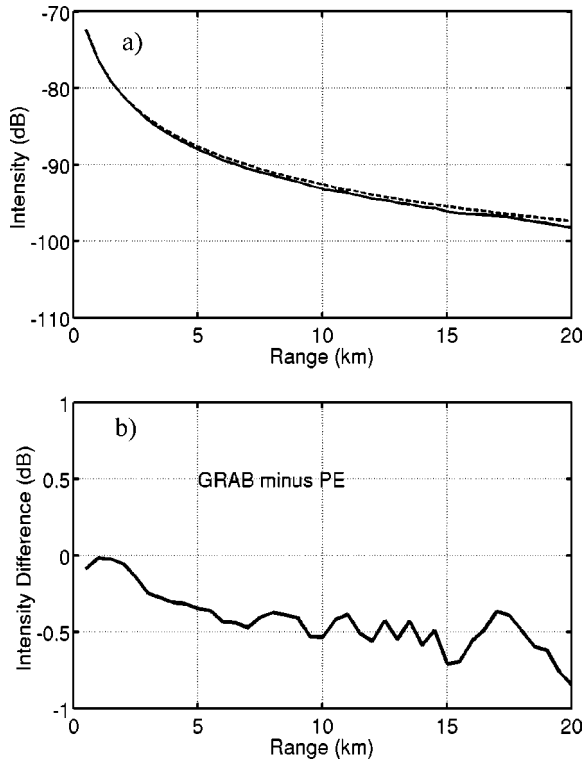


FIG. 4. (a) PE (dashed line) and GRAB (solid line) mean intensities (i.e., $10\log_{10}$ of the average intensity computed over a vertical slice) as a function of range. As in Fig. 3, PE and GRAB agreement deteriorates as the range increases. (b) Difference between the PE and GRAB curves in Fig. 4(a). The maximum difference is less than 1 dB for the range plotted.

IV. SIMPLE OCEAN SURFACE WAVE SPECTRUM MODEL

The 1-D spectrum used here is derived from a simplified version of the 2-D Pierson–Moskowitz surface wave spectrum for a fully developed sea.[5] (The effects of swell or limited fetch could be easily obtained in the same manner as done here, assuming an appropriate spectral representation of surface roughness was defined.)

The isotropic 2-D Pierson–Moskowitz spectrum can be written as[5]

$$W(\mathbf{K})=\left[\frac{\alpha}{4\pi K^4}\right]\exp\left(-\frac{K_L^2}{K^2}\right), \tag{3}$$

where $\mathbf{K}$ is the surface wave vector, $K=|\mathbf{K}|=\sqrt{K_x^2+K_y^2}$, $K_L=(\sqrt{\beta}g)/U^2$, $g=9.81\,\mathrm{m/s^2}$, U is the wind speed at 19.5 m above the mean water interface, $\alpha=0.0081$, and $\beta=0.74$. To obtain an analytical form for the 1-D spectrum, it is convenient to simplify this somewhat further and make the low wavenumber roll-off a hard cutoff at $K=K_L$:

$$W(\mathbf{K})=0 \quad \text{for } K<K_L, \quad W(\mathbf{K})=\frac{\alpha}{4\pi K^4} \quad \text{for } K>K_L. \tag{4}$$

The spectrum is normalized such that

$$\int W(\mathbf{K})d\mathbf{K}=h^2, \tag{5}$$

where h is the rms wave height. When Eq. (4) is substituted into Eq. (5) one obtains

$$h^2=\frac{\alpha}{4K_L^2}. \tag{6}$$

This result is the same as would be derived using the full Pierson–Moskowitz spectrum [Eq. (3)].

There is not a unique prescription for defining a 1-D roughness spectrum from a 2-D spectrum for application to scattering problems. Here we define the 1-D spectrum as

$$W(K_x)=\int W(\mathbf{K})dK_y. \tag{7}$$

Surface realizations obtained from this 1-D spectrum will correspond to 1-D cuts through 2-D surface realizations obtained from the 2-D spectrum. (This prescription differs from that used in Ref. 5.) Substitution of Eq. (4) into Eq. (7) gives

$$\begin{aligned} W(K_x)&=\left[\frac{\alpha}{8|K_x|^3}\right]F(K_x) \quad \text{for } |K_x|<K_L, \\ W(K_x)&=\left[\frac{\alpha}{8|K_x|^3}\right] \quad \text{for } |K_x|>K_L, \end{aligned} \tag{8}$$

with

$$F(K_x)=\frac{2}{\pi}\left[\arcsin\left(\frac{|K_x|}{K_L}\right)-\frac{|K_x|\sqrt{K_L^2-K_x^2}}{K_L^2}\right]. \tag{9}$$

Equations (8) and (9) were used to determine individual realizations of the surface wave field[6] for the PE calculations and to calculate the CSRC for the small slope and perturbation approximations. Equation (6) was also used in determining both the small slope and Kirchhoff approximation coherent reflection losses.

V. COHERENT SURFACE REFLECTION COEFFICIENT (CSRC) MODELS

GRAB provides several options for a reflection coefficient model for the surface of the ocean. Choices include empirically derived models, a combination of empirical and theoretical results called the ''Beckmann–Spizzichino'' model within the GRAB software package for loss due to scattering from the rough air/water interface, and a model for loss due to bubble absorption. The one of interest in the present study is the Beckmann–Spizzichino model. PE/GRAB comparisons in the next section will include results using this model.

An alternative approach is to calculate a CSRC (due to scattering from a 1-D surface) starting with the definition [Eq. (B6), Ref. 5]

$$\langle T(k_{sx},k_{ix})\rangle=R_A(k_{ix})\,\delta(k_{ix}-k_{sx}), \tag{10}$$

where T is the transition matrix (or T matrix) quantifying scattering of an incident plane wave into a particular scattering direction, k_{ix} and k_{sx} are the horizontal components of the incident and scattered wave vectors, respectively, R_A is identified as the CSRC, and δ is the delta function. Below, the basic equations are given for the CSRC derived using the

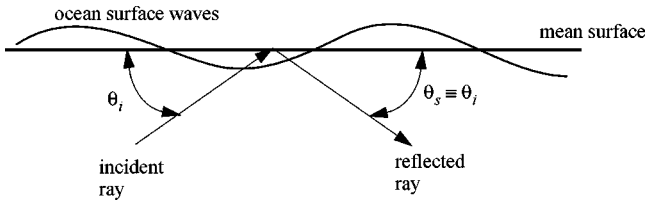


FIG. 5. Definition of incident grazing angle θ_i and angle of reflection $\theta_s \equiv \theta_i$ used in determining some of the parameters of the small slope approximation.

small slope, Kirchhoff, and perturbation approximations. The CSRC derivable from the Kirchhoff approximation is identifiable as one term of the small slope result. The CSRC derived using perturbation theory is obtainable from the small slope result via an expansion of an exponential for small arguments.

The first step is to approximate $\langle T \rangle$ using the first two terms of the small slope approximation,

$$\langle T \rangle \approx \langle T_0 \rangle + \langle T_1 \rangle, \tag{11}$$

and then obtain R_A from Eq. (10). This approximation is good to second order in surface slope. Details of the derivation of T_0 and T_1 as well as the meaning of second order in slope are given in Ref. 7.

The terms $\langle T_0 \rangle$ and $\langle T_1 \rangle$ are given in Eqs. (10) and (11) of Ref. 8 and reproduced here in a slightly different form:

$$\begin{aligned} \langle T_0 \rangle &= -\exp\left[-\frac{\nu_z^2 h^2}{2}\right]\delta(k_{sx}-k_{ix}), \\ \langle T_1 \rangle &= \kappa_{iz}\exp\left[-\frac{\nu_z^2 h^2}{2}\right]\delta(k_{sx}-k_{ix})\int dK_1\, W(K_1) g(K_1). \end{aligned} \tag{12}$$

In Eq. (12) the following definitions apply:

$$\begin{aligned} &\kappa_{iz} = \sqrt{k^2 - k_{ix}^2}, \quad \nu_z = \kappa_{iz} + k_{sz}, \\ &g(K_1) = k\beta_{i+1} + k\beta_{s-1} - \nu_z, \\ &k\beta_{i+1} = [k^2 - (K_1 + k_{ix})^2]^{1/2}, \quad \mathrm{Im}(\beta_{i+1}) > 0, \\ &k\beta_{s-1} = [k^2 - (k_{sx} - K_1)^2]^{1/2}, \quad \mathrm{Im}(\beta_{s-1}) > 0, \end{aligned} \tag{13}$$

and k is the wave number in the water. In addition, Eq. (8) defines W and Eq. (6) defines h. Using Fig. 5 the variables in Eq. (13) can be written in terms of incident and scattered grazing angles and using Eqs. (10)–(12) the CSRC can be determined. The result is

$$R_A(k_{ix}) \approx \exp\left[-\frac{\nu_z^2 h^2}{2}\right]\left[-1 + \kappa_{iz}\int dK_1\, W(K_1) g(K_1)\right], \tag{14}$$

with

$$k_{ix} \equiv k_{sx} = k\cos\theta_i, \quad \kappa_{iz} \equiv |k_{sz}| = k\sin\theta_i, \quad \nu_z = 2k\sin\theta_i. \tag{15}$$

Small slope results shown in the next section used $20\log_{10}|R_A|$ with R_A given by Eq. (14) to calculate a CSRC in dB (as needed by GRAB); a π phase shift upon reflection is also incorporated into the GRAB calculations. The first term on the right-hand side of Eq. (14) is the result for R_A obtained in the Kirchhoff approximation.[5] Kirchhoff results

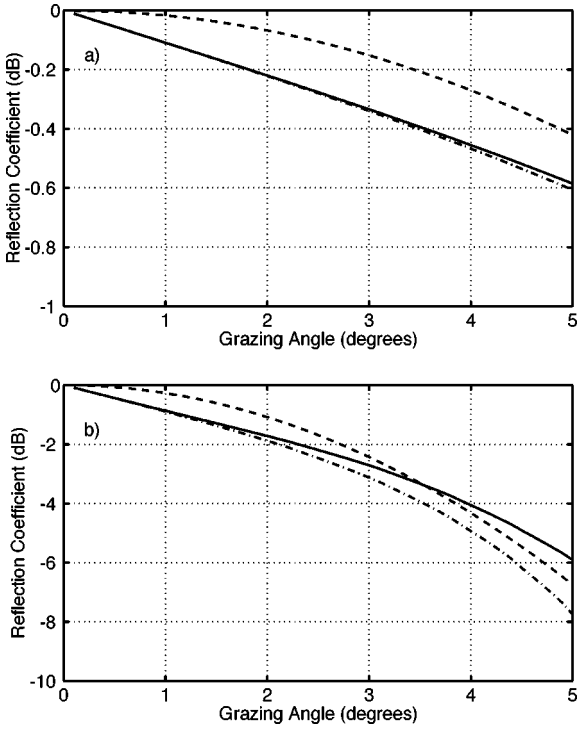


FIG. 6. CSRCs for a frequency of 3.2 kHz and wind speeds 5 m/s (a) and 10 m/s (b) as a function of grazing angle for the small slope (solid lines), perturbation (dash-dot lines), and Kirchhoff (dashed lines) approximations. Only the grazing angle range from 0° to 5° is shown since this angular range is most important for long range propagation.

in the next section use that term in determining a reflection coefficient. Note that the Kirchhoff result requires only the rms wave height h of the surface. For the spectrum used, the rms waveheight [Eq. (6)] is the same as obtained using a 1-D or 2-D (whether isotropic or not) Pierson–Moskowitz spectrum. Thus the Kirchhoff result is that appropriate for a 2-D fully developed sea as modeled by the Pierson–Moskowitz spectrum. The perturbation result

$$R_A(k_{ix}) \approx -1 + 2\kappa_{iz}\int dK_1\, W(k_{ix} - K_1)\sqrt{k^2 - K_1^2} \tag{16}$$

can be obtained by expanding the exponential in Eq. (14) for small arguments, i.e., assuming that $2kh\sin\theta_i \ll 1$. This perturbation theory result is given in the form above in Eq. (13) of Ref. 5.

In calculations of CSRC the input environmental parameter is wind speed. Figure 6 shows the CSRC in dB for wind speeds 5 m/s [Fig. 6(a)] and 10 m/s [Fig. 6(b)] as a function of angle as calculated with the small slope, perturbation, and Kirchhoff approximations. Only the grazing angle range from 0° to 5° is shown since it is this angular range that is most important for the long range propagation calculations of the next section. It is the multiple interaction of rays with the interface that makes an accurate calculation of the CSRC for a single interaction critical to the long range propagation problem.

We now give for reference the perturbation and small slope approximation CSRC expressions for 2-D rough surfaces, which can be obtained by a straightforward generali-

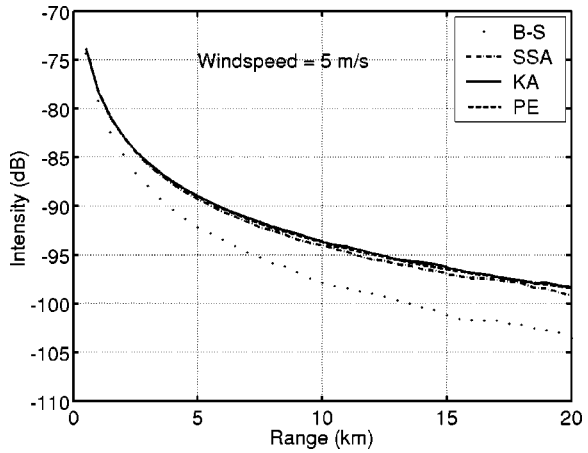


FIG. 7. Comparison at 3.2 kHz of PE depth-averaged intensities as a function of range to GRAB results using three different approximations for the CSRC [Beckmann–Spizzichino (B-S), the small slope approximation (SSA), and the Kirchhoff approximation (KA)] for a 5 m/s wind speed. Considering the differences shown in Fig. 4, the small slope approximation is viewed as being the most accurate model in this case.

zation of Refs. 7 and 8 to 2-D surfaces. The small slope result is given by

$$R_A(\mathbf{K}_i) \approx \exp[-2\kappa_{iz}^2 h^2]\left[-1-2\kappa_{iz}^2 h^2+2\kappa_{iz} \times \int W(\mathbf{K}_i-\mathbf{K}_1)\sqrt{k^2-|\mathbf{K}_1|^2}\,d\mathbf{K}_1\right]. \quad (17)$$

In Eq. (17) we use the wave vector notation $\mathbf{k}_i=\mathbf{K}_i+k_{iz}\mathbf{z}$, where a capital bold symbol denotes the 2-D horizontal component of the corresponding three-vector denoted by a lower case bold symbol. Also, $\kappa_{iz}=|k_{iz}|$, z is a unit vector in the direction of the positive z axis, and $W(\mathbf{K})$ is the 2-D roughness spectrum, which need not be isotropic.

As for the 1-D surface case, the perturbation CSRC can be obtained from the small slope result by a small argument expansion of the exponential factor, keeping terms to second order in kh. The perturbation result is given by

$$R_A(\mathbf{K}_i) \approx -1+2\kappa_{iz}\int W(\mathbf{K}_i-\mathbf{K}_1)\sqrt{k^2-|\mathbf{K}_1|^2}\,d\mathbf{K}_1. \quad (18)$$

This perturbation theory result is given in an equivalent but slightly different form in Eq. (62) in Ref. 9.

VI. GRAB/PE COMPARISONS FOR ROUGH SURFACES

The results within this section compare PE to GRAB at 3.2 kHz using CSRCs calculated using the small slope approximation, the Kirchhoff approximation, and the Beckmann–Spizzichino model. The results shown are similar to those presented for the flat interface case of Figs. 3 and 4, except the PE results are coherent averages (averages of complex pressure) over 50 surface realizations and the resulting intensity is then averaged over depth. The case for $U=5$ m/s is examined first, followed by $U=10$ m/s. For the results within this section perturbation theory [Eq. (16)] produced results indistinguishable from those of the small slope approximation. In the next section the small slope and perturbation approximations are compared under conditions in which their results differ significantly.

Figure 7 shows $10\log_{10}$ of the depth-averaged intensity as a function of range for a wind speed of 5 m/s. Using the PE results as ground truth, it is obvious that the Beckmann–Spizzichino model overestimates the coherent reflection loss. In Fig. 7, the Kirchhoff approximation gives results that are closest to those of PE. However, the differences between PE and GRAB in the flat surface case (Fig. 4) are more closely reproduced by the small slope approximation. As such, the small slope approximation is viewed as being a more accurate CSRC model in this case. This is discussed further below.

Figure 8 shows intensity profiles at ranges of 2.5 and 20 km for a 5 m/s wind speed. The Beckmann–Spizzichino model, the small slope approximation, and the Kirchhoff approximations are compared separately with PE. Again, the Beckmann–Spizzichino model shows obvious errors, especially at 20 km. The small slope and Kirchhoff approximations both compare well with PE though there are differences in the interference patterns seen in those approximations and PE at 20 km.

Figure 9 shows $10\log_{10}$ of the depth-averaged intensity as a function of range for a wind speed of 10 m/s. In addition to the PE results obtained with coherent ensemble averaging that have been used to this point, Fig. 9 also shows the result of averaging the field intensity over the ensemble of surface realizations. This incoherent average of the total field (scattered plus reflected) at each range cannot be obtained with a ray code using a CSRC. However, it is important to note that the coherent and incoherent averages of PE propagated fields become very close to each other as range increases, implying that the field is predominately coherent at long ranges. Thus results from ray-based propagation codes may (for the conditions studied here) be used to represent the total field with little error at long ranges (many water depths). If the total field is needed at short ranges, however, a different approach will be needed.

In Fig. 9 the differences between PE and GRAB in the flat surface case (Fig. 4) are, again, more closely reproduced by the small slope approximation. As such, the small slope approximation is viewed as being a more accurate model in this case also. To quantify this statement, Fig. 10 shows the difference between GRAB and PE for the flat interface case [thick solid line—also given in Fig. 4(b)], for the small slope approximation at 5 m/s (medium thickness solid line) and 10 m/s (thin solid line), and for the Kirchhoff approximation at 5 m/s (medium thickness dashed line) and 10 m/s (thin dashed line). The small slope results follow closely the difference in the flat case while the Kirchhoff results diverge from the flat interface difference with the divergence being more severe at 10 m/s than at 5 m/s. Thus the small slope approximation leads to little additional error in GRAB results relative to the flat interface case.

Figure 11 shows intensity profiles at ranges of 2.5 and 20 km for a 10 m/s wind speed. The 20 km range results clearly reiterate the problem with Beckmann–Spizzichino

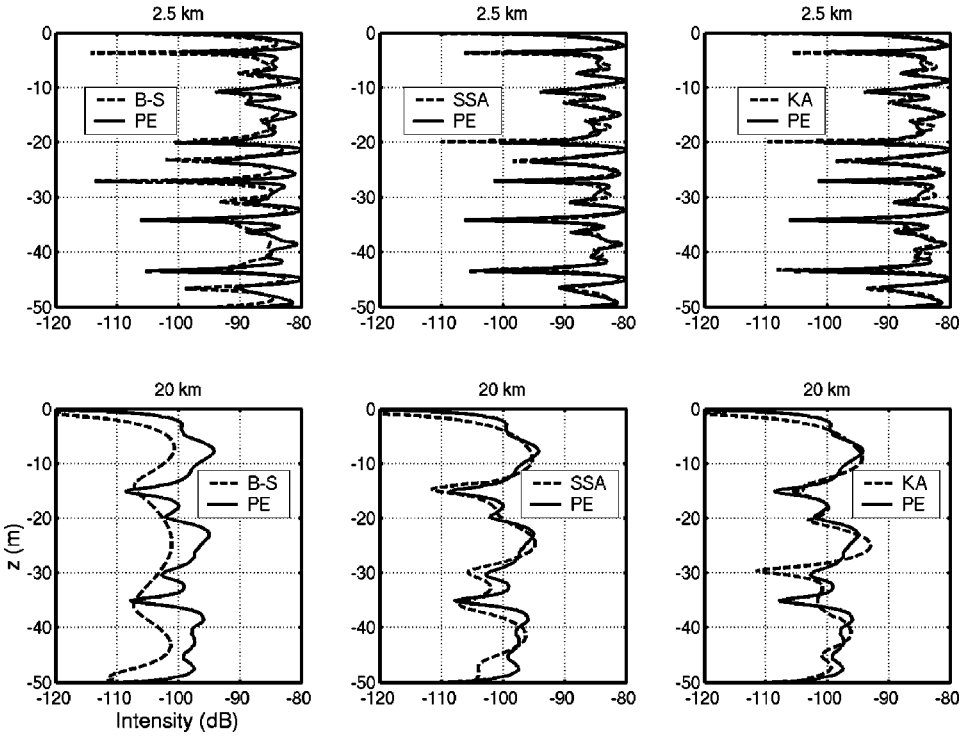


FIG. 8. Intensity profiles at ranges of 2.5 and 20 km for a 5 m/s wind speed. The Beckmann–Spizzichino model, the small slope approximation, and the Kirchhoff approximations are compared separately with PE.

model. The Kirchhoff results at 20 km are consistent with the results in Fig. 10 in that they show intensity levels higher than predicted using PE.

VII. FURTHER EXAMINATION OF SMALL SLOPE AND PERTURBATION CSRC

For the situations examined in the previous section either the small slope or perturbation approximations can be used with equivalent accuracy. At higher frequencies or wind speeds, however, the expansion of the exponential in Eq. (14) leading to Eq. (16) can become inaccurate even for shallow grazing angles. As an example, the CSRC for the small slope, perturbation, and Kirchhoff approximations are shown in Fig. 12 for a source operating at 6 kHz and a wind speed of 10 m/s. Above a grazing angle of 2° the small slope and perturbation results begin to diverge. In the results that follow, the perturbation theory reflection coefficient at grazing angles greater than 2.8° is set to its value at 2.8° since the upturn in reflection coefficient in Fig. 12 above 2.8° is a result of violating the assumptions used in deriving Eq. (16), i.e., higher-order terms in the perturbation expansion cannot be ignored.

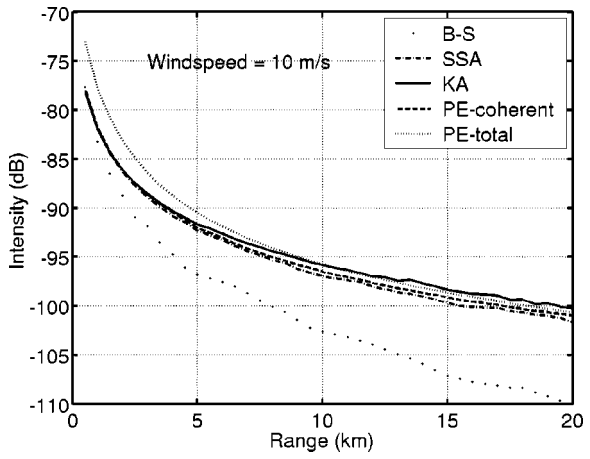


FIG. 9. Comparison at 3.2 kHz of PE depth-averaged intensities as a function of range to GRAB results using three different approximations for the CSRC [Beckmann–Spizzichino (B-S), the small slope approximation (SSA), and the Kirchhoff approximation (KA)] for a 10 m/s wind speed. The PE depth-averaged intensities were calculated from both coherent ensemble averages (PE-coherent) as done in previous figures (and is most appropriate for GRAB comparisons) and by averaging the intensity of the field (PE-total). See text for discussion. As in Fig. 7, the differences between PE (PE-coherent) and GRAB in the flat surface case (Fig. 4) are more closely reproduced by the small slope approximation.

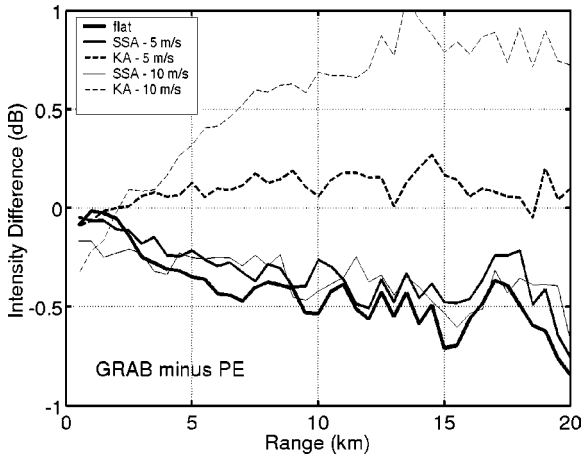


FIG. 10. The difference between GRAB and PE depth-averaged intensities for the flat interface case [thick solid line—also given in Fig. 4(b)], for the small slope approximation at 5 m/s (medium thickness solid line) and 10 m/s (thin solid line), and for the Kirchhoff approximation at 5 m/s (medium thickness dashed line) and 10 m/s (thin dashed line). The small slope results follow closely the difference in the flat case while the Kirchhoff results diverge from the flat interface difference with the divergence being more severe at 10 m/s than 5 m/s.

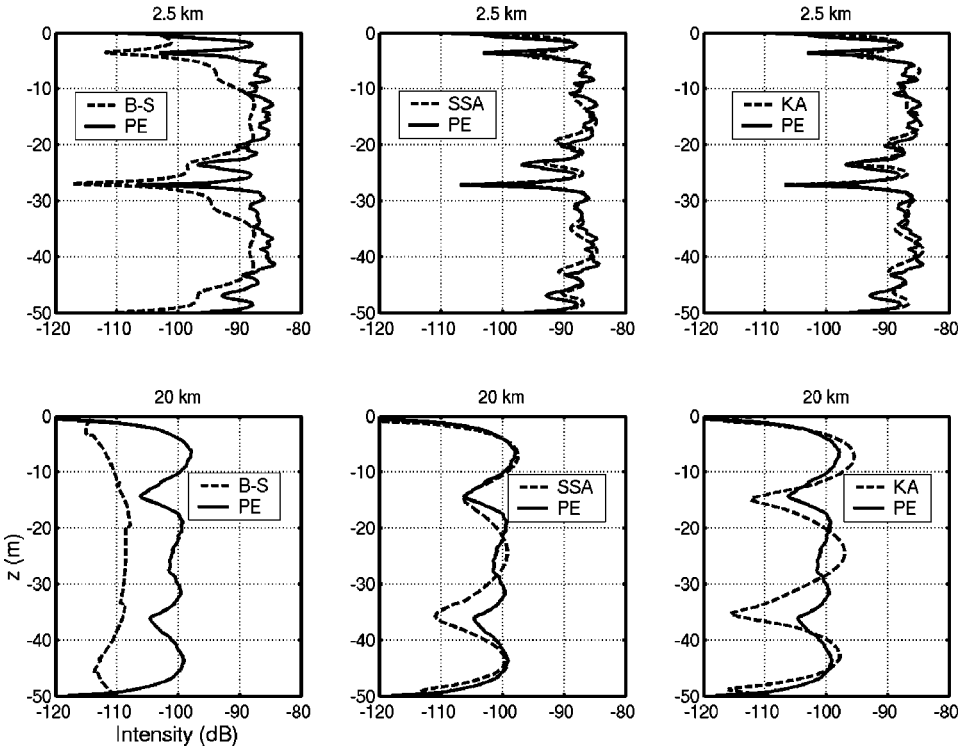


FIG. 11. Intensity profiles at ranges of 2.5 and 20 km for a 10 m/s wind speed. The Beckmann–Spizzichino model, the small slope approximation, and the Kirchhoff approximations are compared separately with PE.

Figure 13 shows the difference between GRAB and PE depth-averaged intensities for this 6 kHz, 10 m/s case. The small slope results follow closely the difference in the flat case (Fig. 4) while the divergence of the Kirchhoff results from the flat interface case is similar to Fig. 10. The perturbation results diverge from the flat interface case for ranges below 7 km. This is a direct result of the inaccuracy of perturbation theory as indicated by the divergence of the small slope and perturbation CSRCs in Fig. 12 as the grazing angle increases. Figure 14 shows intensity profiles at ranges of 1.0 and 20 km for a 10 m/s wind speed. The 1.0 km range results clearly reiterate the problem with perturbation theory at short range. Note that at 20 km none of the GRAB results go to zero at the air/water interface as the boundary condition dictates. This is due to the large (negative) value of the CSRCs at very shallow grazing angles (see Fig. 12—note that even at 1° there are 2 dB losses). Thus the incident and reflected ray amplitudes do not cancel in the GRAB results.

As a further diagnostic of the difference between perturbation and small slope results the rays propagating at different angles were output from GRAB. The ray intensity as a function of range (with spreading removed) is shown in Fig. 15 for rays at six grazing angles (1°–6°). The figure can be understood by examining the two curves labeled 1° (designating the ray traveling at a 1° grazing angle). The thick curve is the small slope result and the thin curve is the perturbation result. The step decreases indicate a boundary interaction (reflection), the small step is a bottom reflection

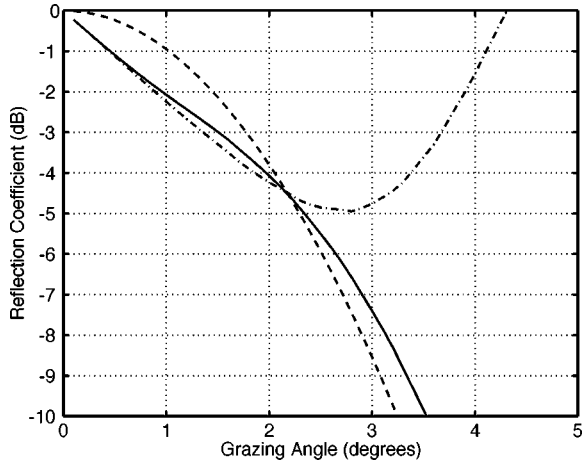


FIG. 12. The coherent surface reflection coefficient for a frequency of 6 kHz and wind speed of 10 m/s as a function of grazing angle for the small slope (solid line), perturbation (dash-dot line), and Kirchhoff (dashed line) approximations.

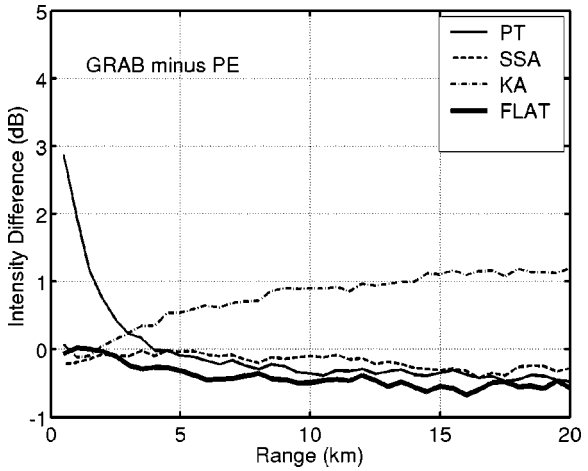


FIG. 13. The difference between GRAB and PE depth-averaged intensities for a source operating at 6 kHz: flat interface (thick solid line); a wind speed of 10 m/s for the small slope approximation (dashed line), the Kirchhoff approximation (dash-dot line), and the perturbation theory approximation (thin solid line).

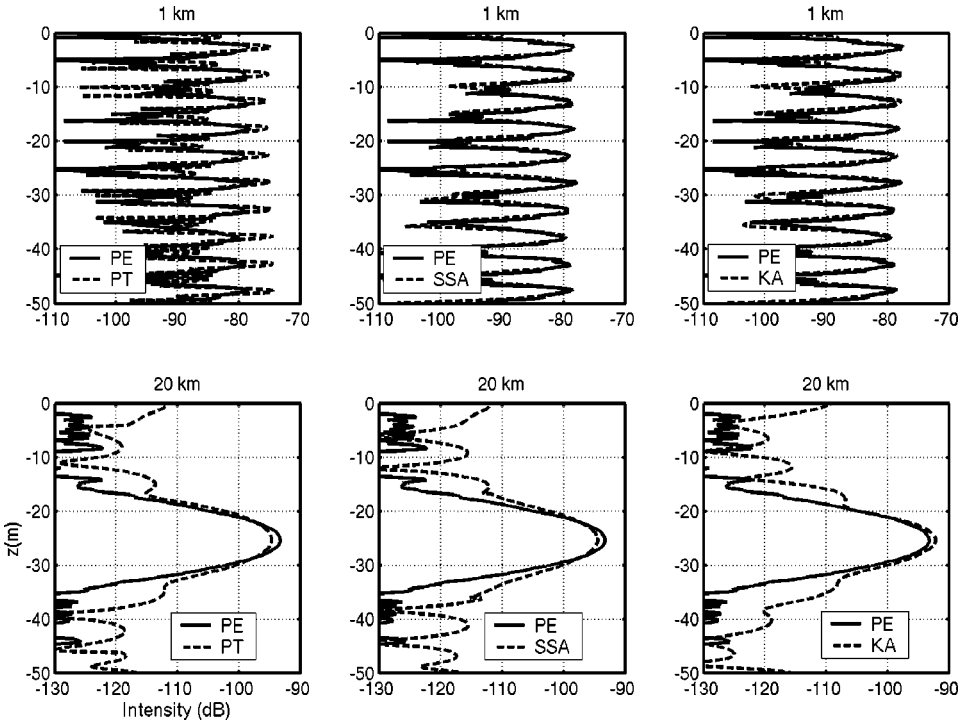


FIG. 14. Intensity profiles at ranges of 1.0 and 20 km for a 6-kHz source and a wind speed of 10 m/s. The perturbation approximation, small slope approximation, and Kirchhoff approximation are compared separately with PE.

and the large one is a surface reflection (note the decrease in level for each surface reflection can be determined by examining Fig. 12). As steeper rays are examined, perturbation and small slope results begin to diverge. In particular, for a 3° grazing angle perturbation theory's underestimate of loss at each surface interaction becomes apparent. The disparity at 4° to 6° has actually been minimized by using the reflection coefficient value at 2.8° (as discussed above). The results in Fig. 15 also lend insight into why the perturbation theory results in Fig. 13 diverge from the small slope results at ranges less than about 7 km.

The small slope approximation is viewed as more reliable than the perturbation approximation since it can accurately determine a CSRC over a larger frequency, wind speed domain.

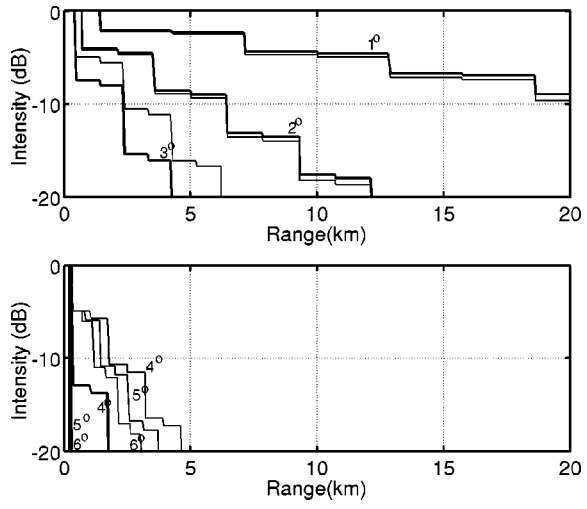


FIG. 15. Ray intensity as a function of range for rays at six grazing angles (1°–6°). The thick curves are the small slope results and the thin curves are the perturbation results. The step decreases indicate a boundary interaction (reflection), the small step is a bottom reflection and the large one a surface reflection. At steeper grazing angles the perturbation and small slope results diverge significantly.

VIII. SUMMARY

A rough-surface PE code[1] was used as ground truth to study acoustic propagation in a waveguide. These PE results allowed a close examination of the ability of a ray code (i.e., GRAB) to accurately estimate coherent field propagation using a CSRC derived from scattering theory. Comparison with PE implies that the Beckmann–Spizzichino model, as given within the GRAB software package, does not give accurate predictions of the coherent field at long ranges. The small slope approximation is the most accurate of the models tested. However, the Kirchhoff approximation is perhaps accurate enough (cf. Figs. 9 and 10) for some purposes and would be much simpler to implement as a module within GRAB. The perturbation approximation can also be a viable option but requires examination of whether the assumption of small roughness is accurate for the particular propagation geometry and frequency being studied. Since there is essentially no computational advantage of using the perturbation approximation over the small slope approximation, use of the small slope approximation would seem to be preferable.

Since the goal here was to test model accuracy, a simplified surface roughness spectrum was used. The acoustic propagation effects of swell or limited fetch could be easily included in the same manner. The focus was primarily on the prediction of the coherent field propagated in a waveguide having a rough air/water interface. However, incoherent averaging of PE results for a source operating at 3.2 kHz and a 10 m/s wind speed implies that, at least for this case, most of the energy is coherent at long ranges. A more general study of field coherence in long range propagation is currently under way. The possibility and utility of an effective energy

reflection coefficient (that accounts for both the scattered and reflected field) for use in ray propagation codes remains to be addressed.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research.

[1] A. P. Rosenberg, ''A new rough surface parabolic equation program for computing low-frequency acoustic forward scattering from the ocean surface,'' J. Acoust. Soc. Am. **105**, 144–153 (1999).
[2] H. Weinberg and R. E. Keenan, ''Gaussian ray bundles for modeling high-frequency propagation loss under shallow-water conditions,'' J. Acoust. Soc. Am. **100**, 1421–1431 (1996).
[3] M. D. Collins, ''A split-step Padé solution for the parabolic equation method,'' J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[4] F. S. Henyey, D. Tang, S. A. Reynolds, and E. I. Thorsos, ''Vertical modes with sediment absorption in general shallow water waveguides,'' submitted to J. Acoust. Soc. Am.
[5] E. I. Thorsos, ''Acoustic scattering from a 'Pierson-Moskowitz' sea surface,'' J. Acoust. Soc. Am. **88**, 335–349 (1990).
[6] E. I. Thorsos, ''The validity of the Kirchhoff approximation for rough surface scattering using a Gaussian roughness spectrum,'' J. Acoust. Soc. Am. **83**, 78–92 (1988).
[7] E. I. Thorsos and S. L. Broschat, ''An investigation of the small slope approximation for scattering from rough surfaces. Part I. Theory,'' J. Acoust. Soc. Am. **97**, 2082–2093 (1995).
[8] S. L. Broschat and E. I. Thorsos, ''An investigation of the small slope approximation for scattering from rough surfaces. Part II. Numerical studies,'' J. Acoust. Soc. Am. **101**, 2615–2625 (1997).
[9] W. A. Kuperman, ''Coherent component of specular reflection and transmission at a randomly rough two-fluid interface,'' J. Acoust. Soc. Am. **58**, 365–370 (1975).

Coherent bottom reverberation: Modeling and comparisons with at-sea measurements

Bernard Cole,[a] James Davis, William Leen, William Powers, and John Hanrahan
Planning Systems Incorporated, Two Shaw's Cove, Suite 200, New London, Connecticut 06320

(Received 10 December 2002; revised 30 May 2004; accepted 30 May 2004)

Shallow water bottom reverberation results collected with a hull-mounted sonar exhibit coherence effects that manifest themselves as reverberation patterns in time—analogous to the well-known Lloyd mirror effect that appears in transmission loss as a function of range for shallow sources or receivers. Moreover, the reverberation peaks and nulls arising from the coherent phasing of the propagation paths are evident both in the reverberation envelopes obtained with short pulses and in the matched filter responses for longer pulse lengths. These ''coherent'' properties of the reverberation deviate from the usual assumptions governing reverberation. Modified assumptions produce coherent reverberation intensity modeled results that match the coherent effects observed in measurements. Finally, the coherent peaks and nulls arising from the phenomenon are shown to produce non-Rayleigh reverberation amplitude distributions in both the measured reverberation and modeled results (increasing the probability of false alarms in this type of interference) unless the reverberation is detrended by the Lloyd mirror pattern.
[DOI: 10.1121/1.1776191]

PACS numbers: 43.30.Gv [WMC] Pages: 1985–1994

I. INTRODUCTION

Traditionally, reverberation (surface, volume, and bottom) has been assumed to be due to a completely random process and, as a consequence, has been modeled as incoherent. However, our previous work in tactical sea trials over 15 years ago in the Persian Gulf suggested that bottom reverberation could exhibit coherence effects. Those particular coherent effects manifested themselves as a reverberation pattern in time analogous to the well known ''surface image interference'' or Lloyd mirror effect that appears in transmission loss as a function of range. While the literature contains a plethora of papers concerning examples of the Lloyd mirror interference in propagation situations (see papers by Young,[1] Urick,[2] and Carey *et al.*,[3] for example), the appearance of such phase patterns in reverberation seems largely unreported in the acoustic community. Two notable exceptions were found: the presence of Lloyd mirror phasing in reverberation was noted in the National Defense Research Committee studies[4] back in 1943 and was later identified and discussed by Mackenzie[5] in his 1960 paper on deep water bottom reverberation. When we discovered that reverberation results collected during a number of other shallow water trials, including a 1995 U.S.-Australian shallow-water sonar trial in the Timor Sea,[6] all exhibited similar patterns, it was concluded that evidence of coherent phase effects in reverberation was more widespread than previously thought. At that point, investigative efforts were launched to incorporate the characteristics of this coherence in a reverberation model.

A theoretical model to compute coherent bottom reverberation intensity was developed (see Appendix A for the equations pertaining to the formulation). The model was then exercised to explore the dependencies of the Lloyd mirror pattern in the reverberation and investigate the conditions under which such patterns might be clearly exhibited in reverberation measurements. Since we hypothesized that such coherent peak and null patterns could be one source of a Rayleigh-deviant amplitude distribution function sometimes observed in reverberation[7–9] [often shown as a change in the tail of the probability of false alarm (P_{fa}) distribution function], a pressure time series model that incorporated coherent phasing in the reverberation was developed as well. (In order to be clear on the terminology, as used in the present context, coherence in the acoustic sense is the pattern of peaks and valleys in the reverberation and coherence in the signal processing sense refers to the characteristics found in the output of a matched filter.)

In this paper, we present evidence for the coherent phase pattern structure in reverberation and seek to explore its characteristics and dependencies in different manifestations by invoking measurement results, modeling results, and comparisons of the two. The modeling assumptions used to derive a formulation for coherent reverberation intensity are outlined in some detail to facilitate understanding dependencies of the effect as it may be anticipated and/or demonstrated in measured results. The organization of this paper is as follows: some features of the peaks and nulls are examined and shown to be present in reverberation measurements from the Persian Gulf (Sec. II), the basis for modeling a limited form of the coherent reverberation intensity is described (Sec. III and Appendix A), and model predictions are compared with measured reverberation collected at two sites off the U.S. east coast (Sec. IV). Next (Sec. V), the modeling of the pressure time series corresponding to both incoherent and coherent reverberation formulations is discussed and illustrations are given of the resulting amplitude distribution functions based on ten modeled realizations—including examples of the effect of different detrending techniques on

[a]Electronic mail: bfcole@plansys.com

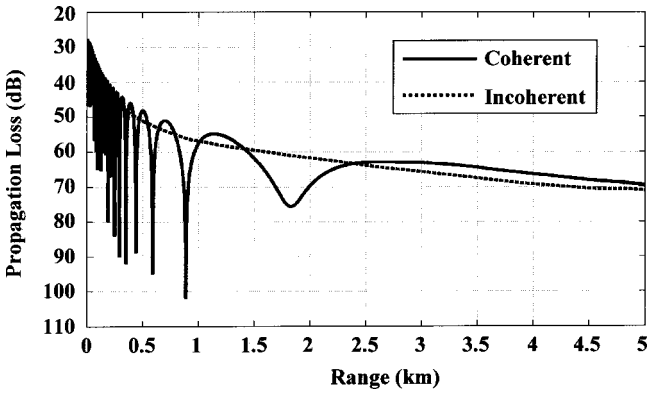


FIG. 1. Typical Lloyd mirror propagation interference pattern to an ocean bottom receiver.

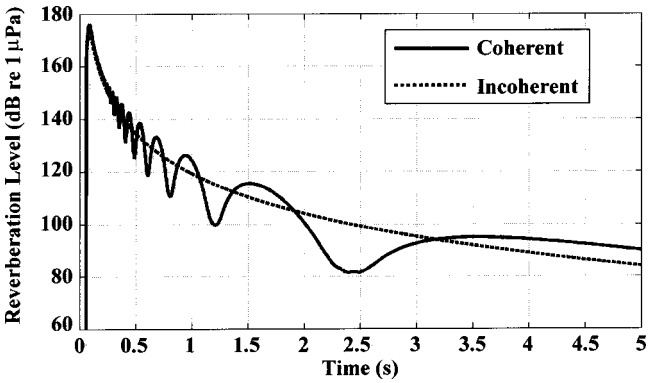


FIG. 3. Representative comparison of coherent and incoherent model results for isovelocity conditions.

these distribution functions. Comparisons of the modeled distributions from this time series simulation with reverberation distributions derived from the one of the measurements off the U.S. east coast are then presented (Sec. VI). Finally, some concluding remarks are given (Sec. VII), followed by appendices containing the mathematical details of the various model formulations.

II. LLOYD MIRROR CHARACTERISTICS AND PERSIAN GULF EXPERIMENT

The classical Lloyd mirror interference effect (formed by the phased interference of the direct and surface reflected paths from a near-surface sonar to a receiver on the ocean bottom) produces a propagation loss dependence in range that has (1) a finite number of constructive peaks; (2) a proportional increase in the spacing of the peaks and nulls with range; and (3) a broadening of the individual peaks and nulls with increasing range, as depicted in Fig. 1. To illustrate how these characteristic features would translate directly to reverberation, were the backscattering coming from a single point on the bottom, an analogous plot of ''source level minus two-way propagation loss'' is shown as a function of time in Fig. 2. As anticipated from Fig. 1, the effective doubling of the pattern amplifies the difference between peaks and nulls such that the levels can differ by almost 100 dB. In the reverberation formulation we have adopted, the scattering is modeled as occurring from a very small but finite area, which maintains the peak structure of the pattern shown in Fig. 2 but results in the nulls of the pattern being partially destroyed, as is shown in Fig. 3. While this may seem a subtle distinction, in our view it is an important one when distinguishing between ''coherent propagation effects'' manifested in reverberation (as illustrated in Fig. 2) and ''coherent reverberation'' where the phase structure is smeared but still persists in scattered returns. This notwithstanding, the recognition of such general interference characteristics (i.e., the proportional spacing and peak spreading) prompted the Lloyd mirror interpretation for the reverberation envelopes measured in the Persian Gulf in 1988.

Examples of different phase patterns exhibited by the reverberation envelopes collected with a hull-mounted sonar progressing along a track in the Persian Gulf are presented in Fig. 4. Each of the curves shown represents a reverberation time series that was collected with short, 10 ms pulses, was time-averaged with a 6 ms time window and then ping-averaged over 18 consecutive pings. The individual 18-ping reverberation collections along the track were segmented by geographical location and were assigned the run number designators. Since some of the collections were performed using different types of wave forms, those envelopes have not been included here and only the short pulselength reverberation envelopes corresponding to Run Nos. 1, 2, 7, 8, 10, 11, and 16 are presented in Fig. 4. Taken collectively, these curves serve to illustrate both the spatial/temporal variability of the reverberation envelopes and the overall persistence of some semblance of phase interference structure in all the results. It

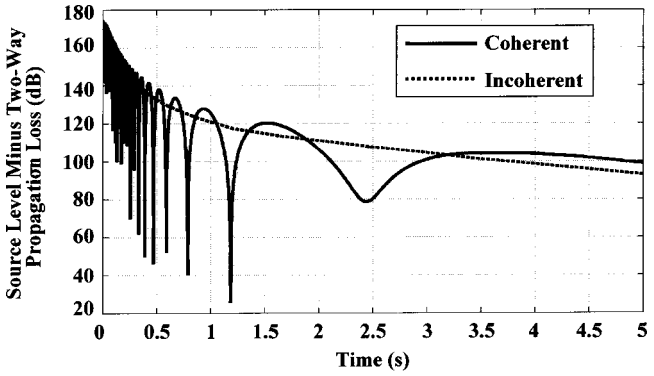


FIG. 2. Bottom reverberation from a single point is analogous to source level minus two-way propagation loss.

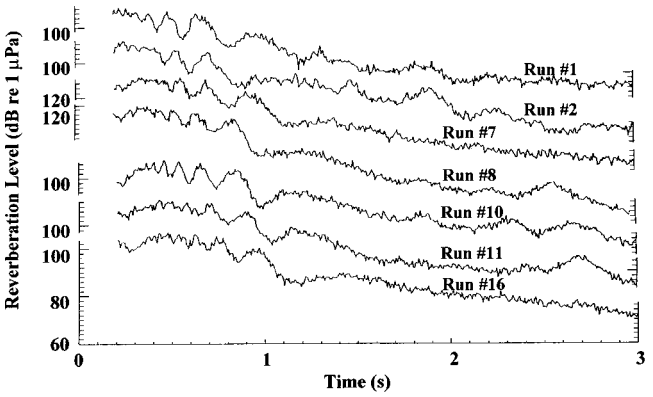


FIG. 4. Reverberation envelopes measured in the Persian Gulf depict persistence and run-to-run variability of Lloyd mirror interference.

is immediately obvious that several of the curves have major peaks and valleys at later times. In this Persian Gulf environment, the peaks and valleys shown arriving in the first second after transmission are backscattered directly from the intersection of the sonar beam and pulselength with the bottom and contain no influences from bottom interacting/forward propagating paths. Notice that some of the measurements contain peaks and valleys at reception times of 1.25 s and beyond while others do not. The reverberation structure at these longer reception times contains bottom interacting forward paths and is sensitive to the magnitude of the bottom reflection loss values as well as the bottom backscattering values of the specific environment. In areas possessing various combinations of higher bottom loss, lower backscattering strengths and significant biological or other noncoherent forms of scattering, the persistence of the pattern will be altered. Significant to the discussion here, many reverberation models (e.g., Sundvik *et al.*)[10] would calculate a smooth reverberation decay curve for these conditions. Certainly, the reverberation shown contains significant fluctuations and is far from smooth.

The basis for modeling such fluctuations by invoking coherent reverberation intensity in the form of the Lloyd mirror effect is presented in the next section. The characteristic dependencies of this reverberation are first discussed for a simple isovelocity case as a prelude to comparison with and explanation of selected measured reverberation samples in Sec. IV.

III. COHERENT REVERBERATION INTENSITY MODELING

As evidenced earlier, bottom reverberation observations have shown characteristics similar to the pattern shown by transmission loss based on the coherent addition of the appropriate paths. The issue involved is how transmission loss based on the coherent addition of paths plays a role in the reverberation, when it will play a role, and what parameters govern its behavior. A completely coherent formulation of the reverberation poses problems in normalization and creates spurious behaviors in time. Such a completely coherent formulation is available in the ray theory-based generic sonor model (GSM)[11] and the early versions of the comprehensive acoustics sonar simulation (CASS) model (see Weinberg[12] and Keenan),[13] which uses the Gaussian ray bundle propagation model (described by Weinberg and Keenan)[14] as one of its propagation capabilities. (During discussions with Dr. Weinberg, the authors have learned that later versions of the CASS model will be modified to include the capability of computing reverberation using the coherent formulation presented in this paper).[15]

From our studies, a restricted form of coherent reverberation appears to give meaningful results. This restricted form is described here for the simplest possible case: an isovelocity water column and a nonreflecting bottom. Both linear freqeuncy modulation (LFM) and continuous wave (cw) pulses are modeled. For this simple case there are only four paths involved in scattering from an elemental bottom patch: (1) source-bottom-source, (2) source-surface-bottom-source, (3) source-bottom-surface-source, and (4) source-surface-bottom-surface-source. The source is a cw pulse having a time dependence of $\exp(-i\omega t)$. At any time, t, the reverberation consists of all multipath contributions scattered from an area or ''patch'' that satisfies the geometrical and pulse length constraints. The usual assumption is that these all add together incoherently. For an unconditional coherence formulation these would all add together coherently. The basis for the present formulation is the following assumption:

''Only contributions from those paths whose pulse duration permits arrivals from a small elemental area of the classical pulse length projected area add together coherently. These contributions then add incoherently with all the other contributions, including other adjacent elemental areas that comprise the classic pulse length projection on the boundary in a given sonar beam.'' (Similarly, adjacent elemental areas in other beam directions also add incoherently.)

This assumption is equivalent, in some sense, to an extremely small spatial coherence length. However, at the same time, it should be noted it also implicitly implies a finite angular correlation length. The resulting reverberation displays a coherent pattern over time of reception provided that:

(1) the pulse length is long enough to include the interfering paths, and
(2) the pulse length is not so long that several cycles of the pattern can interfere.

When the earlier conditions are not satisfied, the reverberation pattern is essentially the same as found in ''incoherent reverberation.'' As a consequence of these conditions, the reverberation exhibits no coherent effects as the pulse becomes infinitely long.

A mathematical formulation of this limited form of coherence is given for a simple case in Appendix A. To facilitate reverberation calculations for various sound speed profile and bottom loss conditions, modified research versions of both the GSM and CASS models were generated by the authors specifically to compute coherent reverberation based on these assumptions. An example of the coherent reverberation result produced by the GSM model for isovelocity conditions was presented in Fig. 3. Among the specific assumptions used for the computation shown in Fig. 3 were: a 20 ms cw pulse at 3500 Hz, a source and receiver depth of 7.62 m, a water depth of 50 m, and a bottom backscattering coefficient of -27 dB. Before considering the impact of more complicated sound speed profile conditions, it is beneficial to first review some of the characteristics evidenced in Fig. 3 and examine some of the behaviors and dependencies of coherent reverberation for some specific variations of this base line case.

Returning to Fig. 3, note the interference pattern, with well-defined peaks and nulls, in the coherent reverberation. This pattern of reverberation as a function of time is the analog of the Lloyd mirror pattern formed by the transmission loss as a function of range. The locations of the nulls and peaks can be approximated by the times, t_n, given by

$$2kWS/ct_n = n\pi/2, \tag{1}$$

where n is even and odd, respectively. In this equation, k is the wave number, W is the water depth, S is the source depth

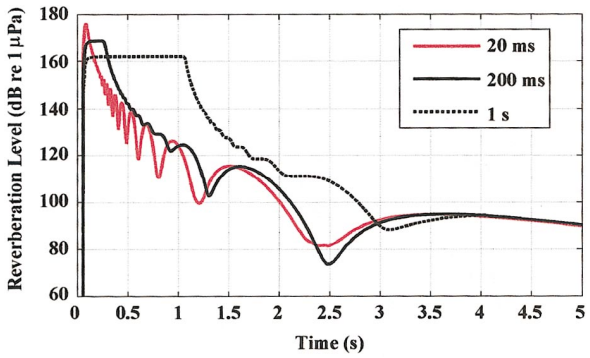


FIG. 5. Dependence of coherent reverberation on pulse length (20, 200 ms, and 1 s) for a cw pulse.

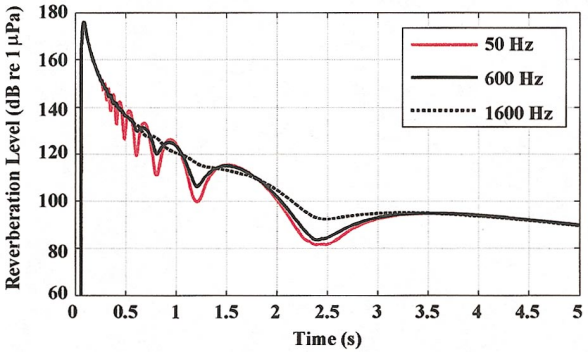


FIG. 6. Dependence of coherent reverberation on bandwidth (50, 600, and 1600 Hz) for a LFM pulse.

(equal to the receiver depth), and c is the sound velocity. For the specific modeling conditions given, the last null in the pattern occurs at approximately 2.4 s and the last peak at approximately 3.7 s. An increase in frequency, water depth, or source and receiver depth will cause the pattern exhibited in Fig. 3 to stretch out in time—analogous to the stretching of the Lloyd mirror propagation pattern in range that would occur with these same changes. Note, however, that the coherent reverberation peak and null patterns have additional dependencies not normally associated with the Lloyd mirror propagation: e.g., pulse length, wave form type (cw, LFM, hyperbolic frequency modulation (HFM)), bandwidth, processing and/or matched filter resolution, and bottom reflection loss. Also notice that the coherent reverberation can differ from the incoherent reverberation by as much as 15–20 dB. The absence of a pattern at the shorter times (less than approximately 0.3 s) is because a 20 ms pulse is too short to include all of the interfering paths. As a consequence, the four paths cannot add coherently.

An explicit example of the dependence on pulse length, computed using the same inputs as Fig. 3 except varying the pulse length, is shown in Fig. 5 at 3500 Hz for cw pulses of 20, 200 ms, and 1 s. For ease of comparison, the 200 ms result has been shifted down by 10 dB and the 1 s result shifted down by 17 dB. Focusing on the time regions beyond the initial reverberation return from the bottom (i.e., after approximately 300 ms for the 200 ms pulse and 1.1 s after the 1 s pulse), notice that the clean pattern for a pulse length of 20 ms becomes weaker and starts to disappear as the pulse length increases. A coherent pattern is also present for a LFM pulse and depends on bandwidth as well as pulse length, as shown in Fig. 6 for a 20 ms LFM pulse with bandwidths of 50, 600, and 1600 Hz. There are several points to be made from this figure. First, the 50 Hz bandwidth result is identical to that shown for the cw pulse in Fig. 3. Second, since the location of peaks and nulls are frequency dependent, smearing (or an averaging) of the pattern might be expected as the bandwidth increases. This clearly is seen to be the case in Fig. 6 where the peak and null amplitudes are dampened for the 600 Hz bandwidth and totally smoothed out for the 1600 Hz bandwidth. (It should be noted, however, that the coherent pattern re-emerges if these wider bandwidth pulses are subsequently subjected to pulse-compression processing using a matched filter with replica correlation.)

The pulse length and bandwidth sensitivities exhibited in Figs. 5 and 6 are for an idealized isovelocity water column case, however, the results suggest a possible reason why the patterns have not been noted more widely in shallow water sonar trials: the longer cw and frequency modulation (FM) pulse lengths (and FM bandwidths) typically used for sonar detection would smear the pattern and make it unrecognizable. In such circumstances, the smeared pattern may not be evident but presumably the underlying physics of the interference pattern should still be present and could reappear at the output of the matched filter with an associated impact on detection problems. Specifically, the target signal may have one range pattern of peaks and nulls and the reverberation another pattern. The range registration (i.e., the range alignment of the peaks associated with the target with respect to the nulls/peaks in the reverberation) of these patterns can produce zonal detection regions or, at a minimum, a more rapidly varying signal excess with range than anticipated.

Some sample comparisons of measurement observations with model computations are provided and discussed in the next section.

IV. REVERBERATION MEASUREMENT AND MODEL COMPARISONS

As part of the study of Lloyd mirror reverberation patterns, sea tests were designed and performed at several locations off the east coast of the U.S. in conjunction with some sonar detection trials specifically designed to demonstrate environmental influences on sonar performance. In this paper, we only present reverberation times series measurement and modeled results for two of those locations, designated as site A (32.58°N, 78.17°W) and site B (33.03°N, 77.53°W). Nevertheless, these measurements are representative since the observations from the other locations produced comparable reverberation peak and null patterns. The sound speed profiles applicable to site A and site B measurements are presented in Fig. 7. As shown, the profiles at both sites were predominantly downward refracting, as is typical of many shallow water areas during the summer and fall seasons.

A. Site A results

The water depth at site A is 253 m. As depicted in Fig. 7, the sound speed for this measurement was largely downward refracting but has a surface duct down to 12 m. The sonar

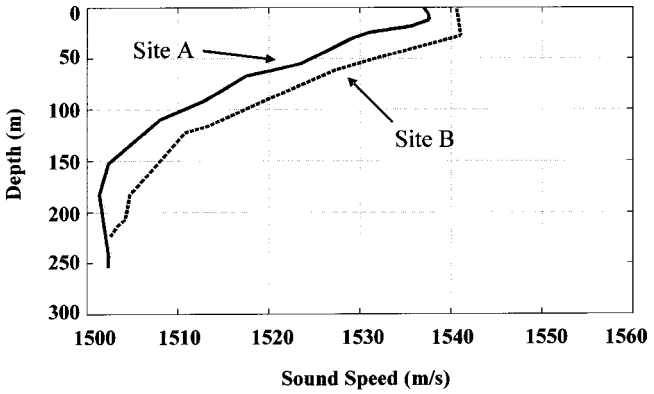


FIG. 7. Sound speed profiles for test sites A and B.

transmitting-receiving arrays are steered down to 3° and both are in the duct at 7.62 m (25 ft). The measured reverberation due to a 10 ms cw pulse at 2730 Hz was time averaged over 12.5 ms and power averaged over ten pings. The reverberation result shown in Fig. 8 clearly contains the amplitude fluctuation pattern characteristic of the Lloyd mirror interference up to approximately 2.5 s. The modeling results shown in Fig. 8 agree quite well with the measurements in this region of time.

Some comments are in order regarding both the comparison and the modeling result. Since the sound speed is not isovelocity, the model computation was made using the specially modified version of GSM (i.e., containing our coherent reverberation formulation). The model computation could have also been made using our similarly modified version of CASS. Because the sound speed profile is downward refracting there is a shadow region bounded by a limiting ray and this is evident in the reverberation at approximately 2.5 s. The downward refracting profile compresses the pattern that is seen under isovelocity conditions (as illustrated by comparing Fig. 8 with Fig. 3). Additionally, because the source and receiver are both in the surface duct the last peak in the isovelocity case is weakened to the extent that it is not discernable in Fig. 8.

For convenience of reference, the following convention will be adopted. The first order reverberation return consists of all paths from the source to the scattering patch that do not contain a forward reflecting interaction with the bottom and, similarly, of all paths back from the scattering patch to the receiver that do not contain a forward reflecting bottom interaction. The second order return consists of all paths that contain one bottom forward reflecting interaction going from the source to the scattering path and of all paths that consist of one forward reflecting bottom interaction going back to the receiver. Similarly for higher orders, hybrid returns consist of combinations such as, for example, no bottom reflection interaction path going out and one bottom reflection interaction on the path returning and also one bottom reflection interaction path going out and no bottom reflection interaction path returning.

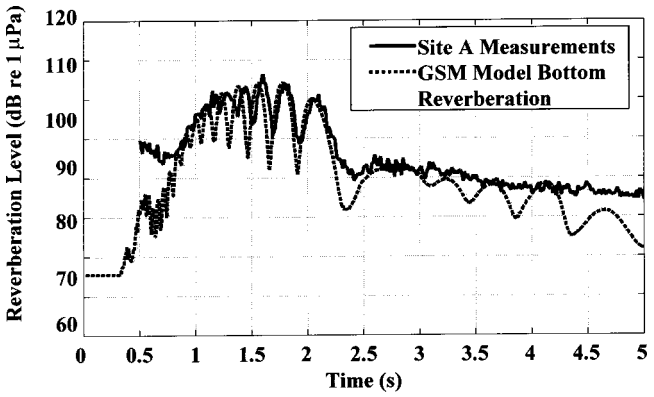


FIG. 8. Comparison of measured reverberation at site A with coherent modeling result (based on modification to GSM).

The reverberation shown in Fig. 8 is dominated by the first order return up to 2.5 s and is therefore independent of the bottom loss in the area. The reverberation from 2.5 to 5 s is dominated by the second order and higher order returns (as well as hybrids) and is therefore strongly dependent on bottom loss. The bottom loss used in the model computation was actually extracted from broadband reverberation time series data collected at a location about 20 km west of the site A area but nevertheless its use in the reverberation prediction seems to provide a reasonable (but slightly underestimated) representation of the level of the measured reverberation. One final point to be made from the modeled result in Fig. 8 is that there is structure in the reverberation beyond 2.5 s that is not evident in the measurements. This structure is due to the pattern formed by coherence in the higher order and hybrid returns. Notice that the measured results from 0.5 s to approximately 0.75 s are higher than the modeled results. It is possible, but has not been proven, that this level is due to biological scattering known to be present in the area. Such biological reverberation, if it persisted out to 5 s, would account for the lack of structure in the measured observation in contrast to that exhibited by the model computation assuming only bottom reverberation.

B. Site B results

The water depth at site B is 223 m and the appropriate sound speed profile is also presented in Fig. 7. This sound speed profile is also downward refracting but now has a surface duct down to 24 m, deeper and stronger than at site A. Again, the sonar transmitting-receiving arrays are steered down to 3° and both are in the duct at 7.62 m (25 ft). The measured reverberation collected with a 10 ms cw pulse at 2730 Hz was time averaged over 12.5 ms and power averaged over ten pings. These reverberation results from site B are compared with the GSM modeled reverberation envelope in Fig. 9. While some of the peaks and nulls contained in the first order pattern are nearly in register, the model and the measurements show an offset for the broad peak in the vicinity of 2 s. In addition, the model depicts some coherent structure between 3 and 4 s, which is not evident in the measurements. Conversely, the measured results contain coherent structure in the 5–6 s interval that is not well replicated by the model. As alluded to earlier, the exact structure for the time intervals beyond 2 s (depending on water depth, etc.) can depend significantly on the bottom reflection loss properties of the area as well as on the bottom backscattering properties. Accordingly, the structure observed in the at-sea

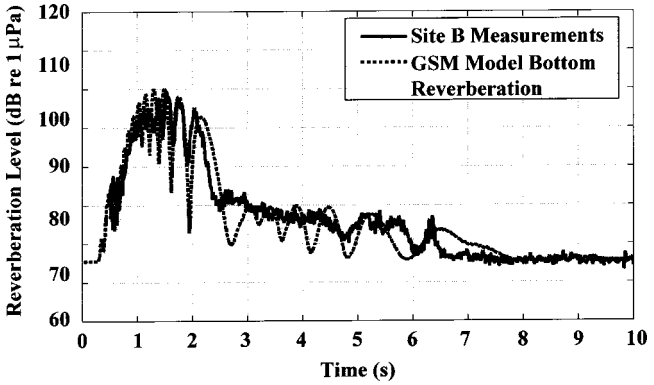


FIG. 9. Comparison of measured reverberation at test site B with coherent modeling result (based on modification to GSM).

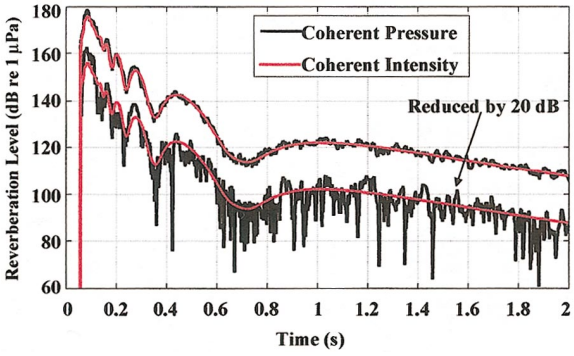


FIG. 10. Comparisons of the reverberation based on one realization and ten realizations of the coherent pressure formulation with reverberation based on the coherent intensity formulation.

measurements provides encouragement for the continued research aimed at developing methods of exploiting such detailed structures to determine the magnitude of the bottom reflection loss.

V. COHERENT REVERBERATION PRESSURE MODELING

The received level at a hydrophone (or an array of phones) consists of the reverberation pressure time series (and the other components of the masking interference, as well as any signal that might be present) but model predictions, such as in the preceding sections, can generally only be made for the reverberation intensity. Simulation of the time series information is important for filtering (e.g., matched filtering) and determining reverberation fluctuations and statistics for purposes of, for example, P_{fa} predictions. Accordingly, studies were deliberately undertaken to characterize the coherence of the reverberation pressure time series. It was found possible to develop a formulation that incorporates incoherent and coherent effects into the reverberation pressure time series and then modify the simple model described in Sec. III (as well as in the GSM and CASS models) to generate realizations of the reverberation pressure time series. For present purposes, only the coherent formulation will be discussed.

Simple illustrative examples of the consequences will again be based on the modeling parameters used earlier, except that now the frequency will be 1000 Hz. The reverberation based on a single realization of the coherent pressure formulation at this frequency, along with the reverberation based on the coherent intensity formulation, are presented as the lower set of curves in Fig. 10. Both of these curves have been shifted down in level by 20 dB to also facilitate inclusion of a similar comparison for the ten-ping pressure realization, shown by the upper two curves. Notice that the single realization of the coherent pressure reverberation is varying in a random fashion about the coherent intensity reverberation and therefore has random and deterministic features. This result for a single realization can be thought of as the reverberation arising from a single ping. Computations for other realizations (not shown) have different random behaviors but all have in common that the variations are about the coherent intensity reverberation. It is to be expected, therefore, that the power average of the intensity over a large number of realizations or pings would approach the coherent intensity reverberation. As demonstrated by the upper two curves in Fig. 10, which compare the expected value based on ten realizations to the coherent intensity reverberation, this is, in fact, the case.

Consider now the fluctuation statistics of the reverberation for this simple baseline case. For the statistics to be meaningful the reverberation time series should be detrended to remove any deterministic bias (e.g., cylindrical spreading). There do not appear to be well-defined guidelines for detrending. It would appear that if all deterministic patterns are removed, then the distribution of amplitude fluctuations should be Rayleigh, unless for example, there are bimodal distributions. However, if some of these patterns are not identified and removed, then the statistics should be non-Rayleigh or Rayleigh deviant. As will be shown, the features of the Lloyd mirror pattern that are present in the coherent pressure reverberation formulation are not completely random and if they are not removed will lead to a Rayleigh deviant P_{fa}.

Although they are not shown, the distributions for each of the realizations comprising the ten-realization average shown in Fig. 10 are Rayleigh deviant. In the present case there are two ways that the reverberation can be detrended. The first method recognizes that realizations of the coherent pressure reverberation have a systematic cylindrical falloff with time much like the incoherent intensity reverberation shown in Fig. 3. This persistent falloff with time of a single realization of the coherent pressure reverberation can be removed by using the incoherent intensity reverberation to detrend (i.e., normalize the one curve by the other). The log of the P_{fa} for each realization based on this detrending is shown as a red curve in Fig. 11. The Rayleigh distribution is shown for comparison. Notice that while no two are identical they are all Rayleigh deviant.

It is clear that the coherent pressure reverberation depicted in Fig. 10 has more than just a persistent falloff with time. As shown in that figure, it varies randomly about the coherent intensity reverberation. Accordingly, the next method is to detrend the coherent pressure reverberation by the coherent intensity reverberation. The results obtained by detrending each of the ten realizations by this method are

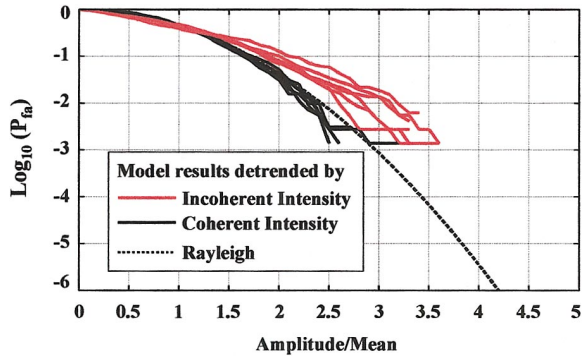


FIG. 11. Comparison of $\log_{10}(P_{fa})$ distributions for ten coherent pressure model realizations, detrended by the coherent and incoherent intensity reverberation, with the Rayleigh distribution.

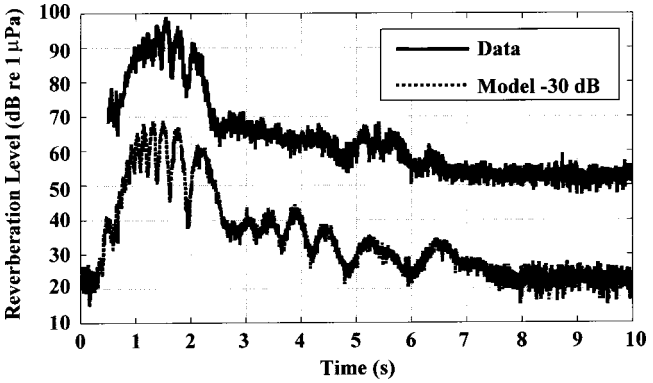


FIG. 12. Comparison of measured reverberation at site B (identical to that in Fig. 9) and average of ten ping realizations of coherent pressure modeling computations (based on modification to GSM).

included in Fig. 11 as well, as depicted by the family of black curves. Notice that while no two are identical they are now all close to the Rayleigh distribution.

These are very interesting model results. They suggest the following generalizations:

* If the Lloyd mirror interference effect is present in reverberation time series data then the P_{fa} will be Rayleigh deviant,

and

* If this pattern can be identified and removed and if there are no other patterns or correlations the P_{fa} will be Rayleigh.

An important point to recognize is that while these results show that Lloyd mirror effects yield a Rayleigh deviant P_{fa}, the converse is not proven. That is, these findings do not suggest that a Rayleigh deviant P_{fa} necessarily implies that a Lloyd mirror effect must be present. With this insight (i.e., from the simple isovelocity case for 1000 Hz) as background, the next section will now examine the extensibility of these observations to some measured and modeled reverberation fluctuation statistics for the shallow water conditions at site B, where the effect is already known to be present.

VI. COMPARISON OF MEASURED AND MODELED REVERBERATION FLUCTUATIONS

The site B measured reverberation result at 2730 Hz presented earlier in Fig. 9 is reproduced again in Fig. 12. However, the modeled result differs from the one presented in Fig. 9 and was generated by averaging ten individual ping realizations produced by the coherent pressure times series model (which again used our modified version of GSM). To facilitate a visual comparison of the measured results with the model results, the pressure time series modeled result has been shifted down in level by 30 dB. Since the modeled averaged time series for the ten pings closely resembles the intensity modeled reverberation output, many of the same discrepancies between the measured and modeled results noted in conjunction with Fig. 9 are still evident here. The advantage of using the pressure time series model is that it provides a mechanism of generating ping-to-ping modeling estimates of the reverberation distribution functions for comparison with those observed in reverberation measurements.

In Fig. 13, reverberation distributions corresponding to two different detrending methods are shown and compared to the Rayleigh distribution. In one instance, the individual time series for ten consecutive pings of 10 ms cw pulse reverberation collected at site B were each detrended by normalizing by a running averaging window of 600 ms, then the resulting mean of each was used to form one of the reverberation distributions shown by the family of red curves in Fig. 13. The family of black curves in the illustration were processed identically, but used a running averaging window of 100 ms. Notice that the measurement results processed with the 600 ms averaging deviate significantly from the Rayleigh distribution, while the distributions obtained with the 100 ms averaging essentially follow the Rayleigh curve. This can be attributed to the longer averaging time window producing (and detrending by) a mean that is similar to the incoherent reverberation curve. In contrast, the shorter integration time window produces a mean that tracks the Lloyd mirror fluctuations and, as such, leads to the reverberation now being detrended by the peaks and nulls of the interference pattern. In short, the distributions for the site B reverberation measurements exhibit behavior similar to that noted in the isovelocity modeling result. This having been shown, we now examine the site B modeling results to determine if the distributions derived from the coherent pressure reverberation calculations also exhibit the same behavior.

The ten realizations produced by the reverberation pres-

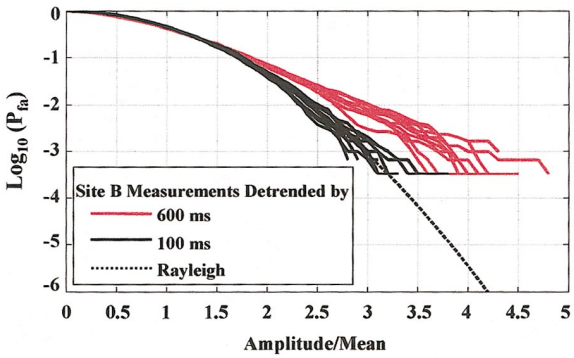


FIG. 13. Comparison of $\log_{10}(P_{fa})$ distributions of ten site B measurement pings, detrended by 100 and 600 ms averaging windows, with Rayleigh distribution.

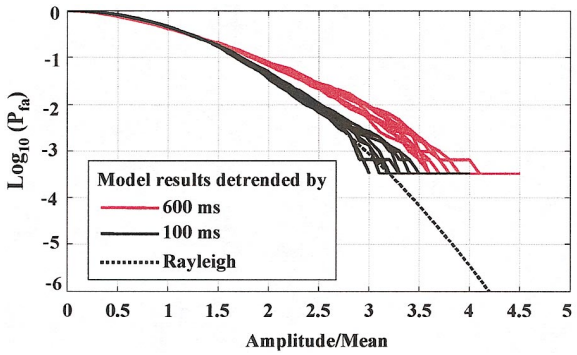


FIG. 14. Comparison of $\log_{10}(P_{fa})$ distributions of ten model generated pings, detrended by 100 and 600 ms averaging windows, with Rayleigh distribution.

sure time series model for the environmental conditions at site B were subjected to the same 600 and 100 ms averaging and detrending processing that was applied to the site B measured time series. The resultant model-generated distributions are shown in Fig. 14. As before, the results based on the 600 ms averaging are shown by the red curves and the results for the 100 ms averaging depicted by the black curves. Again, the Rayleigh distribution is included for comparison purposes. As with the site B measurements, the modeled results processed using the 600 ms averaging deviate significantly from the Rayleigh, whereas those for the 100 ms averaging follow the Rayleigh distribution. For the case of the 600 ms averaging, also note that the envelope formed by the collection of the individual pings for the modeled results tends to be more tightly grouped than the corresponding envelope formed by the collection of curves derived from the measurements and shown in Fig. 13. Moreover, the modeled results (taken as a whole) are, on the average, slightly less deviant from the Rayleigh curve than the measurements. As yet, it has not been determined if subtle differences such as these are significant. A different (second) set of ten-ping realizations from the pressure time series model might produce a set of curves that more closely replicates the reverberation distributions observed in the measurements. Alternatively, another ten-ping set of reverberation time series from the experiment could produce a set of curves that more closely replicates the reverberation distributions generated by model for the first ten realizations. (Since no such perturbations were attempted in our studies, this is merely a speculation by the authors.)

Such subtle differences notwithstanding, these analyses of the measured time series and the modeled pressure reverberation suggest that short interference estimation windows are highly desirable when coherent interference effects are present in the reverberation background, unlike the noise interference condition (where longer interference estimation windows are often desirable to obtain a better estimate of the mean). In addition, these measurement and modeling results are very encouraging, demonstrating that while model generated predictions of reverberation containing Lloyd mirror effects might not agree in detail in reproducing the coherent pattern possessed by the measurements, they nevertheless can still provide representative measures of the reverberation fluctuation statistics for realistic shallow water environments.

VII. DISCUSSION AND CONCLUSIONS

We have shown that the Lloyd mirror pattern is indeed present and can represent a characteristic feature of reverberation measurements collected in shallow water with a hull-mounted (or other near-surface) sonar. The comparisons of model predictions (based on our formulation of coherent reverberation) to measured results have shown good-to-excellent agreement. The sensitivities of the pattern to various factors predicted by the modeling suggest that the effect could be widely prevalent in many reverberation data sets but is not recognized due to it being obscured by the long pulse length and/or wide bandwidth pulses generally transmitted.

Modeled reverberation amplitude fluctuations, derived from GSM complex reverberation pressure time series containing the interference effect, compare favorably to the amplitude fluctuations observed in reverberation measurements. It has been demonstrated by model computations and confirmed, in part, by reverberation observations that: (1) reverberation that contains the Lloyd mirror effect will have a Rayleigh deviant P_{fa}, (2) reverberation that has a Rayleigh deviant P_{fa} does not necessarily contain the Lloyd mirror form of coherence, and (3) if there are no other effects that lead to a deviant behavior, detrending by the Lloyd mirror interference pattern, if identified, leads to a Rayleigh P_{fa}.

Based on the limited measurement observations discussed herein, the periodic reverberation peaks and valleys caused by Lloyd mirror effects, aside from their implications on detection performance, would appear to possess characteristics that can be exploited for environmental adaptation. Alternatively, their presence, if unaccounted for, could lead to errors in the values of extracted parameters and to incorrect assumptions relating to the reverberation statistics. Either of these can directly result in inaccurate detection performance characterizations when the sonar is near the ocean surface.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The authors would like to thank Ken Dial for his suggestions and recommendations on the research contained in this paper.

APPENDIX A

A. Intensity formulation

In this appendix, a coherent reverberation intensity formulation (i.e., reverberation intensity as a function of time) is presented for an omnidirectional source and receiver located in an isovelocity water column having an infinite bottom loss. Appendix B contains a formulation for the pressure time series (pressure as a function of time) corresponding to either the coherent or incoherent reverberation intensity. A monostatic geometry and cw pulse is considered. The gener-

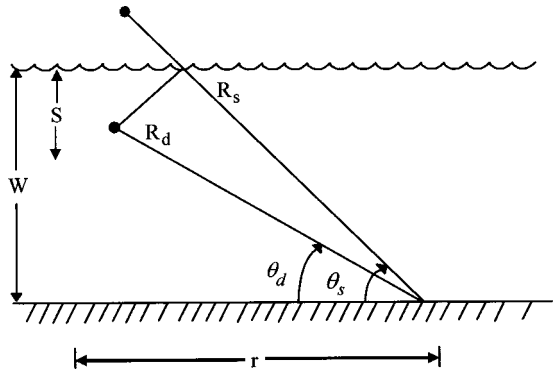


FIG. 15. Geometry and nomenclature assumed in the Lloyd mirror equation derivation.

alization to directional source and receivers, LFM and HFM pulses, bistatic geometries and finite bottom loss is easily accomplished.

B. Lloyd mirror

The geometry for the case considered is shown in Fig. 15, where r is the range, S is the source-receiver depth, W is the water depth, θ_d is the grazing angle at the bottom of the direct path, θ_s is the grazing angle of the surface reflected path, R_d and R_s are the one-way direct and surface reflected path lengths, respectively, and are given by

$$R_d = \sqrt{(r^2 + (W-S)^2)}, \tag{A1}$$

and

$$R_s = \sqrt{(r^2 + (W+S)^2)}.$$

The coherent pressure field for the one-way path can be expressed as

$$P(r,t) = p(r)f(t) = p_0[(e^{ikR_d}/R_d) - (e^{ikR_s}/R_s)]e^{-i\omega t}, \tag{A2}$$

where t is the time, ω is the angular frequency, p_0 is the pressure source level, and k is the wave number

$$\omega = 2\pi f, \quad k = \omega/c, \tag{A3}$$

and c is the sound speed. The usual Lloyd mirror pattern is contained in

$$-10\log|p|^2. \tag{A4}$$

The two way pressure field, based on the earlier notation, is given by

$$P_{tw}(r,t) = p(r)^2 f(t), \tag{A5}$$

and the two-way equivalent of the Lloyd mirror pattern by

$$-20\log|P_{tw}(r)|^2. \tag{A6}$$

C. Coherent reverberation

Having established the notation, the formulation for coherent backscattering will now be presented. A Lambert's law behavior for the backscattering strength, B_s, is assumed, i.e.:

$$B_s = \mu \sin(\theta_d)\sin(\theta_s). \tag{A7}$$

The appropriate form of the two-way scattered pressure field can then be written as

$$P(r,t) = p_0\sqrt{\mu}[SD(1/R_d)e^{ikR_d} - SS(1/R_s)e^{ikR_s}]^2 e^{-i\omega t}, \tag{A8}$$

where $SD = \sqrt{\sin\theta_d}$ and $SS = \sqrt{\sin\theta_s}$. Notice that there are actually four paths involved: (1) source-bottom-source, (2) source-surface-bottom-source, (3) source-bottom-surface-source, and (4) source-surface-bottom-surface-source. An extension of the formulation to a receiver at a different depth than the source is easily implemented.

The time dependence at the source is assumed to be a cw pulse of the form

$$e^{-i\omega t} \quad \text{for } 0 < t < \tau \quad \text{and zero otherwise.} \tag{A9}$$

In this expression, τ is the pulse length. A LFM pulse can also be modeled by using the appropriate form. There are conceptual difficulties in modeling full coherence that involve normalization and spurious behaviors in time. Instead, a special form of coherence is assumed. In the simple case considered, the reverberation at any time t arises from the contribution of the multipaths from all elemental scattering areas that satisfy the travel time requirements determined by the geometry and the transmitted pulse length. Since the elemental areas are at different ranges and the respective multipaths, in general, have different contributing arrivals can differ by as much as the duration of the transmitted pulse. In the usual reverberation formulation these contributions are all added together incoherently. In a completely coherent computation, these contributions are all added together coherently. In the present formulation it is assumed that only those contributions arriving back at the receiver from the same elemental scattering area at time t will add together coherently and these will form the total reverberation.

Such an elemental area is given by $rdr(d\theta)$ located at range r and azimuthal angle θ. As the angle θ changes so does the location of the elemental area and, as a consequence, contributions arising from different azimuthal angles will add together incoherently. Since only a monostatic situation is considered here and since there is not environmental or geometric dependence on azimuthal angle, these contributions are independent of that angle. Integration over the azimuthal angle introduces a factor of 2π and, as a consequence, the two-way backscattered intensity at time t for a source level intensity I_0 is given by

$$I_{\text{coh}}(t) = 2\pi\mu I_0 \int rdr|G_d(r,t) + G_{sd}(r,t) + G_s(r,t)|^2, \tag{A10}$$

where

$$G_d(r,t) = (SD/R_d)^2 e^{i(2kR_d - \omega t)}, \tag{A11}$$

$$\text{for } 2R_d/c < t < 2R_d/c + \tau \quad \text{and zero otherwise,} \tag{A12}$$

$$G_{sd}(r,t) = -2[SD\cdot SS/(R_d R_s)^2]e^{i[k(R_d+R_s)-\omega t]}, \tag{A13}$$

$$\text{for } (R_d+R_s)/c < t < (R_d+R_s)/c + \tau \quad \text{and zero otherwise,} \tag{A14}$$

and

$$G_s(r,t)=(SS/R_s)^2 e^{i(2kR_s-\omega t)}, \tag{A15}$$

for $2R_s/c<t<2R_s/c+\tau$ and zero otherwise. (A16)

In contrast, the appropriate form for incoherent reverberation is given by

$$I_{\text{inc}}(t)=2\pi\mu I_0 \int r dr\{|G_d(r,t)|^2+(1/2)\cdot|G_{sd}(r,t)|^2 + |G_s(r,t)|^2\}. \tag{A17}$$

Notice that Eq. (A17) contains no phase information.

Based on the earlier coherent formulation, it may be seen that at any range for which

$$\tau<[(R_s(r))-R_d(r)]/c, \tag{A18}$$

there can be no overlap in time of arrivals and therefore all paths add incoherently. This is one extreme. The middle ground is when

$$\tau>[(R_s(r))-R_d(r)]/c, \tag{A19}$$

and then arrivals from the direct and the surface reflected paths will overlap and add coherently. The nonoverlap portions (beginning for the direct and ending for the surface reflected) will add incoherently. Another extreme is when the overlap condition is satisfied but τ is very large. In this case, the interference pattern is smeared and the reverberation approaches the incoherent behavior.

APPENDIX B

A. Pressure formulation

The two new formulations for the reverberation time series developed in this section will be referred to as coherent pressure and incoherent pressure. In order to specify the time series, it is necessary to discretize the range r_i, and consider the pressure at time t to be the contribution from all such elements Δr centered about r_i where

$$r_{i+1}=r_i+\Delta r. \tag{B1}$$

A particular realization of the pressure time series of the coherent reverberation can be written as

$$p_{\text{coh}}(t)=\sqrt{2\pi\mu}p_0\sum_i [G_d(r_i,t)+G_{sd}(r_i,t) + G_s(r_i,t)]e^{i2\pi\phi(r_i)}\sqrt{r_i\Delta r} \tag{B2}$$

and a particular realization of the pressure time series of the incoherent reverberation can be written as

$$p_{\text{inc}}(t)=\sqrt{2\pi\mu}p_0\sum_i [G_d(r_i,t)e^{i2\pi\phi_d(r_i)} + G_{sd}(r_i,t)\cdot(1/2)\cdot(e^{i2\pi\phi_{sd}(r_i)}+e^{i2\pi\phi_{ds}(r_i)}) + G_s(r_i,t)e^{i2\pi\phi_s(r_i)}]\sqrt{r_i\Delta r}. \tag{B3}$$

In these expressions, the various $\phi(r_i)$ are independent pseudorandom variables having a uniform probability density between 0 and 1. The definitions of the various G contained in Eqs. (A11)–(A16) are unchanged except that R_d, for example, as defined in Eq. (A1), is replaced by $R_d(r_i)$, where

$$R_d \Rightarrow R_d(r_i)=\sqrt{r_i^2+(W-S)^2}. \tag{B4}$$

The expectation value of p_{coh} is given by

$$\langle|p_{\text{coh}}(t)|^2\rangle=2\pi\mu I_0\Big\langle \sum_i |G_d(r_i)+\ldots|^2 r_i\Delta r + \sum_{i\neq j}\{G_d(r_i)+\ldots\}\cdot\{G_d(r_j)+\ldots\}^* \times e^{i2\pi[\phi_i(r_i)-\phi_j(r_j)]}\sqrt{r_i r_j}\cdot\Delta r\Big\rangle \tag{B5}$$

and in the limit as $\Delta r\to 0$

$$\langle|p_{\text{coh}}(t)|^2\rangle_{\lim\Delta r\to 0}=I_{\text{coh}}(t) \quad \text{of Eq. (A10)} \tag{B6}$$

Similarly

$$\langle|p_{\text{inc}}(t)|^2\rangle_{\lim\Delta r\to 0}=I_{\text{inc}}(t) \quad \text{of Eq. (A17).} \tag{B7}$$

Numerical examples of the coherent reverberation time series are shown in Fig. 10.

[1] R. Young, "Image interference in the presence of refraction," J. Acoust. Soc. Am. **19**, 1–7 (1947).
[2] R. Urick, *Principles of Underwater Sound for Engineers* (McGraw-Hill, New York, 1967), pp. 110–113.
[3] W. Carey, I. Gereben, and B. Brunson, "Measurement of sound propagation downslope to a bottom-limited sound channel," J. Acoust. Soc. Am. **81**, 244–257 (1987).
[4] Lyman Spitzer, Jr., *Physics of Sound in the Sea* (Peninsula, Los Altos, CA, 1989), p. 301.
[5] K. MacKenzie, "Bottom reverberation for 530- and 1030-cps sound in deep water," J. Acoust. Soc. Am. **33**, 1498–1504 (1961).
[6] L. Dilman and J. Bishop (private communication).
[7] J. Preston, D. McCammon, F. Desharnais, and D. Ellis, "Some shallow water inferences from the Celtic duet experiment," SPAWAR CST/LLFA-WP-EVA-44, August 1996. (Available DTIC, ADA 314014).
[8] S. McDaniel, "Seafloor reverberation fluctuations," J. Acoust. Soc. Am. **88**, 1530–1535 (1990).
[9] M. Gu and D. Abraham, "Using McDaniel's model to represent non-Rayleigh reverberation," IEEE J. Ocean. Eng. **26**, 348–357 (2001).
[10] M. Sundvik, E. Rabe, and M. Vaccaro, "Mapping seabed acoustic reflection and scattering coefficients using tactical sonar systems," *Proceedings of the MTS/IEEE OCEANS'99 Conference*, 1999, pp. 1070–1075.
[11] H. Weinberg, "Generic sonar model," *Proceedings of the MTS/IEEE OCEANS'82 Conference*, 1982, pp. 201–205.
[12] H. Weinberg, "CASS roots," *Proceedings of the MTS/IEEE OCEANS 2000 Conference*, 2000, Vol. 2, pp. 1071–1076.
[13] R. Keenan, "An introduction to GRAB eigenrays and CASS reverberation and signal excess," *Proceedings of the MTS/OCEANS 2000 Conference*, 2000, Vol. 2, pp. 1065–1070.
[14] H. Weinberg and R. Keenan, "Gaussian ray bundles for modeling high-frequency propagation loss under shallow-water conditions," J. Acoust. Soc. Am. **100**, 1421–1431 (1996).
[15] H. Weinberg (private communication).

Extracting coherent wave fronts from acoustic ambient noise in the ocean

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

(Received 16 May 2003; revised 24 June 2004; accepted 16 July 2004)

A method to obtain coherent acoustic wave fronts by measuring the space–time correlation function of ocean noise between two hydrophones is experimentally demonstrated. Though the sources of ocean noise are uncorrelated, the time-averaged noise correlation function exhibits deterministic waveguide arrival structure embedded in the time-domain Green's function. A theoretical approach is derived for both volume and surface noise sources. Shipping noise is also investigated and simulated results are presented in deep or shallow water configurations. The data of opportunity used to demonstrate the extraction of wave fronts from ocean noise were taken from the synchronized vertical receive arrays used in the frame of the North Pacific Laboratory (NPAL) during time intervals when no source was transmitting. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1797754]

PACS numbers: 43.30.Pc, 43.50.Rq, 43.60.Fg [WLS] Pages: 1995–2003

I. INTRODUCTION

Acousticians have used incoherent processing of ambient noise to achieve small scale imaging in the ocean[1] in the same way as optical imaging is performed from incoherent light. Here, we show that it is feasible to transform a collection of apparently incoherent noise sources into a coherent, large scale, imaging field. Actually, every individual source of noise (e.g., a collapsing bubble) in the ocean generates an acoustic field that is potentially coherent when received between two points after long range propagation. However, this small coherent component at each receiver point is buried in the spatially and temporally incoherent field produced by all the widespread noise sources distributed over the ocean. We demonstrate in this paper that a long-time cross-correlation process extracts coherent wave fronts from ambient noise without the support of any identifiable source. This means that noise could be used as a potential coherent source in the ocean leading to the concept of a self-imaging process. These general results reveal the potential information content of random noise fields in a natural environment.

The sources of ocean surface noise[2–4] (natural and man-made) as well as the subsequent average spatial distribution of ocean noise[5–7] have been studied extensively. However, because the instantaneous distribution of all the mutually incoherent sources is extremely variable in space and time, a robust, *space–time* observable of ocean noise is difficult to identify. In this work, we derive and verify with data a (space–time) wave-front coherence property of surface noise not previously explored. With simple signal processing these wave fronts that are strongly related to the time-domain Green's function (TDGF) between observation points, are easily observable. Though incoherent imaging with ambient noise has been demonstrated,[1,8] the goal of this research is to lay the foundation for substituting ambient noise in coherent imaging procedures such as tomography[9] that typically require an active source or other noise based imaging methods in need of some coherence (e.g., Ref. 10).

The results presented here find their origin in the helioseismologic research of Rickett and Claerbout.[11] Their work is based on the following conjecture: ''By cross-correlating noise traces recorded at two locations (...), we can construct the wavefield that would be recorded at one location if there was a source at the other.'' This statement has been experimentally confirmed by the ultrasonic research of Weaver and Lobkis,[12,13] who have shown that the long-time, two-point correlation of random (Brownian motion) noise in an aluminum block cavity yields the deterministic time-domain Green's function between the two points. Further, there have been recent results of a related nature by Campillo and Paul[14] based on the cross correlation of the diffuse coda of identifiable seismic events. Here, we demonstrate that the temporal arrival structure of the two-point acoustic TDGF can be analogously approximated in the ocean using ambient noise. In particular, the long-time correlation between a receiver and elements of a vertical array of receivers yields a wave-front arrival structure at the array that is identical to the structure of the TDGF except that the amplitudes of the individual wave fronts are shaded by the directionality of the noise sources. The Green's function emerges in both the cavity and ocean cases from those correlations that contain *noise sources whose acoustic field passes through both receivers*. The Weaver result relies on noise sources that are contributing to the construction of the TDGF that are distributed over three dimensions (3D) leading to the so-called modal equipartition in the cavity. The Campillo result is also based on modal equipartition,[15] in this case due to multiple scattering in the earth upper crust, but they build up their TDGF from a set of identifiable events. For the ocean environment considered here, there is no significant 3D scattering in the fre-

[a]Electronic mail: philippe@mpl.ucsd.edu
[b]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.

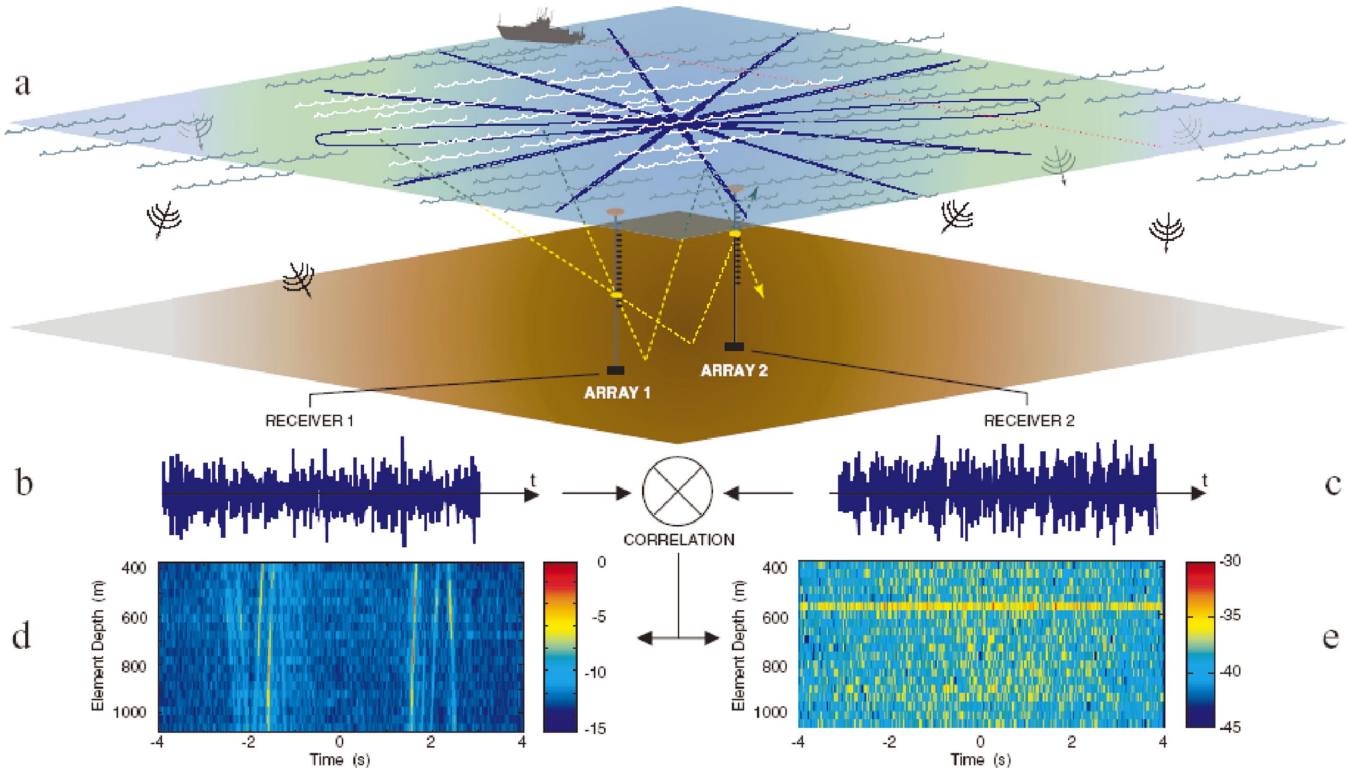


FIG. 1. (a) Two arrays are depicted at a separation distance R. A schematic of the directivity pattern of the time-domain correlation process between two receivers on each array is projected on the ocean surface. Only a discrete set of lobes have been displayed that correspond to noise sources whose emission angle is equal to $-60°$, $-30°$, $0°$, $30°$, $60°$, and $90°$. Each angular lobe depends on the central frequency and bandwidth and corresponds to a delay time in the correlation function. For the case of equally distributed ambient noise sources, the broad end-fire directions will contribute coherently over time to the arrival times associated with the TDGF while the contribution of the narrow off-axis sidelobes will average down. For the case of shipping noise, coherent wave fronts emerge only when there is sufficient intersection of the shipping paths with the end-fire beams. However, if there is a particular loud shipping event, it will dominate so that either impractically long correlation times are needed, or discrete events should be filtered out. (b) and (c) The correlation process is done using time-domain ambient noise simultaneously recorded on two receivers in arrays 1 and 2. (d) Spatial temporal representation of the wave fronts obtained from the correlation process between a receiver in array 1 at depth 500 m and all receivers in array 2 separated by a distance $R=2200$ m. The arrival structure of the correlation function is composed of the direct path, surface reflected, bottom reflected, etc., as expected in the TDGF. Because data were taken in 20 min segments, the result is a combination of three segments in order to accumulate the wave-front structure from what turns out to be shipping noise (see Fig. 7 for further explanation). The correlation function is plotted in a dB scale and normalized by its maximum. (e) The same correlation processing is performed on data that have not been recorded at the same time on the two arrays. In (d) and (e), the x and y axes correspond to the time axis of the cross-correlation function and receiver depth, respectively. The correlation functions are plotted in a dB scale and normalized by the maximum of (d).

quency regime of the available data. Without the presence of identifiable events, it is shown that only noise sources aligned along the line between the receivers contribute over a long-time correlation.

This paper is structured as follows. In Sec. II, we describe the basic principles of the ambient noise cross-correlation technique in the ocean. Section III presents experimental results using a known shipping source verifying that the residual of the time-averaged correlation function indeed comes from noise sources located in the endfire direction of the receivers. Section IV starts with a theoretical derivation of the ambient noise time-domain correlation function with noise sources distributed either in the volume or at the surface of the waveguide. The case of shipping noise is also explored. Simulations in shallow water point out the similarities and differences with the TDGF. Finally, we present in Sec. V coherent wave fronts obtained from ambient noise data simultaneously recorded on four coplanar hydrophone arrays.

II. BASIC PRINCIPLES AND EXPERIMENTAL DEMONSTRATION

The overall concept is summarized in Fig. 1 in which ambient noise at two receivers [Figs. 1(b) and 1(c)] in array 1 and array 2 are pairwise cross-correlated. In Figs. 1(d) and 1(e), this correlation is a function of delay time and vertical position (depths of receivers in array 2) as is the TDGF between a position in array 1 and the receivers of the array 2. The directivity pattern of the correlation process for a set of incoming angles is schematically projected on the ocean surface [Fig. 1(a)]. For each incident angle, the directivity beam depends on the central frequency and bandwidth (see Sec. III for details) and corresponds to a delay time in the correlation function. This directivity pattern shows how noise sources all over the ocean surface participate to the noise cross-correlation function. Noise sources inside the same directivity beam add coherently while noise sources inside different directivity beams average incoherently (i.e., there is a different delay time associated with each beam). In Fig. 1(a), we see that two broader beams (so called end-fire beams) are

aligned on the axis between the two arrays. Because of their size, these beams yield a larger contribution to the noise correlation function. For example, the two dashed lines in Fig. 1(a) show the coherent contribution from surface noise sources that travel through both receivers. We will show in Sec. III that noise sources inside the end-fire beams provide the residual time-averaged coherence between the two arrays.

In Fig. 1(d) we show a display that is an actual multiday composite of three 20 min segments of simultaneous recordings of data discussed in Sec. IV. The wave fronts in the display obtained from the cross-correlation process are symmetric in time with respect to the zero time of the correlation function because noise sources were distributed on both sides of the arrays. Note that there is no correlation between data not recorded at the same time [Fig. 1(e)] confirming the hypothesis that coherent wave fronts built up over time from individual noise sources whose acoustic field propagates through both receivers.

The basic difference between the Weaver cavity configuration and ocean noise is the 3D reverberation physics of the cavity versus the two-dimensional ocean waveguide physics of traveling waves in the horizontal direction. For the latter, this means that a ray aligned along the receiver axis will pass through both receiver points by reflection or refraction; however, if the ray has a horizontal component not along the horizontal line between the receivers, it cannot be reflected back to the second receiver and therefore cannot contribute to building up the coherent wave fronts between the receivers. This geometrical interpretation does not apply to the cavity but is still valid in the case of Campillo's multiple scattering earth model. Indeed, even if the seismic events are not aligned with the two seismometers, we believe that the main contribution to the average correlation function comes from the scatterers present inside the end-fire beams. Indeed, the direct path is built from those scatterers in the upper crust of the earth that behave as secondary sources and re-direct the incident field so that part of the wave travels on a straight line through both receivers.

III. EXPERIMENTAL DEMONSTRATION OF LONG-TIME TWO-POINT CORRELATION PROCESSOR

The underlying physics of this technique relies on long-time cross correlation of ambient noise data. Cross correlation of acoustic data between hydrophones is a very common signal processing tool to detect and locate sources in the ocean. The difference here is that we are not interested in the actual noise sources but in the residual coherence between the hydrophones. This coherence, which corresponds to the temporal arrival structure of the TDGF, is extracted when the cross correlation is performed on long-duration time series. In the case of nonstationary noise sources (surface noise or noise generated by a ship in motion), the signature of the noise sources in the cross-correlated signal averages out and disappears while the coherent paths between the two hydrophones remain.

A separate experimental demonstration of this process has been performed from data simultaneously recorded on two sono-buoys at a few hundred meters from each other in a shallow water environment. Noise was generated during 16 min in the 100–300 Hz frequency interval by a ship whose track is represented in Fig. 2. The two 16-min-long time series are then cross correlated using different time windows (Fig. 3). When the correlation is performed on 1 s duration time series [Fig. 3(a)], the ship track is clearly observed. If the length of the cross-correlated time series is increased to 5, 10, 20, and up to 30 s [Figs. 3(b)–3(e), respectively], the signature of the ship track tends to disappear and the only signal left is obtained when the ship crosses the end-fire main lobes (Fig. 2). This signal exhibits different bottom and surface-reflected paths classically found in a shallow water environment. We will show in Sec. IV that it converges to the arrival structure of the TDGF between the two sono-buoys when averaged over several different ship tracks. Obviously, we observe in Figs. 3(a)–3(e) that the longer the correlation window, the higher the signal-to-noise ratio because more acoustic sources participate coherently to correlation function. Then, for longer time series [Fig. 3(f)], the correlation pattern does not change because no more coherent sources were present in the signal. Assuming that the speed of the ship was constant during the track, it generates a uniform density of sources over time. For long time windows, the signal-to-noise ratio of the correlation process can be defined as the ratio of the number of coherent versus incoherent sources inside the recording time window. Following the geometrical interpretation developed in Sec. II, this ratio corresponds to the area enclosed by the end-fire beam to a non-end-fire beam.

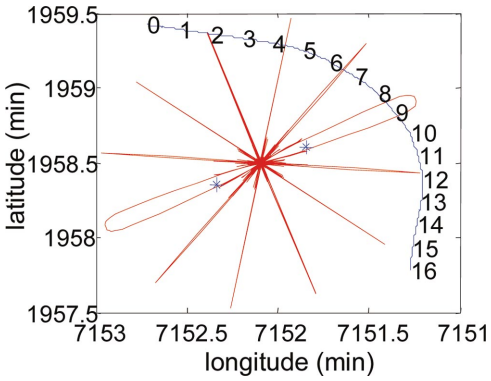


FIG. 2. Representation in latitude-longitude coordinates of the 16-min-long ship track (blue full line) with respect to the sono-buoys location (blue "∗"). The approximate distance between the sono-buoys is $R \sim 650$ m. The average ship speed was constant and equal to 4.8 m/s. The directivity pattern of the time domain cross-correlation process between the two sono-buoys is plotted in red. Only a discrete set of lobes have been displayed that correspond to noise sources whose emission angle is equal to −60°, −30°, 0°, 30°, 60°, and 90°.

The directivity pattern $B(\theta,\theta_0)$ in the direction θ of the correlation function between two receivers for an incident wave in the direction θ_0 can be written as

$$B(\theta,\theta_0) = \int_{\omega-\Delta\omega/2}^{\omega+\Delta\omega/2} \left[1+\exp\left(i\frac{\omega' R}{c}\cos(\theta)\right)\right] \times \left[1+\exp\left(-i\frac{\omega' R}{c}\cos(\theta_0)\right)\right] d\omega', \quad (1)$$

where R is the distance between the two receivers, $\omega = 2\pi f$

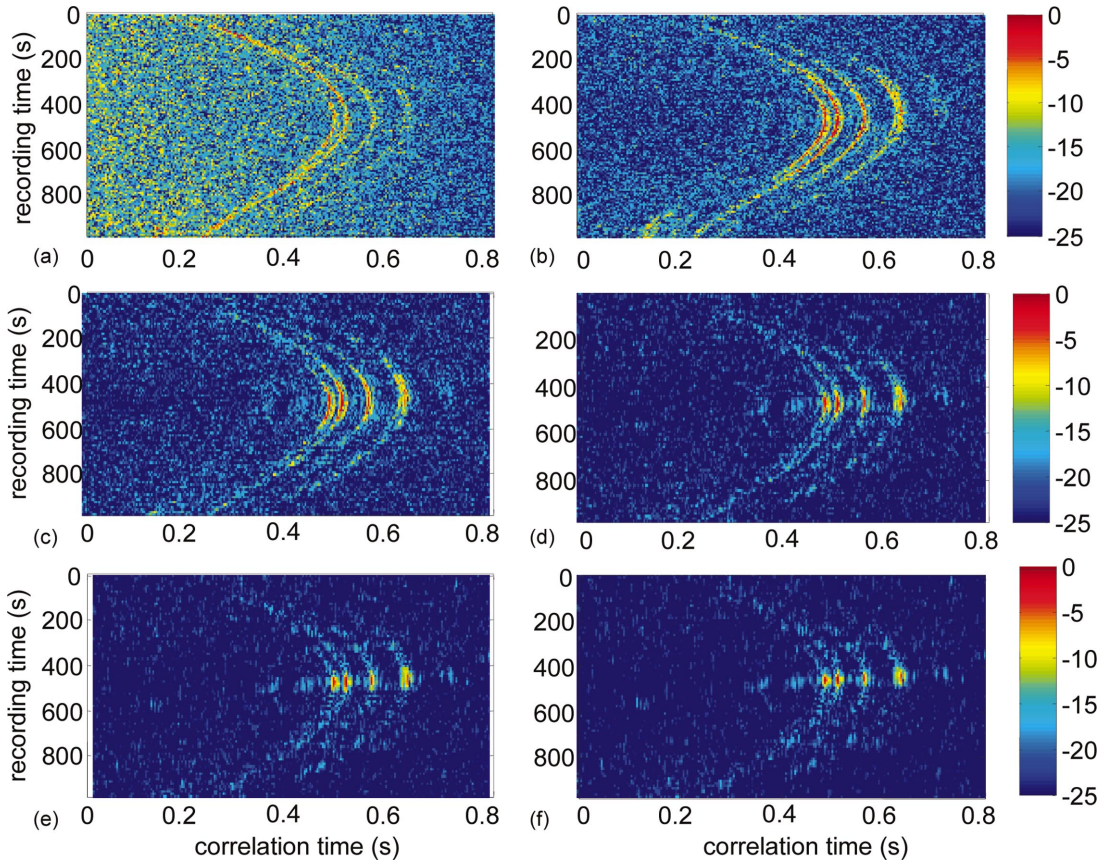


FIG. 3. Representation of the temporal evolution of the time-domain cross-correlation function between the two sono-buoys along the 16-min-long ship track. The *x* and *y* axes correspond to the time axis of the correlation function and the ship position, respectively. The duration of the time windows on which the cross correlation is performed is (a) 1 s, (b) 5 s, (c) 10 s, (d) 20 s, (e) 30 s, and (f) 40 s. Each cross-correlation pattern is normalized by its maximum. The color scales are in dB.

the central angular frequency, and $\Delta\omega$ the frequency bandwidth. The angle θ and θ_0 are defined with respect to the axis of the two receivers. After development and integration, we obtain the following from Eq. (1).

(1) If $\theta_0 \neq 0$ and $\delta\theta \ll \theta_0$,

$$B(\theta_0+\delta\theta,\theta_0) \simeq 1-\frac{\delta\theta^2}{2}\left(\frac{2R\sin(\theta_0)}{c}\right)^2\left(\omega^2+\frac{\Delta\omega^2}{12}\right). \tag{2}$$

(2) If $\theta_0=0$ or $\theta_0=\pi$ and $\delta\theta \ll \theta_0$,

$$B(\delta\theta,0)=B(\delta\theta+\pi,\pi)\simeq 1-\frac{\delta\theta^4}{8}\left(\frac{R}{c}\right)^2\left(\omega^2+\frac{\Delta\omega^2}{12}\right). \tag{3}$$

Equations (2) and (3) show that the end-fire beams (for $\theta_0=0$ or $\theta_0=\pi$) are broader. It follows that the ratio of the area enclosed by the end-fire beam to a non-end-fire beam is

$$\text{Ratio}(\theta_0)=\sqrt{\frac{\omega}{c}R}\sin(\theta_0)\left(1+\frac{1}{12}\left(\frac{\Delta\omega}{\omega}\right)^2\right)^{1/4}. \tag{4}$$

Taking $R\sim 650$ m, $f=\Delta f=200$ Hz, $c=1500$ m/s we get a maximum signal-to-noise ratio of 25 dB, which is in good agreement with the data [Fig. 3(e)]. The maximum signal-to-noise ratio is obtained for a time window $T\sim 30$ s that is approximately the time spent by the ship in the end-fire beam.

IV. THEORY AND SIMULATION

The analogy to the Weaver volume cavity noise would ideally be ocean volume noise sources uniformly distributed throughout the water column. However, ocean noise is dominated by surface noise sources that are typically uniformly distributed over the ocean surface as one goes to higher frequencies (>1 kHz, see Ref. 2). For the lower frequency case (~100 Hz), data are dominated by shipping noise, and while the concept remains the same, the relative amplitudes of the wave fronts that become observable will be dependent on the specific shipping distribution during the recording time interval. As a matter of fact, it is shown here by theory and

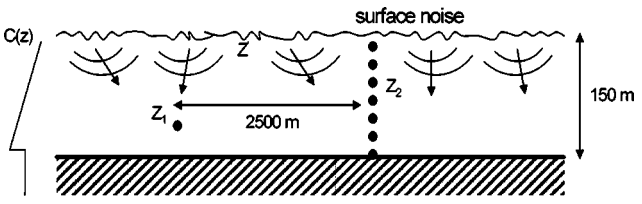


FIG. 4. Schematic of the waveguide in which simulations are performed. A source receiver in $z_1=100$ m and an array of receivers are located at a distance $R=2500$ m from each other in a 150-m-deep shallow water environment. The surface noise sources are at depth $z'=1$ m. The sound-speed profile decreases linearly from 1500 m/s at the surface to 1480 m/s at the bottom. The bottom sound speed, density, and attenuation are 1800 m/s, 1800 kg/m^3, and 0.05 db/λ, respectively.

simulation that shipping produces similar results to the surface noise case but that distant shipping emphasizes more horizontally traveling wave fronts than nearby ships for which the wave fronts are more vertical.

The ocean acoustic environment is typically treated as a waveguide[16] such that the propagation between points 1 at depth z_1 and 2 at depth z_2 separated by horizontal range R (Fig. 4) is given, in its simplest form, by a normal mode expansion:

$$G_\omega(R,z_1,z_2)=\frac{iS(\omega)}{4\rho}\sum_n U_n(z_1)U_n(z_2)H_0^{(1)}(k_nR), \quad (5)$$

where $U_n(z)$ is the depth-dependent eigenfunction associated with wave number k_n, ρ is the density at the source location, and $S(\omega)$ is the source spectrum. When integrated over the frequency bandwidth, Eq. (5) becomes the TDGF:

$$G_t(R,z_1,z_2)=\int d\omega\, G_\omega(R,z_1,z_2)\exp(-i\omega t). \quad (6)$$

The wave-front structure of the Green's function results from modes with similar group speeds constructively interfering over frequency; mathematically the wave fronts can be shown to emerge, for example, from a stationary phase evaluation[17] of Eq. (6) that results from the condition $d(k_nR-\omega t)=0$.

A. Volume-noise case

In order to demonstrate the connectivity between the volume-cavity result of Weaver and the correlation of ocean waveguide noise as formulated by Kuperman and Ingenito,[6,16] we first modify the latter theory to include volume sources. This shows how the waveguide TDGF emerges in analogy to the cavity case. Following Ref. 6, the modal decomposition in the frequency domain of the volume noise cross-correlation function between points 1 and 2 is

$$C_\omega(R,z_1,z_2)=\frac{i\pi Q^2(\omega)}{k^2(\omega)}\int\frac{1}{\rho^2(z')}dz'\sum_{n,m}U_n(z')U_n(z_1)\times U_m(z')U_m(z_2)\frac{1}{k_n^2-k_m^{*2}}\times[H_0^{(1)}(k_nR)-H_0^{(1)}(-k_m^*R)], \quad (7)$$

where $Q^2(\omega)$ is the power spectrum of the noise sources and z' the depth of the noise sources. To obtain Eq. (7), we assumed that each sheet of noise sources at depth z' are uncorrelated and that noise sources are uniformly distributed in the whole water column. The spatial integration over the depth of the waveguide can be performed using the orthogonality condition of modes

$$\int\frac{U_n(z')U_m(z')}{\rho(z')}dz'=\delta_{nm}.$$

Doing so, we neglect the tail of the mode in the bottom of the waveguide and we assume a constant density in the water column $\rho(z')=\rho$. We also suppose that k_n is a complex number of the form $k_n=K_n+i\alpha_n$ (with $K_n\gg\alpha_n>0$), where α_n is the modal attenuation coefficient. The final expression for the ambient noise correlation function is then:

$$C_\omega(R,z_1,z_2)=\frac{\pi Q^2(\omega)}{4\rho k^2(\omega)}\sum_n U_n(z_1)U_n(z_2)\frac{1}{\alpha_nK_n}\times[H_0^{(1)}(k_nR)-H_0^{(1)}(-k_n^*R)]. \quad (8)$$

After frequency integration, the time-domain ambient noise correlation function is obtained from $C_t(R,z_1,z_2)=\int d\omega\, C_\omega(R,z_1,z_2)\exp(-i\omega t)$. The two Hankel functions in Eq. (8) represent two wave fronts traveling between receivers 1 and 2 in opposite directions. Physically speaking, the two wave fronts arise from a uniform volume noise distribution so that at any point, noise is coming from all directions. Invoking the modal normalization condition, $\int[U_n^2(z)/\rho(z)]dz=1$, the amplitude factor

$$\alpha_nK_n=\int\alpha(z)\frac{\omega}{c(z)}\frac{U_n^2(z)}{\rho(z)}dz$$

(where α(z) is the depth-dependent absorption[16]) is slowly dependent on mode number. Thus, the modal decomposition in Eq. (8) is very close to the Green's function decomposition as written in Eq. (5). This means that the correlation function obtained from volume ambient noise recorded at two receivers in a waveguide is a good approximation of the Green's function between the two points. The reasoning done through Eqs. (7) and (8) is the waveguide equivalent of Weaver's cavity approach in which he supposed the modes equipartition due to an uniform noise distribution in the cavity.

B. Surface-noise case

Assuming a sheet of noise sources located at a given depth z' only, Kuperman and Ingenito obtained the following expression for the noise correlation function:

$$C_\omega(R,z_1,z_2)=\frac{\pi Q^2(\omega)}{4\rho^2(z')k^2(\omega)}\sum_n U_n(z_1)U_n(z_2)\times\frac{U_n^2(z')}{\alpha_nK_n}[H_0^{(1)}(k_nR)-H_0^{(1)}(-k_n^*R)]. \quad (9)$$

In the case of noise sources distributed at the surface of the ocean, a monopole source below the pressure release ocean surface behaves as a dipole structure. The amplitude factor $U_n^2(z')/\alpha_nK_n$ in Eq. (9) results then from a combination of the dipole behavior of the noise sources and the effect of attenuation over long ranges. Formally, the factor arises because there is no integral over the depth of sources as in the volume case; that is the only difference between the volume and surface cases. Since this amplitude term will not affect the stationary phase argument that synthesizes the wave fronts,[18] the time-domain surface noise correlation function $C_t(R,z_1,z_2)$ will exhibit the same wave-front structure as the two point Green's functions though the amplitude of the wave fronts will be different. We therefore conclude that

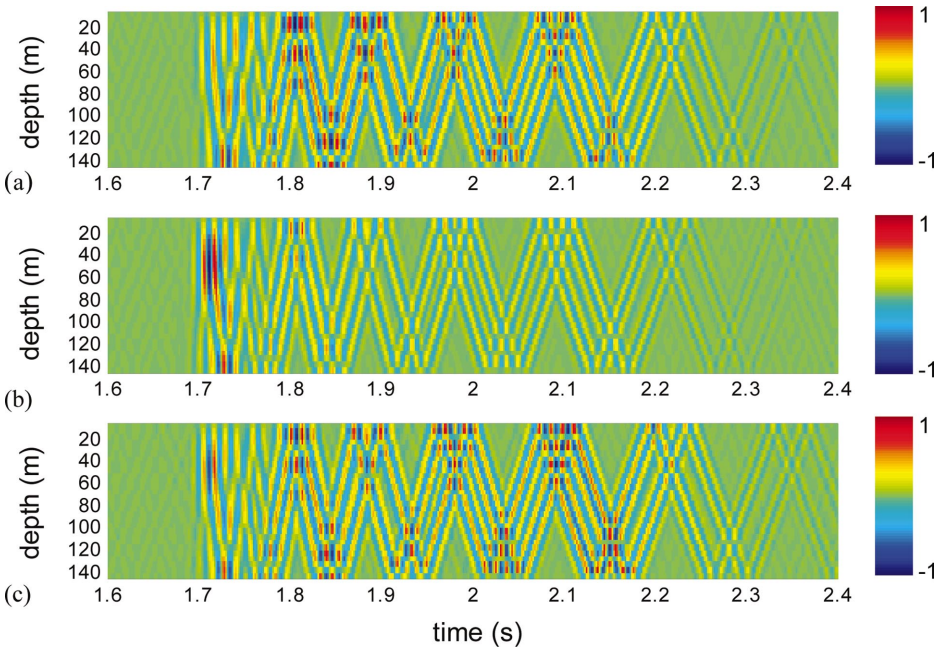


FIG. 5. Spatial-temporal representation in a linear scale of (a) the time-domain cross-correlation function of surface noise computed between a receiver in $z_1 = 100$ m and a receiver array at a distance R in a shallow water waveguide (see Fig. 4). The x and y axes correspond to the time axis of the correlation function and the receiver depth, respectively. (b) The time domain Green's function (TDGF) computed between a source in z_1 and a receiver array at the same distance R; (c) the TDGF computed in the same configuration as in (b) for a vertical dipole source at z_1. Simulations have been performed in the [50–150 Hz] frequency bandwidth.

after the temporal averaging underlying Eq. (9), we obtain coherent wave fronts from ocean surface noise. However, these coherent wave fronts only constitute an approximation of the TDGF, the excitation of each mode being weighted by a dipole shading. This approach is similar to Campillo's result in the sense that only the Rayleigh wave was reconstructed between two seismometers using the late coda of seismic events. The longitudinal and shear waves that classically participate to the TDGF between two points at the earth surface are missing in the correlation function because they are not properly excited by the scatterers present in the upper crust of the earth.

In Fig. 5, we confirm this theoretical approach with simulations using the spectral model[19] performed in a shallow water environment in the 50–150 Hz frequency bandwidth (Fig. 4). The time-domain correlation function of surface noise [Fig. 5(a)] is compared to the actual Green's function [Fig. 5(b)] between one source-receiver and an array of receivers. As expected, the same coherent wave fronts are observed but the amplitude of the higher-order reflected paths is different. Consistent with the above-mentioned results, the Green's function computed with a vertical dipole source instead of an omnidirectional monopole excitation [Fig. 5(c)] shows an obvious similarity with the time-domain correlation function.

C. Shipping-noise case

For the pure shipping case, the correlation function results from the product of two Green's functions each of the form given in Eq. (5) but evaluated at the radial distances R_1 and R_2 between the ship and receivers 1 and 2, respectively:

$$C_\omega(R,z_1,z_2) \propto \frac{\tilde{Q}^2(\omega)}{\rho^2(z')} \sum_{n,m} U_n(z') U_n(z_1) U_m(z') \times U_m(z_2) H_0^{(1)}(k_n R_1) H_0^{(1)}(-k_m R_2), \tag{10}$$

where $\tilde{Q}^2(\omega)$ is the power spectrum of the shipping noise. As explained in Sec. III, the correlation process emphasizes the contribution from ships located in the end-fire beams. When a ship is located in the end-fire direction of the two receivers, we have $|R_1 - R_2| = R$. Evaluating the average contribution from a ship involves integrating Eq. (10) over the length L of the ship track inside the end-fire beam [see Fig. 1(a)] yielding an amplitude term $\sin c((k_n - k_m)(L/2))$, where $\sin c(\pi x) = \sin(\pi x)/\pi x$ stands for the sinus cardinal function. When the integration is performed on several ship paths whose accumulated path L is such that

$$L \gg \frac{1}{\min_{n,m}(|k_n - k_m|)},$$

it follows:

$$\langle C_\omega(R,z_1,z_2) \rangle \propto \frac{\tilde{Q}^2(\omega)}{\rho^2(z')} \sum_n \frac{U_n^2(z')}{\sqrt{k_n}} U_n(z_1) \times U_m(z_2) H_0^{(1)}(k_n R), \tag{11}$$

where the angular brackets $\langle\ \rangle$ refer to the average over an accumulation of ship events. When ship paths cover the whole ocean surface, the propagating wave in Eq. (11) is changed into two propagating waves in opposite directions as in Eq. (9). Equations (11) and (9) then become very similar in the sense that they both correspond to an amplitude-shaded approximate version of the TDGF.

Though we get the same forms for either broadband shipping or surface ambient noise, the signal-to-noise ratio (emergence of the wave fronts) of the two cases will be different as the schematic of Fig. 1 indicates and as discussed further in the following. For example, as opposed to ambient noise sources, shipping events could be possibly identified, normalized in amplitude, and integrated separately in the correlation function. This would balance the contribution of close/faraway ships with loud/low source levels. Furthermore, since shipping is episodic, it would be helpful to elimi-

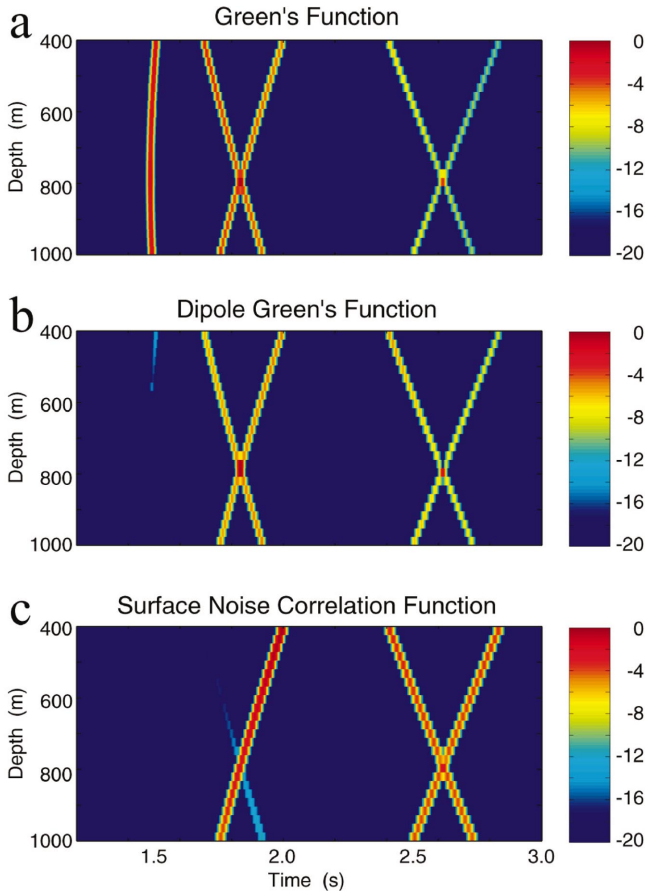


FIG. 6. Simulations of the Green's function (TDGF) and the time domain correlation function in an oceanic environment close to the experimental data. (a) Spatial-temporal TDGF from a source at depth 800 m to an array at a distance of 2400 m. (b) The same TDGF but source shaded by a dipole pattern similar to the radiation pattern of near-surface sources. (c) Surface noise correlation function between the same position of the source and the array showing a close similarity to (b). The *x* and *y* axes correspond to the time axis of the correlation function and the receiver depth, respectively. The color scales are in dB.

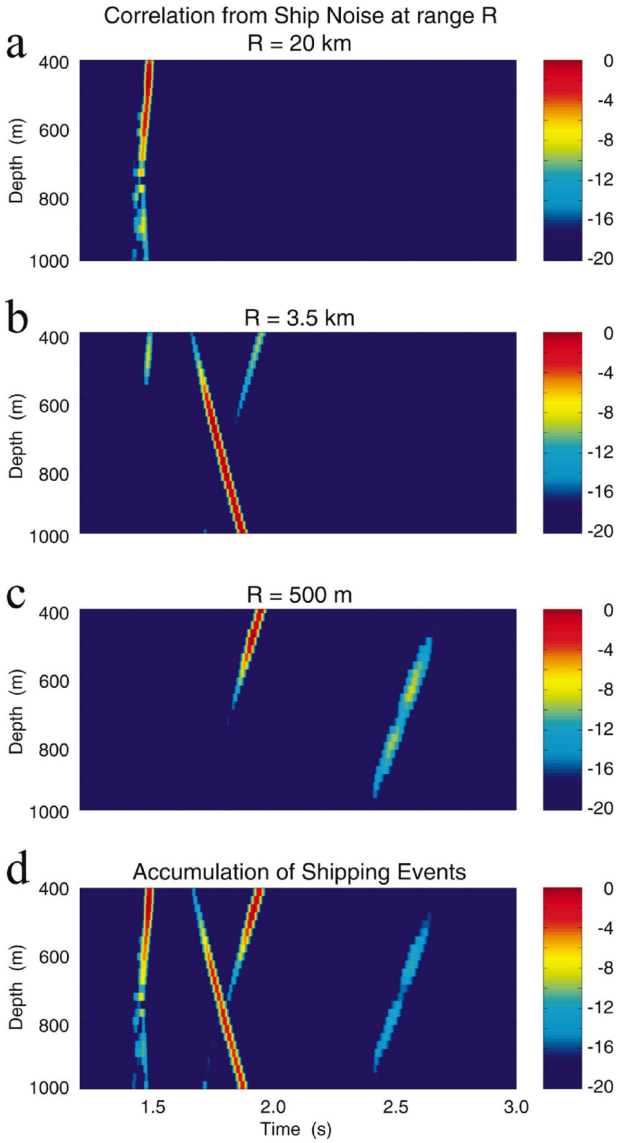


FIG. 7. Spatial temporal representation of the correlation function obtained in the case of near-surface sources (shipping noise) at various ranges in an oceanic environment close to the experimental data. The *x* and *y* axes correspond to the time axis of the correlation function and the receiver depth, respectively. (a) An $L=500$ m track for a ship at a 20 km distance. (b) An $L=500$ m track for a ship at a 3.5 km distance. (c) An $L=500$ m track for a ship at a 500 m distance. (d) Composite of (a), (b), and (c) indicating what one could expect from an accumulation of shipping episodes. The color scales are in dB.

nate high amplitude events from the data and thereby ''homogenize'' the noise in order to enhance the convergence to the expected correlation function. The best data set displays no specific events and is approximated well by the theory of uniformly distributed surface noise sources.

Simulations of the above-noted processes are given in Figs. 6 and 7 in which the TDGF is numerically computed with the surface noise correlation function or the shipping events correlation function. The ocean environment is close to the experimental configuration described in Sec. V [Fig. 8(a)]. As in Fig. 5, when the Green's function [Fig. 6(a)] is combined with the directivity of a dipole [Fig. 6(b)], the depth-time dependent pattern of the Green's function and the noise correlation function [Fig. 6(c)] look similar. However, the amplitudes of the wave fronts are still different because a dipole at 600 m does not have the same directivity pattern as a dipole at the surface in a depth-dependent sound speed profile.

Further, Fig. 7 refers to an end-fire ship track (source contributions from other directions vanish over long correlation time window, as shown in Fig. 3) at different ranges. In Figs. 7(a)–7(c), we observe that the wave-front arrivals have different amplitude emphasis—shorter ranges favoring more vertical paths in the correlation function. Figure 7(d), which is a composite of the various ranges, exhibits, as expected, the same arrival structure as in Fig. 6(a). Hence, in the shipping case, a collection of random shipping events will fill in the whole TDGF pattern over time. Roux and Fink[20] have theoretically examined the case of averaging correlations of deterministic sources over depth that theoretically results in modal equipartition.[12,14] However, as an impractical matter in ocean acoustics, it requires the insertion of active sources and is more akin to the above-discussed hypothetical volume source problem.

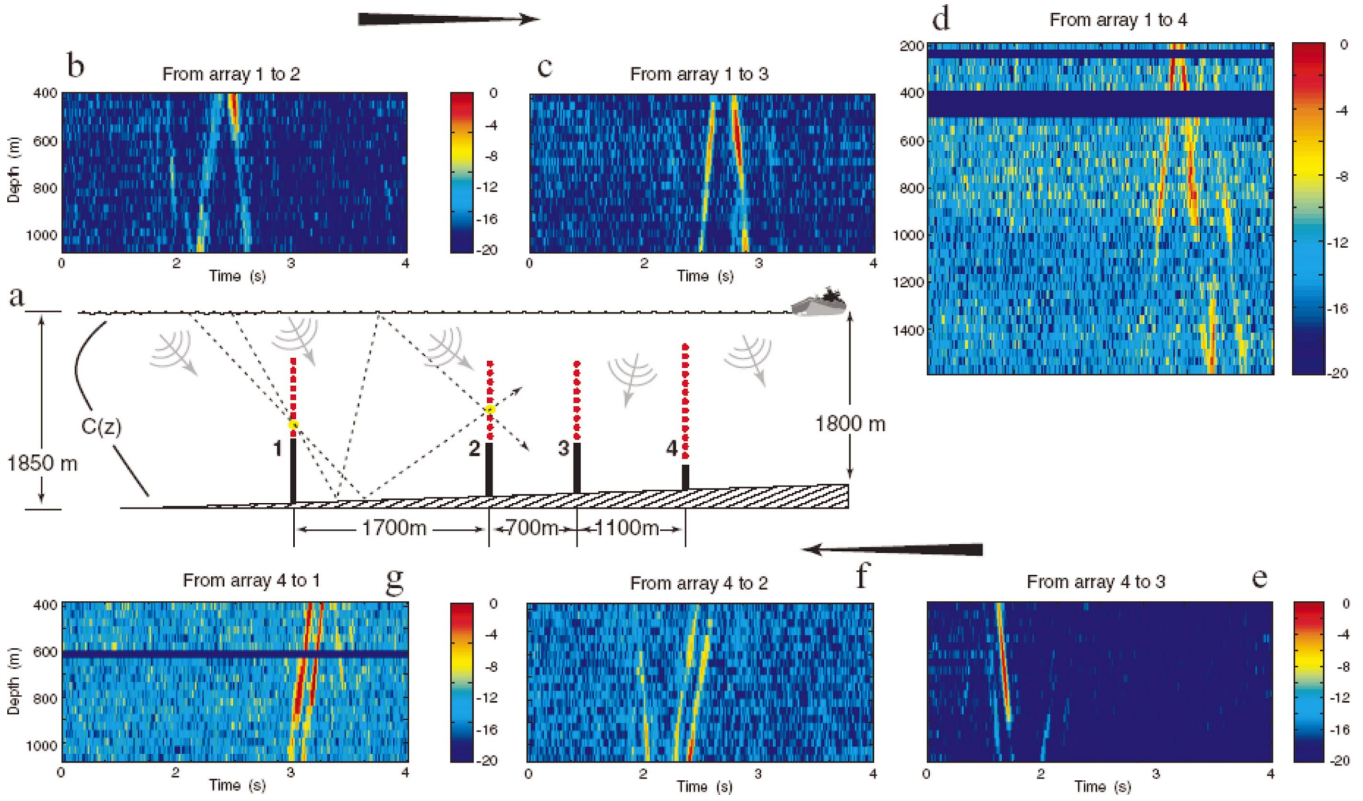


FIG. 8. The noise-derived amplitude-shaded TDGF extracted from the NPAL data. (a) The array geometry indicating a sloping bottom. Note that array 4 is made of twice as many elements as arrays 1–3. (b)–(d) Time-domain correlation functions between a receiver at depth 500 m in array 1 and all receivers in the other arrays. Traveling wave fronts are clearly observed in the direction of the arrow as if they emanated from the receiver in array 1. (e), (f), and (g) Time-domain correlation functions between a receiver at depth 500 m in array 4 and all receivers in the other arrays. Here we see traveling wave fronts in the direction of the arrow, opposite case (b), (c) and (d), as if they emanated from the receiver in array 4. The wave fronts for this direction are more vertical because of the slope effect—further confirming the correct extraction of the arrival structure of the TDGF. In (b)–(g), the x and y axes correspond to the time axis of the correlation function and the receiver depth, respectively. The color scales are in dB.

V. EXPERIMENTAL RESULTS WITH HYDROPHONE ARRAYS

The actual measurement and signal processing that correspond to the theoretical results of Eq. (9) is done in the time domain where the correlation function $C_{12}(t)$ is measured using $C_{12}(\mathrm{t})=\int S_1(\tau)S_2(t+\tau)d\tau$, where $S_1(t)$ and $S_2(t)$ are the ambient noise received on receivers 1 and 2 at time t. Note that the correlation processing requires data measurement that have a common clock time. We use data of opportunity from the NPAL program[21] originally taken for other purposes. Data correspond to different sets of 20 min simultaneous recording of ambient noise on four vertical arrays, filtered between 70 and 130 Hz. Despite the obvious presence of shipping noise in this frequency bandwidth, high amplitude signals were not identifiable in the spectrograms at the receivers. Using four coplanar arrays [Fig. 8(a)] enables us to measure the noise correlation function with respect to the travel time separation between one receiver in array 1 and all receivers in arrays 2–4 as shown in Figs. 8(b)–8(d). Note in Figs. 8(a) and 8(d) that array 4 has twice as many elements as arrays 1–3. We observe from the correlation lag times that we have extracted wave fronts as they would propagate from a point source to ranges of 1700, 2400, and 3500 m, respectively. We also show that we recover similar wave fronts for the opposite direction by correlating one receiver in array 4 with all receivers in arrays 1–3 [Figs. 8(e)–8(g)]. The sloping environment results in the asymmetry between the two directions, i.e., upslope increases the reflection angle.

From the physical picture and the measurements presented above we can derive insight into the major components governing the rate of emergence of the coherent wave fronts for a homogeneous distribution of random sources. However, we can only draw some qualitative insight into our specific data of opportunity that is dominated by episodic shipping. Consider two receivers separated by a distance R. The signal-to-noise ratio (SNR) at the output of the correlation depends on three physical phenomena. First, it depends on the part of the signal that contributes to the correlation compared to the uncorrelated ambient noise, either acoustic or electronic. The uncorrelated acoustic ambient noise corresponds to the field that does not reach the two receivers because of the medium attenuation or above critical grazing angle transmission into the bottom. That is, the noise field consists of a local and some nonlocal components which vary with overall attenuation. Second, the correlation function is built from the contribution of noise sources located in the end-fire beams versus noise sources in the non-end-fire beams [cf. Eq. (4)]. Last, for a bandwidth Δf, the SNR associated with a correlation process grows with recording time T and is given by $\sqrt{T\Delta f}$.[18,22] In any event, the total SNR at the end of the correlation process is related to the

time bandwidth product, the spatial structure of the correlator as described in Eq. (4), and an environmental factor (expressions for local versus long range contributions given in Refs. 23 and 24), the latter typically being not known for an arbitrary location. Without specific environmental knowledge, the rate of emergence TDGF is a measured parameter that can be used to estimate the bottom geophysics (since volume attenuation is typically known). These arguments are for homogeneous ambient noise; the additional complication of shipping noise requires a long enough measurement period that demonstrates convergence.

VI. CONCLUSION

Our results demonstrate the potential information content of a random noise field. In particular, we have demonstrated through theory and data analysis that we can recover coherent deterministic wave fronts related to the structure of the time domain Green's function using measurements of ocean ambient noise between vertical arrays. In the NPAL data analysis, we have used the 20 min data blocks that were available and it is not likely that we have done sufficient time averaging for the optimal result for a shipping dominated environment. Since shipping noise is dominant at lower frequencies ($<$1000 Hz), we expect that high frequency will yield the most complete, uniformly converging, two-sided wave fronts predicted by theory. Further experiments in this higher frequency regime are expected to provide insights into the convergence time of this process. The results presented here are a first step toward passive tomographic imaging.

ACKNOWLEDGMENTS

This research was supported by the Office of Naval Research. The authors would like to thank Erin Oleson for her help in acquiring sonobuoy data during a Scripps experiment conducted by John Hildebrand's group and W. S. Hodgkiss, H. C. Song, and K. Sabra for valuable discussions.

[1] M. J. Buckingham, B. V. Berkhouse, and S. A. L. Glegg, "Passive imaging of targets with ambient noise," Nature (London) **365**, 327–329 (1992).
[2] G. M. Wenz, "Acoustic ambient noise in the ocean: Spectra and sources," J. Acoust. Soc. Am. **34**, 1936–1956 (1962).
[3] R. K. Andrew, B. M. Howe, J. M. Mercer, and M. Dzieciuch, "Ocean ambient sound: Comparing the 1960s with the 1990s for a receiver off the California coast," Acoustic Res. Letters Online, **3**, 65–70 (2002)]. In this paper, Wenz's results have been compared against more recent data.
[4] R. J. Urick, *Ambient Noise in the Sea* (Peninsula, Los Altos, 1986).
[5] D. Ross, *Mechanics of Underwater Noise* (Peninsula, Los Altos, 1987). Estimates of shipping densities given extrapolate to a distribution with more than 1 ship per square degree which suggests (and has been confirmed for many years in the field of Underwater Acoustics) that except for nearby specific ship tracks, distant shipping can be considered to be smeared out over the large surface of the ocean, albeit with a directional dependence.
[6] W. A. Kuperman and F. Ingenito, "Spatial correlation of surface generated noise in a stratified ocean," J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[7] C. H. Harrison and D. G. Simons, "Geoacoustic inversion of ambient noise: A simple method," J. Acoust. Soc. Am. **112**, 1377–1389 (2002).
[8] C. L. Epifanio, J. R. Potter, G. B. Deane, M. L. Readhead, and M. J. Buckingham, "Imaging in the ocean with ambient noise: The ORB experiments," J. Acoust. Soc. Am. **106**, 3211–3225 (1999).
[9] W. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* Cambridge (Cambridge University Press, Cambridge, 1995).
[10] N. C. Makris, F. Ingenito, and W. A. Kuperman, "Detection of a submerged object insonified by surface noise in an ocean waveguide," J. Acoust. Soc. Am. **96**, 1703–1724 (1994).
[11] J. Rickett and J. Claerbout, "Acoustic daylight imaging via spectral factorization: Helioseismology and reservoir monitoring," The Leading Edge **18**, 957–960 (1999).
[12] R. L. Weaver and O. J. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[13] 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).
[14] M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[15] R. Hennino, N. Tregoures, N. M. Shapiro, L. Margerin, M. Campillo, B. A. van Tiggelen, and R. L. Weaver, "Observation of equipartition of seismic waves," Phys. Rev. Lett. **86**, 3447 (2001).
[16] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer, New York, 2000).
[17] L. Brekhovskikh, *Waves In Layered Media* (Academic, New York, 1980). Of course, the wave fronts also emerge from a ray picture, but the ensuing frequency and bandwidth effect then become an additional complication to compute the complete path structure.
[18] 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).
[19] H. Schmidt and F. B. Jensen, "A full wave solution for propagation in multi-layered visco-elastic media with application to Gaussian beam reflection at fluid-solid interfaces," J. Acoust. Soc. Am. **77**, 813–825 (1985). See also Ref. 16, pp. 211–263.
[20] 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). This theoretical paper actually averages over depth for an inserted source and in a sense yields results closer to the Weaver equipartition results.
[21] 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 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 [ATOC Consortium, "Ocean climate change: comparison of acoustic tomography, satellite altimetry and modeling," Science **281**, 1327–1332 (1998)].
[22] The data occasionally showed distinct correlation building up over 20 s intervals. A typical ship might have a source level of 165 dB and at 100 Hz, a transmission loss of order 80 dB (see Ref. 5, p. 12). The ambient noise is about 90 db so that the SNR ~ -5 dB. With a time-bandwidth product $\sqrt{50\,(\mathrm{Hz}) \times 20\,(\mathrm{s})} \rightarrow 30$ dB, we definitely should see occasional distinct correlation patterns compressed by the ship bearing projection onto the line between the two receivers.
[23] H. Schmidt and W. A. Kuperman, "Estimation of surface noise source level from low-frequency seismo-acoustic ambient noise measurements," J. Acoust. Soc. Am. **84**, 2153–2162 (1988).
[24] J. S. Perkins, W. A. Kuperman, F. Ingenito, L. T. Fialkowski, and J. Glattetre, "Modeling ambient noise in three-dimensional ocean environments," J. Acoust. Soc. Am. **93**, 739–752 (1993).

Long-range propagation of finite-amplitude acoustic waves in an ocean waveguide

Kaëlig Castor,[a] Peter Gerstoft, Philippe Roux, and W. A. Kuperman
Marine Physical Laboratory, University of California San Diego, La Jolla, California 92093-0238

B. E. McDonald[b]
Naval Research Laboratory, Acoustics Division, Washington D.C. 20375

(Received 3 November 2003; revised 5 April 2004; accepted 5 April 2004)

A hybrid method coupling nonlinear and linear propagation codes is used to study the nonlinear signature of long-range acoustic propagation for high-amplitude sources in an ocean waveguide. The differences between linear and nonlinear propagation are investigated in deep and shallow water environments. The spectral reshaping that occurs in nonlinear propagation induces two main effects: in shallow water, an unusual arrival time structure in the lowest order modes is observed, and in both shallow and deep water environments, there is a tendency to have acoustic energy more uniformly distributed across modes. Further, parametric low-frequency generation in deep water is a candidate for the coupling between water and sediments for T-wave formation. [DOI: 10.1121/1.1756613]

PACS numbers: 43.30.Lz, 43.30.Bp, 43.30.Qd, 43.25.Cb. [AIT] Pages: 2004–2010

I. INTRODUCTION

Our goal in this paper is to study how to characterize acoustic signals generated by a high intensity source and propagated over several thousand kilometers in an ocean waveguide.[1,2] In a high-energy event, such as an explosive shock wave, nonlinear processes induce changes in the ocean's acoustic properties. The nonlinear signature of an acoustic waveform propagated over long ranges is investigated in this paper.

Several approaches have been put forward to model explosion waveforms and to estimate source levels.[3–6] In this article we do not deal with the phenomena that occur directly at the source. The simulations presented here use a narrow-band source in order to focus on signal evolution during nonlinear propagation over very long ranges. Two propagation models are used here. First, a time–domain code based on the Nonlinear Progressive Wave Equation (NPE)[7] is used to propagate a finite-amplitude acoustic wave field. Second, this code is coupled to the linear KRAKEN[8] normal mode code to propagate this field over very long oceanic paths beyond a certain distance for which the amplitude of the acoustic field is sufficiently low.

The results in this paper show that an initially high-amplitude acoustic wave retains a nonlinear signature, characteristic of its nonlinear origin, after becoming linear and propagating over great distances. One of the significant transformations that occurs during nonlinear propagation and still remains at very long ranges is in the waveform spectrum. The spectrum modification is associated with a frequency–mode coupling that is an energy redistribution upon both modes and frequencies during propagation. This frequency–mode coupling leads to a particular shape in the modal structure in terms of arrival time and energy distribution. Two types of long-range paths are analyzed in deep water that correspond respectively to a source located in shallow or deep water. Other results to be demonstrated are the following.

(1) Nonlinear effects are smaller in deep water than in shallow water due to the higher geometrical spreading that strongly diminishes the amplitude.
(2) In deep water, a nonlinear signature is characterized by energy more uniformly spread out over a large number of modes. In shallow water, this redistribution can be clearly observed, but fewer modes are excited.
(3) Nonlinearity in shallow water can be high enough to strongly alter the modal dispersion and thus change the arrival time of the modes.
(4) Nonlinear frequency–mode coupling can be responsible for an acoustic coupling from the earth to the SOFAR channel, which for long-range propagation is commonly designated as a T wave. Indeed, low parametric frequency excitation in deep water, that can be restricted in shallow water because of the waveguide cutoff frequency, gives rise to lower-order mode excitation. These lower modes are spread over the water column, and therefore a strong interaction with the bottom can contribute to an increased sea/bottom coupling.

In Sec. II, the NPE model used in the simulations is presented. In Sec. III we will present, for a narrow-band source, the nonlinear spectrum reshaping during nonlinear propagation, and then its effect on modal dispersion, by pointing out the differences due to the environment. Finally, the low-order mode excitation due to the low parametric frequency generated in deep water will be examined because of its possible influence on T-wave formation.

[a]Electronic mail: kaelig.castor@cea.fr
[b]Electronic mail: mcdonald@sonar.nrl.navy.mil

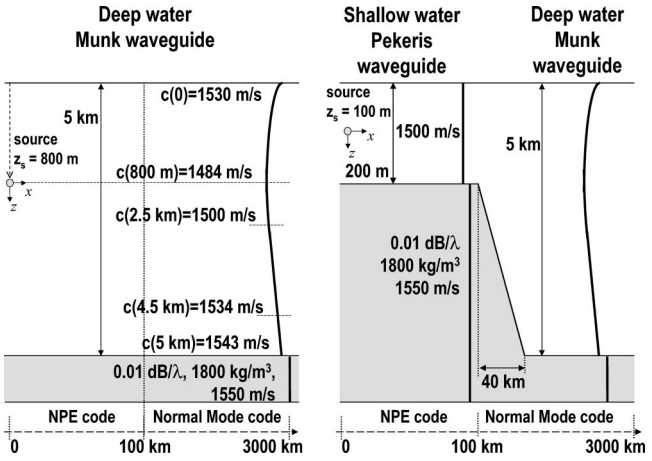


FIG. 1. Schematic diagrams of the environments used in this work.

II. EQUATIONS OF NONLINEAR ACOUSTICS AND WEAK SHOCK PROPAGATION

The NPE was developed by McDonald and Kuperman[7,9,10] to compute a time–domain solution for the acoustic field in a waveguide, including dominant nonlinear effects. The model is derived from the Euler equations of fluid dynamics retaining lowest-order nonlinearity augmented by an adiabatic equation of state relating pressure and density. The NPE is cast in a wave-following coordinate system moving at a nominal average sound speed c_0 in the longitudinal propagation direction x. The NPE describes nonlinear propagation of compression waves in the time–domain and is expressed as

$$\frac{\partial p}{\partial t} = -\frac{c_0}{2}\int_{\infty}^{x}\nabla_{\perp}^2 p\, dx - \frac{\partial}{\partial x}\left(c_1 p + \frac{\beta}{2\rho_0 c_0}p^2\right) + \frac{\delta}{2}\frac{\partial^2 p}{\partial x^2}, \tag{1}$$

where p is the acoustic pressure, and subscript 0 denotes ambient values. The constants δ and β are thermoviscous absorption and nonlinearity parameters ($\beta = 3.5$ for water), c_1 is the environmental sound speed fluctuation about c_0. One can reformulate a similar NPE[7,11] in terms of a dimensionless overdensity variable $R = p/(\rho_0 c_0^2)$, and the initial source amplitude is referred as the maximum overdensity R_m in the following. The terms on the right-hand side of Eq. (1) represent from left to right, diffraction, refraction, nonlinear steepening, and thermoviscous dissipation. Since the attenuation of low-frequency sound (<1 kHz) in seawater is very small, the thermoviscous dissipation term in the NPE is neglected here. However, a porous medium attenuation,[12] which is approximately linear in frequency, is included in the sediment layer adding at the right-hand side of Eq. (1), a quasi-Cauchy integral expressed as

$$0.02\alpha\frac{c_0}{2}\int_0^{\infty}\frac{\partial_x p(x+x')}{x'+\Delta x}dx', \tag{2}$$

where Δx and α are, respectively, the x grid spacing and attenuation in dB/wavelength (Fig. 1).

Recently, the well-known KZK[13,14] equation has been expressed in a form similar to the NPE. Thus, to simulate sonic boom propagation through a turbulent atmosphere, Blanc-Benon *et al.*[15] developed a modified KZK expressed as

$$\frac{\partial p}{\partial x} = \frac{c_0}{2}\int_{-\infty}^{t}\nabla_{\perp}^2 p\, dt + \frac{1}{c_0^2}\frac{\partial}{\partial t}\left(c_1 p + \frac{\beta}{2\rho_0 c_0}p^2\right) + \frac{\delta}{2c_0^3}\frac{\partial^2 p}{\partial t^2}. \tag{3}$$

This KZK equation includes the refractive term that accounts for the environmental variation of the sound speed. The NPE, Eq. (1), and KZK, Eq. (3), formulations are similar with the roles of time and distance reversed ($x_{\mathrm{NPE}} = -c_0 t_{\mathrm{KZK}}; t_{\mathrm{NPE}} = t_{\mathrm{KZK}} + x_{\mathrm{KZK}}/c_0$).[16] The differences between both approaches are only in the numerical algorithms. The KZK uses classical finite difference schemes for all the calculations while the NPE uses an accurate scheme to calculate the refraction and nonlinear steepening terms: a second-order upwind flux corrected transport scheme[17] that accounts properly for shock dissipation automatically avoiding Gibbs' oscillations without the necessity of adding an artificial attenuation like in a classical finite difference scheme.

III. LONG-RANGE NONLINEAR PROPAGATION OF A NARROWBAND SOURCE

A. Nonlinear effects on spectral evolution

The quadratic nonlinearity in Eq. (1) implies that the nonlinear contribution to the local sound speed is $\beta p/(\rho_0 c_0)$. In nonlinear propagation, the acoustic energy losses are increased for high initial amplitudes because of shock dissipation. When the shock wave discontinuity begins, a cascade of higher frequencies is generated. This phenomenon increases entropy locally and constitutes a mechanism of energy dissipation, even in a perfect fluid. The shock wave formation distance decreases as the source level increases.[18] Thus, for a strong explosion, dissipation at the shock front leads to a high rate of energy decay and amplitude saturation of the signal propagated in the waveguide.

The nonlinear propagation of an acoustic wave induces a spectral reshaping due to harmonic generation. Nonlinear propagation of two primary monochromatic waves at frequencies f_1 and f_2 gives rise to induced secondary radiation at frequencies $f_1 \pm f_2$. These secondary waves alter the spectrum of the acoustic field. The sum and difference frequencies are generated by parametric interaction, which has been studied for a long time,[19] especially in tomographic applications,[20,21] and more recently to measure the nonlinear parameter β in liquids.[22] Generally, radiation of sum- and difference-frequency sound from the nonlinear interaction region formed by the intersection of nonplanar modes in the waveguide is referred to as scattering of sound by sound.[18]

The NPE code is initialized by a sine-wave packet, modulated by Gaussian envelopes having scale sizes 600 m in range and 150 m in depth, and centered at the frequency $f = 30$ Hz. Two propagation environments are studied here (Fig. 1). In the first case [Fig. 1(a)], the source is located in shallow water, at a source depth $z_s = 100$ m. The second case

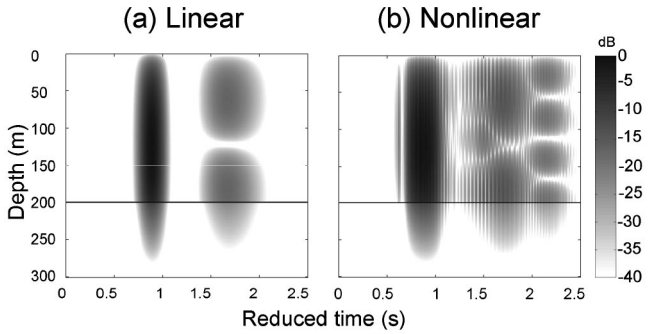


FIG. 2. Time series envelopes in shallow water at 100-km range, for a narrowband source centered at $f_c = 30$ Hz, at 100 m depth, with a maximum overdensity $R_m = 1 \times 10^{-5}$ (a. linear case) and $R_m = 5 \times 10^{-3}$ (b. nonlinear case).

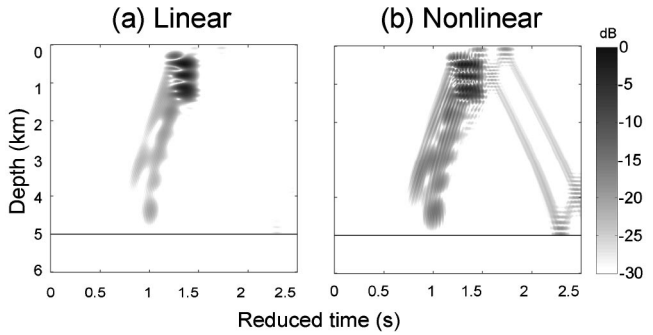


FIG. 3. Time series envelopes in deep water at 100-km range, for a narrowband source centered at $f_c = 30$ Hz, at 800 m depth, with a maximum overdensity $R_m = 1 \times 10^{-5}$ (a. linear case) and $R_m = 5 \times 10^{-3}$ (b. nonlinear case).

studied [Fig. 1(b)] is when the source is placed in deep water, in the middle of the SOFAR channel ($z_s = 800$ m). In both cases, the nonlinear propagation using the NPE code is carried to the 100-km range. Figure 2 and Fig. 3 give the time series envelopes at the 100-km range, resulting from the stated initial conditions. Figure 4 displays the depth-averaged spectrum of the source and at the 100-km range in shallow and deep water. After propagation over the 100-km range, the higher frequencies are damped out and only the first harmonic remains important in both cases. In deep water, the spectrum displays a low-frequency peak near 2–5 Hz due to nonlinear effects that lead to an energy transfer toward all the difference frequency components during propagation. So, this parametric difference frequency wave (DFW) at low frequency is directly related to the source frequency band-

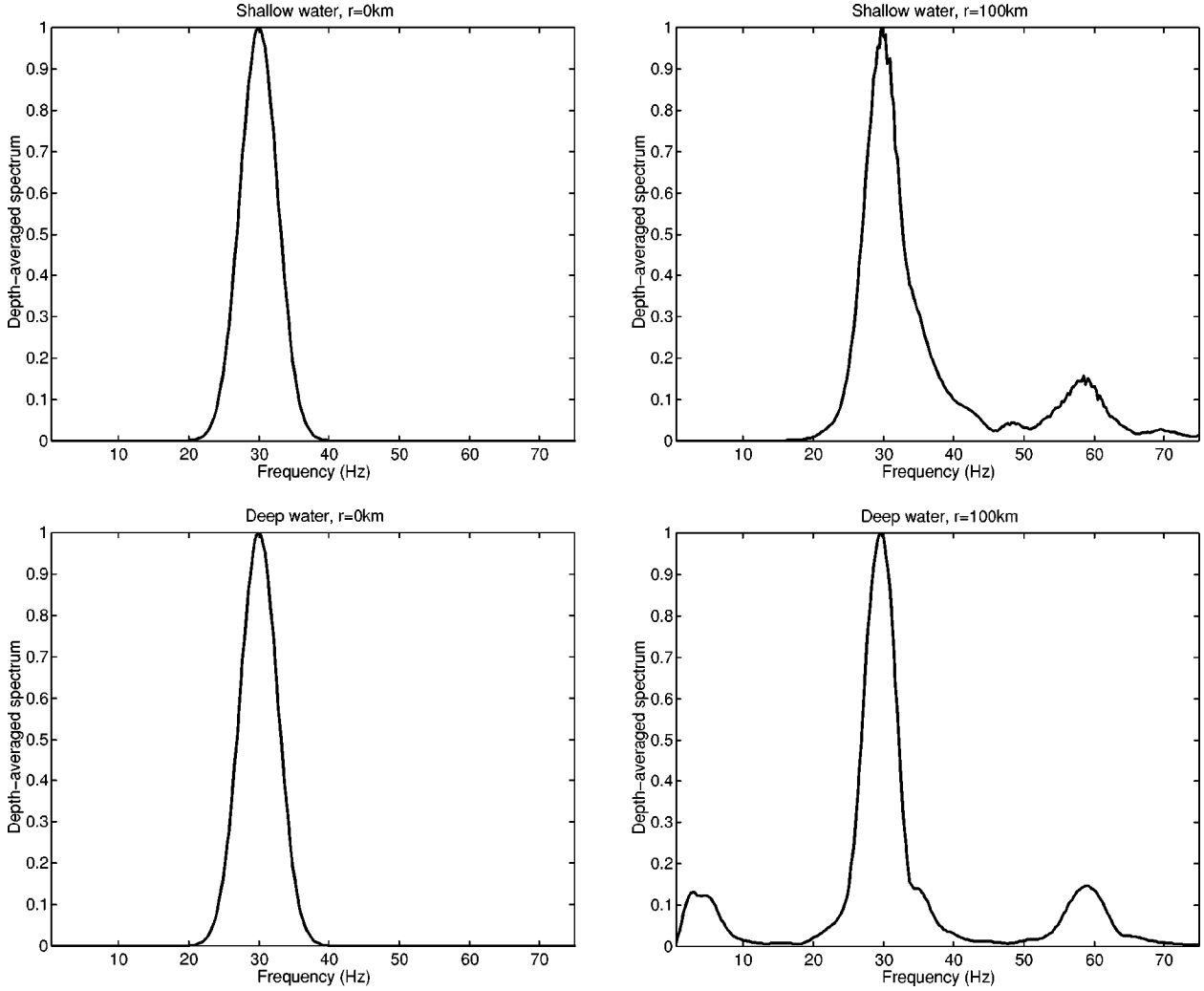


FIG. 4. Normalized depth-averaged spectrum in shallow and deep water for a narrowband source centered at $f_c = 30$ Hz, with a maximum overdensity $R_m = 5 \times 10^{-3}$, at 0 and 100-km ranges.

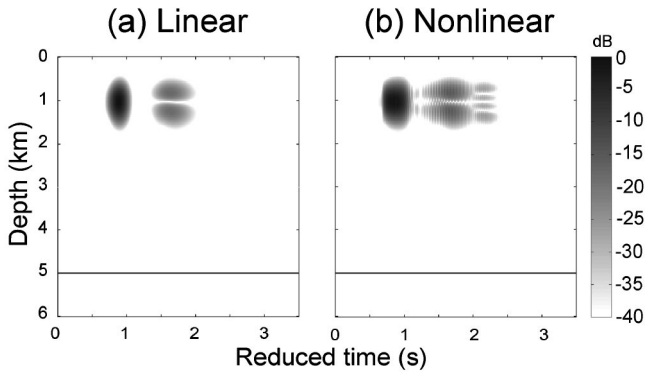


FIG. 5. Time series envelopes in deep water at 3000-km range, for a narrowband source centered at $f_c = 30$ Hz, at 100 m depth in shallow water, with a maximum overdensity $R_m = 1 \times 10^{-5}$ (a. linear case) and $R_m = 5 \times 10^{-3}$ (b. nonlinear case).

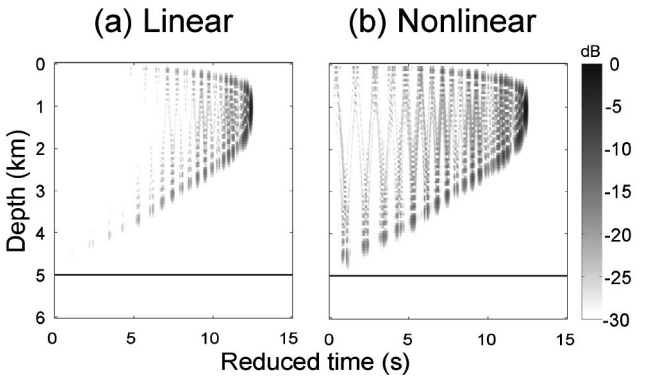


FIG. 6. Time series envelopes in deep water at 3000-km range, for a narrowband source centered at $f_c = 30$ Hz, at 800 m depth in deep water, with a maximum overdensity $R_m = 1 \times 10^{-5}$ (a. linear case) and $R_m = 5 \times 10^{-3}$ (b. nonlinear case).

width. A narrow frequency band source creates a narrow low-frequency peak, whereas a shorter duration source signal would generate a broader low-frequency spectrum. Figure 4(b) shows that this parametric low-frequency peak is truncated in shallow water, since it is below the waveguide cutoff frequency.

For finite-amplitude sound waves, nonlinear effects develop at short ranges. One of our aims in this study is to determine some characteristics at very long ranges of such waves that originate from nonlinear wave propagation. After an initial distance in which the nonlinear wave loses energy to shock processes and increased bottom penetration,[11] its amplitude is sufficiently low so that its interaction with the waveguide becomes essentially linear. Then, a linear normal mode code can be used to study the modal decomposition of the acoustic field and also propagate it to much longer ranges (i.e., several thousand km). Even though the presence of nonlinearity does not lend itself to a straightforward representation in linear normal modes, similarities between the two cases are expected since the nonlinearities are weak.

At the 100-km range, the field is spectrally decomposed into its modal components before the linear code KRAKEN can carry out the propagation to longer ranges frequency by frequency and mode by mode.[8,23] These modal components at very long ranges are finally synthesized into signal arrivals in the time domain. For the range-dependent environment [Fig. 1(b)], the adiabatic approximation is used in the linear normal modes propagation. Figure 5 represents the time series envelopes, for linear and nonlinear propagation paths in this case, at a 3000-km range, after propagation in deep water. Figure 6 represents the time series envelopes for linear and nonlinear cases, at a 3000-km range, when the source is placed in deep water at a source depth $z_s = 800$ m.

B. Influence of the environment on dispersion of modes

Figures 7(a) and 7(b) represent the group speed versus phase speed, at the center frequency $f = 30$ Hz, respectively for the shallow and deep water waveguides (Fig. 1). The lower-order modes correspond to the lower grazing angles and thus the lower phase speeds. Group speeds correspond to time of arrivals for a given range. In shallow water, the lower-order modes travel faster [Fig. 7(a)]. In deep water [Fig. 7(b)], the dispersion of modes is mainly due to refraction: the group-speed order of the lower-order modes is flipped in comparison to the shallow water case [Fig. 7(a)].

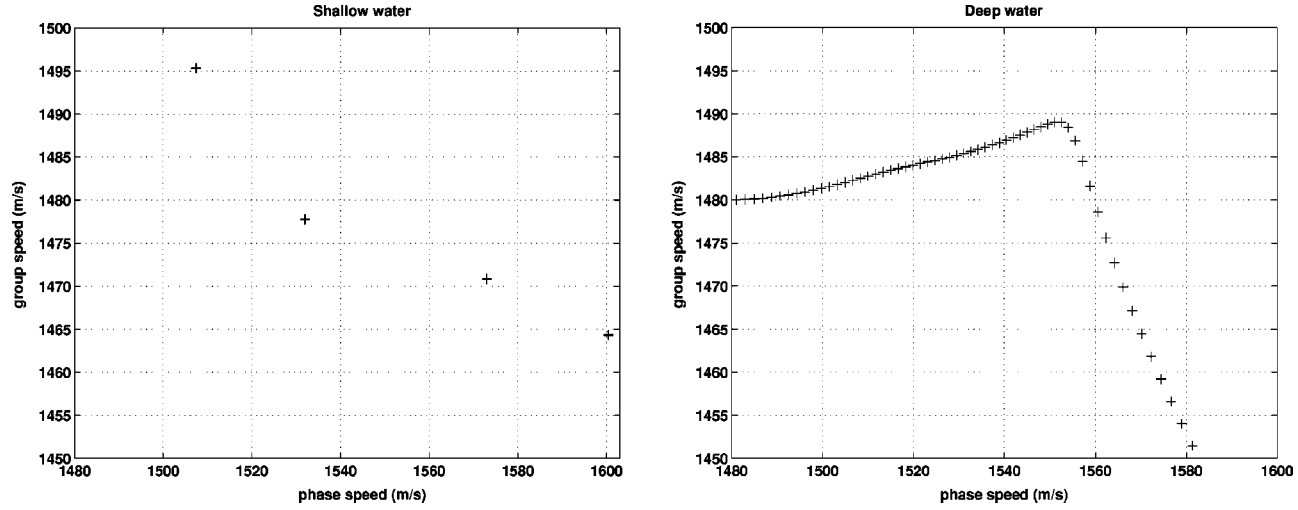


FIG. 7. Group speed vs phase speed at the central source frequency $f = 30$ Hz for shallow and deep water.

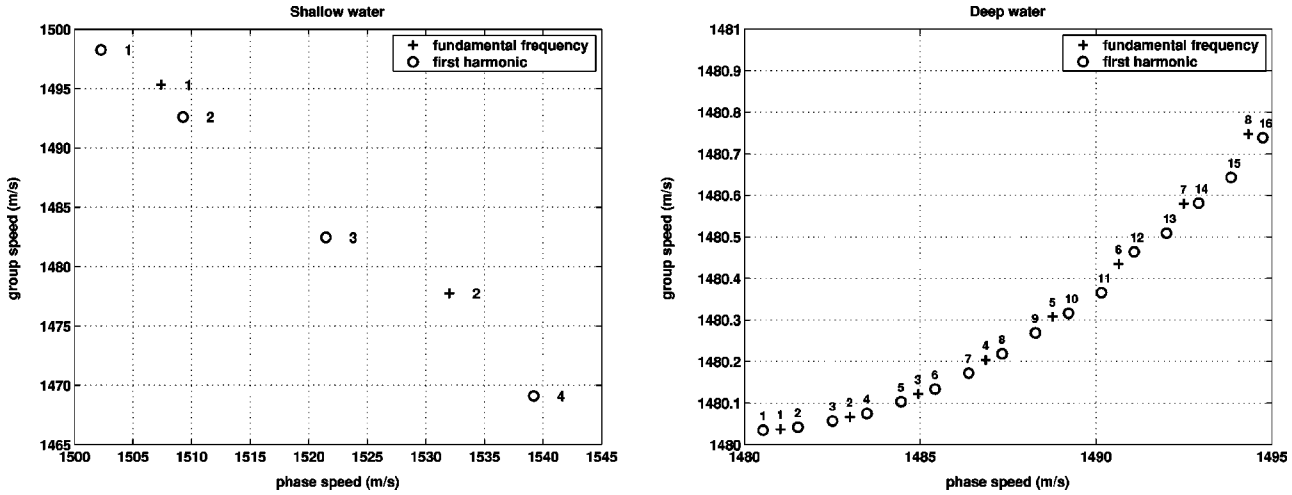


FIG. 8. Group speed vs phase speed, for the central source frequency $f=30$ Hz and the first harmonic $f=60$ Hz, for shallow and deep water. The numbers display the indexes of the lower order modes that are excited.

So, the higher-order modes travel faster in deep water, which is clearly visible on a greatly dispersed signal (Fig. 6 at a 3000-km range). Figure 5 shows that when the signal is propagated to deep water after the 100-km range, the lower-order modes still arrive first at the 3000-km range because only a few modes are excited in shallow water, and the group speeds of these lowest-order modes are almost the same in deep water [Fig. 7(b)]. Indeed, the difference between the group speed for the first mode and the third mode is approximately 25 m/s in shallow water [Fig. 7(a)] and about 0.08 m/s in deep water [Fig. 7(b)]. In shallow water the modes are expected to be separated in time faster than in deep water. Then, for the shallow-to-deep water case, the dispersion of modes mainly occurs in shallow water and the signal keeps approximately the same shape along several hundred kilometers when it propagates in the deep water SOFAR channel (Fig. 5). Note that, at long ranges, even if the time separation of modes is larger in shallow water, the total time spread of the signals will be much larger in deep water than in shallow water since more modes are excited. In shallow water, due to a strong interaction with the bottom, the higher-order modes are attenuated and only few modes survive.

C. Nonlinear effects on modal distribution

Finite-amplitude sound waves interact differently with the ocean/bottom interface than linear waves because of the nonlinear contribution to the local sound speed. The dispersion of a nonlinear pulse is therefore expected to be different from that of a linear pulse. Two differences can be observed between linear and nonlinear propagation: the first one is related to the energy redistributed among modes and the second concerns the modal arrival time. In both environments, the energy initially carried by the source frequency components is redistributed to the other frequencies created by nonlinear effects.

Figures 2 and 5 show that, for a shallow water source, the nonlinear signature results in low-order mode attenuation due to shock formation: the nonlinearities lead to more relative energy for higher-order modes, which are left after the initial mode stripping due to the bottom interaction. When the nonlinearities are strong enough, there is a tendency toward energy equipartition at long ranges leading to an arrival time structure in lower-order modes characterized by an almost uniform weight for each mode. This modal equipartition results from mode coupling at short ranges, where the nonlinearities are sufficiently high and the modes are not dispersed. Figure 8(a) shows, for the modes that are mainly excited (Figs. 2 and 5), the group speed versus phase speed at the fundamental frequency and at the first harmonic. In shallow water, the individual modes at the first harmonic travel faster than the ones at the fundamental. Note that the modes associated with the nonlinear frequency components are excited during propagation. It is expected that an optimum in nonlinear mode excitation is reached at the shock wave formation distance. The phase lag then induced could slightly influence the arrival time of the nonlinear modes by delaying the time separation of modes. But the difference in group speeds between the fundamental and first harmonic, as explained before [Fig. 8(a)], is primarily responsible for the difference between linear and nonlinear arrival time structure. Figure 9 shows the mode amplitudes versus frequency for the linear and nonlinear cases in shallow water at 100-km range showing that the first harmonic leads to an excitation of modes 3 and 4. According to the previous comments, the

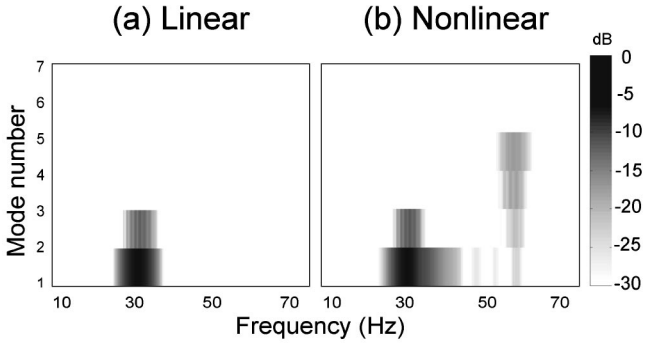


FIG. 9. Mode amplitudes (dB) versus mode-number and frequency, in shallow water at 100 km for a narrowband source centered at $f_c=30$ Hz, with a maximum overdensity $R_m=1\times10^{-5}$ (a. linear case) and $R_m=5\times10^{-3}$ (b. nonlinear case).

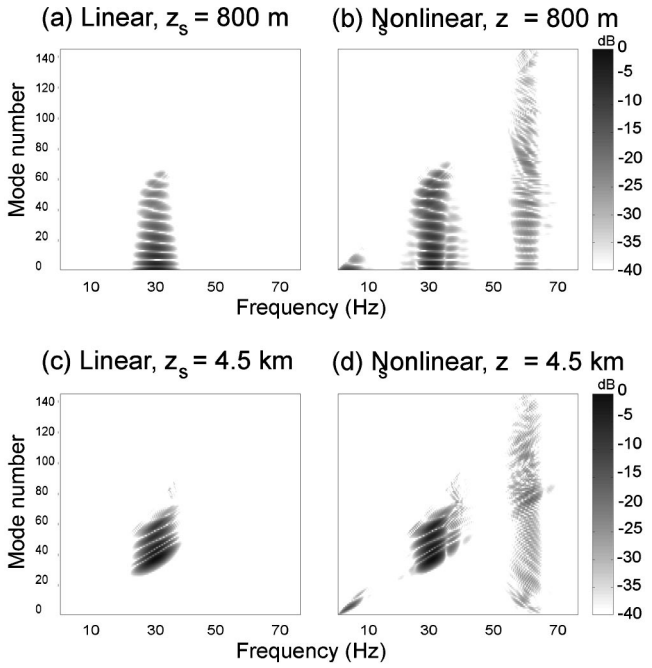


FIG. 10. Mode amplitudes (dB) versus mode-number and frequency, in deep water, for linear (left) and nonlinear (right) for two different source depths [($z_s = 800$ (top) and $z_s = 4.5$ km (bottom)].

nonlinear modes (modes 3 and 4 at 60 Hz) arrive just after the first linear modes (modes 1 and 2 at 30 Hz), respectively (Fig. 2). This example illustrates what we mean by frequency–mode coupling.

D. Influence of parametric mode conversion on sea/bottom coupling

A recent study[24] has reported that sound speed fluctuations due to internal waves are a dominant source of mode coupling in long-range propagation scenarios. Internal-wave-induced scattering eventually results in an equipartition of energy among the lower-order modes in deep water. Also, a broadening of the signals[25] is attributable to the exchange of energy among the modes, caused by the internal waves. Acoustic normal mode propagation is strongly nonadiabatic due to internal waves.[26] The results presented previously (Sec. II C) show that, due to a mode coupling induced by a frequency redistribution of the acoustic energy, the impact of nonlinearities also leads to a result that displays a tendency toward an equipartition and a mixed arrival time structure of the modes. In this section, even if the results presented are still limited to ocean propagation and sources in the ocean, we point out a particular effect of frequency–mode coupling that is an increased sea/bottom coupling.

Figure 10 represents the mode amplitudes versus frequency at a 100-km range for the 5-km deep water waveguide for two source depths ($z_s = 800$ m and $z_s = 4.5$ km) and for linear and nonlinear cases. The number of propagating modes increases with frequency. When the source is at the SOFAR axis ($z_s = 800$ m), all the modes are excited. Whereas if the source is near the bottom ($z_s = 4.5$ km), the SOFAR trapped modes are not excited. Then, in the linear case, only the highest-order modes are excited and propagate over long ranges (Fig. 10). Figure 10 shows that, in a nonlinear case, mode coupling at short ranges leads to an excitation of the lowest-order modes due to parametric low-frequency generation. Even when the source is close to the bottom (Fig. 10 for $z_s = 4.5$ km), the nonlinearities induce low-order mode excitation that contributes to increased bottom penetration.

We have shown that parametric low-frequency generation also increases the sea/bottom coupling. For a wave originating from the earth, we expect to obtain sufficiently high acoustic amplitudes in the water, and consequently, find similar qualitative results to the ones presented here based on a source in the water above the bottom: nonlinearities would redistribute energy toward lower frequencies, and then, more penetrating modes would be excited, leading to increased sea/bottom coupling. Consequently, such nonlinear effects may be involved in the processes of conversion of seismic waves to (and from) acoustic energy, which can result in the efficient excitation of T waves by seismic sources located inland or by underground explosions. From a terrestrial origin, T waves are acoustic energy trapped in the SOFAR channel that can propagate efficiently over a long distance without significant energy loss. Understanding the nature of the coupling between the underwater acoustic field and the land seismic field is important for evaluating the performance of the hydroacoustic stations attempting to detect nuclear explosions.[1,27] The seismic T-wave conversion process at the ocean bottom is not well understood. Two different mechanisms have been identified to explain the T-wave coupling.[28] The first one is ''slope conversion,'' which is controlled by seafloor slope.[29,30] The second mechanism is scattering of energy into the sound channel at bathymetric promontories in close proximity to the source region. This ''seafloor scattering'' is then dominated by the seafloor depth.[31,32] Figure 10 shows that the nonlinearities induce energy transfer toward low-order modes at the parametric difference frequency. These low-order modes excite the water column and interact with the bottom. Consequently, this phenomenon could be responsible for T-wave generation.

IV. CONCLUSION

Nonlinear characteristics of long-range acoustic paths in shallow and deep water environments have been demonstrated in this paper. The results illustrate that nonlinear properties modify the spectral evolution of the acoustic field and its modal distribution. A nonlinear birthmark may therefore be present in hydroacoustic signals recorded at long distance related to underwater explosions. This study should contribute to diagnosing the nonlinear origin of such signals.

In shallow water, the acoustic field is altered by the waveguide interfaces earlier than in deep water, inducing lower geometrical spreading. Hence, the nonlinear effects on acoustical propagation are greater if an explosion occurs in shallow water. In deep water, the lowest-order modes have roughly the same group speed. Therefore, the shallow-to-deep water time series still display an arrival time structure with the lower-order modes arriving first in contrast to the deep water case.

Nonlinearities can lead to similar results to those obtained for long-range ocean acoustic wave propagation in random media: a tendency to have an equipartition of lower-order modes. Finally, the nonlinear process of parametric conversion of bottom interacting modes to lower-order SOFAR trapped modes suggests a potential mechanism for T-wave generation. In summary, after long-range propagation, it may be possible to characterize a strong explosion by studying the remaining low-frequency waves generated by parametric interaction.

ACKNOWLEDGMENTS

This research was sponsored by Defense Threat Reduction Agency, Contract No. DTRA01-00-C-0084.

[1] P. Gerstoft, ''Assessment of hydroacoustic processing in the CTBT release one monitoring software,'' Comprehensive Nuclear-Test-Ban-Treaty Organisation, Vienna, Austria, 1999. http://www-mpl.ucsd.edu/people/gerstoft/

[2] G. L. D'Spain, W. A. Kuperman, J. Orcutt, and M. Hedlin, ''Long range localisation of impulsive sources in the atmosphere and ocean from focus regions in single element spectrograms,'' *Proceedings of the 22nd*, Annual Seismic Research Symposium, 2000.

[3] P. H. Rogers, ''Weak-shock solution for underwater explosive shock waves,'' J. Acoust. Soc. Am. **62**, 1412–1419 (1977).

[4] N. R. Chapman, ''Measurement of the waveform parameters of shallow explosive charges,'' J. Acoust. Soc. Am. **78**, 672–681 (1985).

[5] N. R. Chapman, ''Source levels of shallow explosive charges,'' J. Acoust. Soc. Am. **84**, 697–702 (1988).

[6] T. L. Geers and K. S. Hunter, ''An integrated wave-effects model for an underwater explosion bubble,'' J. Acoust. Soc. Am. **111**, 1584–1601 (2002).

[7] B. E. McDonald and W. A. Kuperman, ''Time-domain formulation for pulse propagation including nonlinear behavior at a caustic,'' J. Acoust. Soc. Am. **81**, 1406–1417 (1987).

[8] M. Porter, The KRAKEN normal mode program, SACLANTCEN SM-245, La Spezia, Italy, 1991.

[9] B. E. McDonald, ''High-angle formulation for the nonlinear progressive-wave equation model,'' Wave Motion **31**, 165–171 (2000).

[10] B. E. McDonald, ''Nonlinear effects in source localization,'' *Ocean Acoustic Interference Phenomena*, edited by W. A. Kuperman and G. L. D'Spain (American Institute of Physics, Melville, NY, 2002).

[11] J. J. Ambrosiano, D. Plante, B. E. McDonald, and W. A. Kuperman, ''Nonlinear propagation in an ocean waveguide,'' J. Acoust. Soc. Am. **87**, 1473–1481 (1990).

[12] J. E. White, *Seismic Waves* (McGraw-Hill, New York, 1965).

[13] X. Zabolotskaya and X. Khokhlov, ''Quasi-plane waves in the nonlinear acoustics of confined beams,'' Sov. Phys. Acoust. **15**, 35–40 (1969).

[14] V. P. Kuznetsov, ''Equations of nonlinear acoustics,'' Sov. Phys. Acoust. **16**, 467–470 (1971).

[15] P. Blanc-Benon, B. Lipkens, 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).

[16] K. Castor, B. E. McDonald, and W. A. Kuperman, ''Equations of nonlinear acoustics and weak shock propagation,'' *16th International Symposium on Nonlinear Acoustics*, Moscow, Russia, 2002.

[17] B. E. McDonald, ''High order upwind flux methods for scalar hyperbolic conservation laws,'' J. Comput. Phys. **56**, 448–460 (1984).

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

[19] P. J. Westervelt, ''Parametric acoustic array,'' J. Acoust. Soc. Am. **35**, 535–537 (1963).

[20] B. K. Novikov, O. V. Rudenko, and V. I. Timochenko, ''Nonlinear underwater acoustics,'' ASA, New York, 1987.

[21] J. Marchal, ''Acoustique non linéaire: contribution théorique et expérimentale à l'étude de l'émission paramétrique,'' Ph.D. dissertation, Université Pierre et Marie Curie-Paris 6, 2002.

[22] C. Barrière and D. Royer, ''Diffraction effects in the parametric interaction of acoustic waves: Application to measurements of the nonlinearity parameter B/A in liquids,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1706–1715 (2001).

[23] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Springer-Verlag, New York, NY, 1994).

[24] K. E. Wage, A. B. Baggeroer, and J. C. Preisig, ''Modal analysis of broadband acoustic receptions at 3515-km range in the North Pacific using short-time fourier techniques,'' J. Acoust. Soc. Am. **113**, 801–817 (2003).

[25] J. A. Colosi, S. M. Flatte, and C. Bracher, ''Internal-wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison with experiment,'' J. Acoust. Soc. Am. **96**, 452–468 (1994).

[26] J. A. Colosi and the ATOC group, ''A review of recent results on ocean acoustic wave propagation in random media: Basin scales,'' IEEE J. Ocean. Eng. **24**, 138–155 (1999).

[27] G. L. D'Spain, L. P. Berger, W. A. Kuperman, J. L. Stevens, and G. E. Baker, ''Normal mode composition of earthquake T phases,'' Pure Appl. Geophys. **158**, 475–512 (2001).

[28] C. H. Li, ''T-waves excited by S-waves and oscillated within the ocean above the southeastern taiwan forearc,'' Geophys. Res. Lett. **28**, 3297–3300 (2001).

[29] J. Talandier and E. A. Okal, ''On the mechanism of conversion of seismic waves to and from T-waves in the vicinity of island shores,'' Bull. Seism. Soc. Am. **88**, 621–632 (1998).

[30] P. F. Piserchia, J. Virieux, D. Rodrigues, S. Gaffet, and J. Talandier, ''Hybrid numerical modelling of T-wave propagation: application to the mid-plate experiment,'' Geophys. J. Int. **133**, 789–800 (1998).

[31] C. D. de Groot-Hedlin and J. A. Orcutt, ''Excitation of T-phases by seafloor scattering,'' J. Acoust. Soc. Am. **109**, 1944–1954 (2001).

[32] C. D. de Groot-Hedlin and J. A. Orcutt, ''T-phase observation in northern california: Acoustic to seismic coupling at a weakly elastic boundary,'' Pure Appl. Geophys. **158**, 513–530 (2001).

A broadband model of sandy ocean sediments: Biot–Stoll with contact squirt flow and shear drag

Nicholas P. Chotiros[a)]
Office of Naval Research, Code 321 OA, 800 North Quincy Street, Arlington Virginia 22217

Marcia J. Isakson
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029

(Received 22 September 2003; revised 23 April 2004; accepted 16 July 2004)

Unlike the application of the Biot model for fused glass beads, which was conclusively demonstrated by Berryman [Appl. Phys. Lett. **37**(4), 382–384 (1980)] using the experimental measurements by Plona [Appl. Phys. Lett. **36**, 259–261 (1980)], the model for unconsolidated water-saturated sand has been more elusive. The difficulty is in the grain to grain contact physics. Unlike the fused glass beads, the connection between the unconsolidated sand grains is not easily modeled. Measurements over a broad range of frequencies show that the sound speed dispersion is significantly greater than that predicted by the Biot–Stoll model with constant coefficients, and the observed sound attenuation does not seem to follow a consistent power law. The sound speed dispersion may be explainable in terms of the Biot plus squirt flow (BISQ) model of Dvorkin and Nur [Geophysics **58**(4), 524–533 (1993)]. By using a similar approach that includes grain contact squirt flow and viscous drag (BICSQS), the observed diverse behavior of the attenuation was successfully modeled. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1791715]

PACS numbers: 43.30.Ma, 43.30.Pc, 43.30.Vh [RAS] Pages: 2011–2022

I. INTRODUCTION

For acoustic purposes, the ocean sediment is often modeled as a fluid or an elastic solid. This allows the extensive knowledge and experience in computing elastic wave propagation to be directly applied. The acoustic properties of ocean sediments have been compiled in a number of landmark publications, such as the series of papers by Hamilton[1] in which empirical relationships were developed to connect sound wave speed and attenuation to sediment type and frequency. The data appeared to show that sound speed is approximately independent of frequency, but attenuation increases linearly with frequency. Both vary with sediment type. Methods for inverting acoustic measurements for sediment properties often rely on these relationships. While there has been some success with soft sediments, the results of such inversions in sandy sediments have been poor. The problem is the underlying model.

The inadequacy of the visco-elastic model, in which the sediment is modeled as a visco-elastic solid with approximately constant p- and s-wave speeds and wave attenuations that increase linearly with frequency, has been demonstrated. Model predictions of reflection loss at normal incidence are shown to be inconsistent with measured data.[2] It will be shown below that the visco-elastic model is unable to accommodate the recently measured sound speed dispersion. The poro-elastic model, as conceived by Biot[3] and reformulated by Stoll,[4] has been put forward as a possibly better alternative. The Biot–Stoll model represents the collection of sand grains as a porous elastic frame in which the pore spaces are filled with water. It is certainly an improvement over the homogeneous visco-elastic models but inconsistencies between model and data remained.

At any single frequency, the Biot–Stoll model can be adjusted to match the measured values of acoustic and shear wave speeds and attenuations.[5–7] When compared to the extant published experimental data in water-saturated sands, the Biot–Stoll model is unable to match the observed frequency dependence of sound speed and attenuation over a broad range of frequencies. In Sec. II, it will be shown that the model is unable to reproduce the measured sound speed dispersion in the two most comprehensive data sets available to date. It will also be shown that the frequency dependence of attenuation does not follow one consistent power law. It is against this backdrop that the following extension of the Biot–Stoll theory was developed. In Sec. III, the Biot model with grain contact squirt flow and shear drag (BICSQS) will be developed from simple physical considerations at the grain to grain contact. BICSQS provides a causal and physically sound model for the poro-elastic frame. In Sec. IV, the properties of the BICSQS model will be explored with representative theoretical examples, and matched with the small but growing broadband data base of measured sediment sound speed and attenuation. In the conclusions, the essential points of the BICSQS model are summarized and the remaining issues are described.

II. BROADBAND ACOUSTIC DATA

Acoustic attenuation in ocean sediments, measured as decibels per meter of distance traveled, appears to increase linearly with frequency, as depicted in the data presented by Hamilton[8] (H). Unfortunately, measurements of attenuation in water-saturated sands and sandy sediments are not always consistent with this model. In the interest of fairness to all

a)Electronic mail: nicholas_chotiros@onr.navy.mil

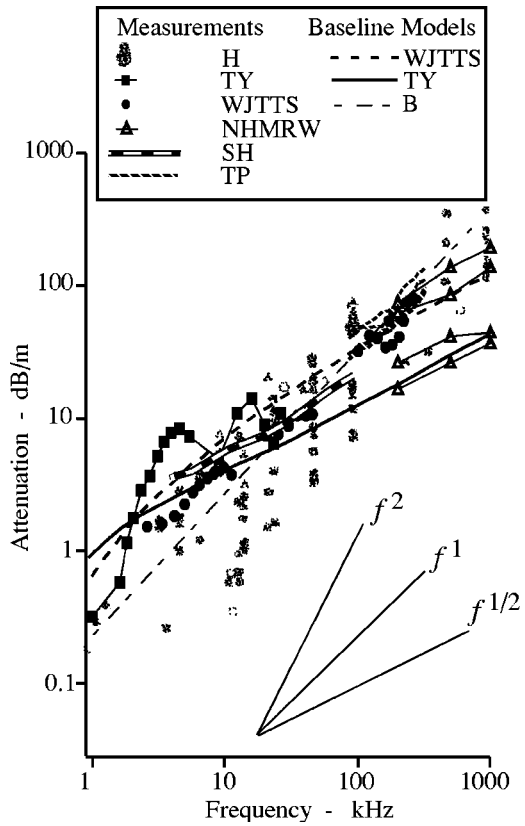


FIG. 1. Measured sound attenuation as a function of frequency from several sources and the baseline model predictions.

contributing authors and for compactness, published experimental results will be referred to by the first letters of the authors' names, as follows. Laboratory measurements by Nolle, Hoyer, Mifsud, Runyan, and Ward[9] (NHMRW), Thomas and Pace[10] (TP), and more recent data from Simpson and Houston[11] (SH), along with *in situ* measurements by Turgut and Yamamoto[12] (TY) and by several participants at the Sediment Acoustics Experiment of 1999 (SAX99) as reported by Williams, Jackson, Thorsos, Tang, and Schock[13] (WJTTS), are overlaid on the historical data from Hamilton in Fig. 1. The trends are confusing. The low frequency data below 5 kHz appear to follow a trend that is closer to f^2. The laboratory data of SH and NHMRW appear to follow a $f^{1/2}$ trend. The remainder appear to be linearly proportional to frequency. It would appear that no single power law could satisfactorily match all of the diverse trends.

The sound speed in the sediment is usually assumed to be independent of frequency, but measurements from TY and WJTTS show quite the opposite. The measurements are reproduced in Fig. 2. Although there are only two such sets of measurements in the published literature, they appear to corroborate each other. Both sets suggest that there may be low and high frequency asymptotic values, and a transition region in the frequency band roughly between 1 and 10 kHz.

It is clear that a simple fluid or elastic model will be incapable of explaining the sound speed dispersion. The recent model by Buckingham[14] (B), which predicts an attenuation that increases as the first power of frequency, also predicts an increase in sound speed with frequency, consistent with Kramers–Krönig. It has five parameters: bulk density, asymptotic low frequency sound speed, frame shear rigidity, frame bulk rigidity, and stress relaxation exponent. The bulk density and the asymptotic low frequency sound speed are measurable. The parameters interact to some extent, but, generally speaking, the frame shear rigidity term controls the shear speed, the average bulk compressive rigidity controls the sound speed, and the stress relaxation exponent controls the magnitude of the attenuation. A realization of this model, using parameter values in Table I, was compared with the data in Figs. 1 and 2. It underestimates the magnitude of the observed sound speed dispersion, and the linear frequency attenuation prediction is unable to match the diversity of frequency dependence observed in the measured data.

TABLE I. Parameter values of the Buckingham model.

Parameter	Units	
c_o, low frequency sound speed	m/s	1600
n, stress relaxation exponent	···	0.090 14
γ_p, compressional rigidity coefficient	GPa	0.248
γ_s, shear rigidity coefficient	GPa	0.001 52
ρ_o, bulk density of saturated medium	kg/m^3	2016

The baseline Biot–Stoll model is taken from Stoll and Kan.[15] Two realizations of the baseline model were constructed using model parameter values from the TY and WJTTS experiments. In the case of TY, the frame moduli were not given, but the shear wave speed and the frame Poisson's ratio were given. Thus, the frame shear modulus was adjusted to fit the measured shear speed of 110 m/s, and the frame bulk modulus was computed using the Poisson's ratio of 0.3. The log decrements were set at 0.15 as given in TY for typical cases. The values of permeability κ, porosity β, added mass coefficient c, and the densities (ρ_r, ρ_f) and bulk moduli (K_r, K_f) of the grain and fluid were given. The pore size parameter a was computed using the Kozeny–Carman equation,[16]

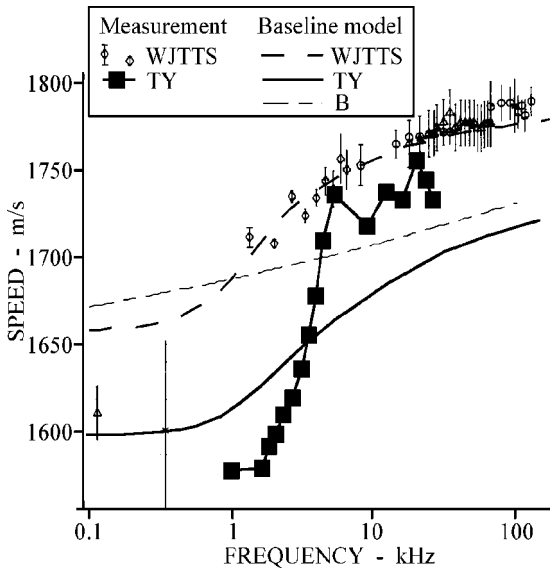


FIG. 2. Measured sound speed as a function of frequency from TY and WJTTS and baseline model predictions.

TABLE II. Baseline Biot–Stoll models from TY and WJTTS.

Parameter	Units	TY	WJTTS
Bulk properties			
β, porosity	⋯	0.44	0.385
ρ_r, grain density	kg/m^3	2650	2690
ρ_f, fluid density	kg/m^3	1000	1023
K_r, grain bulk modulus	GPa	36	32
K_f, fluid bulk modulus	GPa	2.3	2.395
Fluid motion			
η, fluid viscosity	kg/m-s	0.001	0.001
κ, permeability	μ^2	17.5	25
a, pore size	μ	52	26.5
c, added mass coefficient	⋯	1.25	1.35
Frame response			
μ_o, frame shear modulus	GPa	0.024	0.0292
δ_μ, shear log decrement	⋯	0.15	0.194
K_b, frame bulk modulus	GPa	0.052	0.0436
δ, bulk log decrement	⋯	0.15	0.15

$$\kappa = \frac{\beta a^2}{4k_k}, \tag{1}$$

where k_k is the Kozeny parameter, which has a value of 5 for spherical grains. In the case of WJTTS, all of the parameter values were given. The values are shown in Table II. The computed sound speed and attenuation are compared with the data in Figs. 1 and 2. The measured dispersion is clearly greater than the TY baseline Biot–Stoll model. The WJTTS baseline model appears to predict greater dispersion, and comes closer to matching the measured dispersion. However, the near-agreement between model and data was achieved by stretching certain parameter values beyond what is likely to be true. In particular, the value of porosity used, 0.385, was significantly higher than the average measured value[17] from diver cores of 0.366, which may be rounded up to 0.37. The value of the grain bulk modulus chosen was at the lower limit of the 95% confidence interval,[18] i.e., 32 GPa instead of the usual[19] 36 GPa for quartz. With respect to attenuation, both baseline models predict a frequency dependence closer to $f^{1/2}$ over the range of frequencies shown. The baseline model is clearly overwhelmed by the diversity in the experimental trends, which suggests that the underlying physics may be more complicated than the model permits.

III. THE BIOT–STOLL MODEL WITH CONTACT SQUIRT FLOW AND SHEAR DRAG (BICSQS)

The physics of the grain to grain contact is critical to understanding the dispersion and attenuation curves. The baseline model considers the frame to be a monolithic solid that has been hollowed out by interconnected pores, and the properties of the dry frame remain unchanged when fluid is introduced into the pore spaces. This is not the case with fluid-saturated unconsolidated particles because the fluid, through capillary and other short-range forces, alters the mechanical properties of the contact region. In the case of water and sand, the water tends to permeate the grain to grain contact region. Sand grains may be considered as particles with a level of small-scale surface roughness, as illustrated in Figs. 3(a) and (b). Two adjacent sand grains may make solid contact at a few discrete points. Otherwise, there is a layer of fluid between the surfaces. The geometry may be idealized as two flat surfaces separated by a film of fluid, and punctuated by a small solid connection, as in Fig. 3(c).

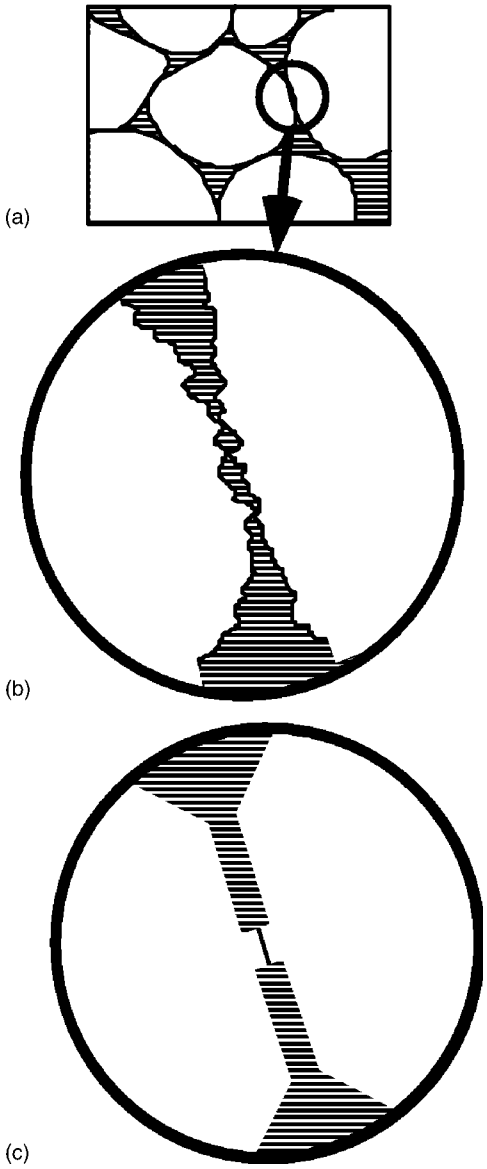


FIG. 3. Idealization of the grain-grain contact.

As the frame is compressed, the two surfaces are pushed closer together. The force between the surfaces is balanced by the stiffness of the solid connection, and by the reaction of the fluid film. With reference to Fig. 4(a), the response to a compressive force F is treated in terms of the change in gap width y, and a change in radial displacement of the gap fluid film r. The response to a shear force S is in terms of the deformation angle θ, as illustrated in Fig. 4(b).

The compressive response of the solid connection is modeled as a spring, and the response of the surrounding fluid as a spring and a dash-pot, representing the compressibility of the fluid and the drag associated with squirt flow in the gap, as shown in Fig. 4(c). There is an additional inertial term due to the mass of the water as it moves into and out of the gap, but it will be neglected because it is comparatively

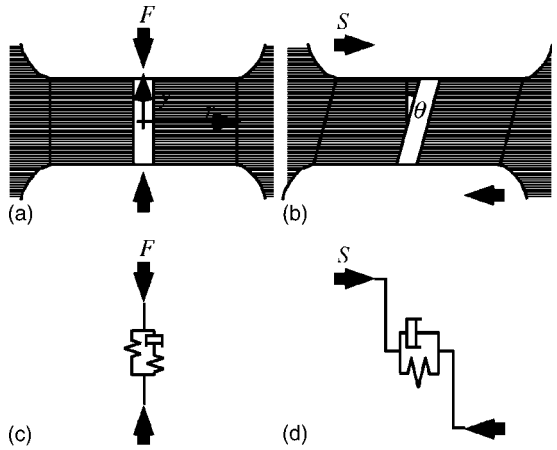


FIG. 4. Grain contact models of (a) compression, (b) shear, and their equivalents (c) and (d).

insignificant, as shown in the Appendix. The pore spaces surrounding the gap are assumed to be much larger than the volume within the gap, so that they may be treated as an infinite reservoir. It will be assumed that the gap is so small that Poiseuille flow will prevail for all frequencies of interest, and that frequency dependent corrections will not be necessary. Without going into the details of the flow pattern and the distribution of flow velocity and fluid density within the gap, a simple first order differential equation with constant coefficients may be applied to model the reactive force between the two surfaces.

In compressive motion, F must equal the elastic reaction of the solid contact, represented by the coefficient k_c, and the bulk reaction of the fluid film, defined by the coefficients k_y and c_r with respect to first order changes in y and r, respectively. The coefficient k_c is related to the size and shape of the solid contact and the elastic properties of the solid material, and similarly k_y and c_r are related to the size and shape of the fluid film and its bulk modulus. A second equation is obtained from the balance between the fluid pressure differential and the drag force represented by the drag coefficient b, which is determined by the viscosity of the fluid and the radial permeability of the gap:

$$F=(k_c+k_y)(y-y_o)+c_r(r-r_o), \tag{2}$$

$$b\frac{dr}{dt}=-k_y(y-y_o)-c_r(r-r_o). \tag{3}$$

All the coefficients are treated as constants, and the time dependence of the solution is assumed to be of the form $e^{-i\omega t}$. The resulting expression for the contact stiffness, i.e., the ratio between applied force and y-displacement amplitudes, is obtained in terms of two constants and a relaxation frequency, ω_k. Setting

$$F=A_f e^{-i\omega t}, \tag{4}$$

$$y-y_o=A_y e^{-i\omega t}, \tag{5}$$

the solution for the compressive stiffness of the gap is

$$\frac{A_f}{A_y}=k_c+\frac{k_y}{1+i(\omega_k/\omega)} \quad \text{where} \quad \omega_k=\frac{c_r}{b}. \tag{6}$$

It shows that at very low frequencies, the solid contact, represented by the spring, will dominate. At very high frequencies, the fluid has not the time to flow into or out of the gap, and, due to the larger contact area of the fluid film, the compressibility of the fluid film will dominate. Between the extremes, there is a transition region, in which viscous drag will dominate.

On a microscopic scale, the frame is stiffened preferentially along the normal of the grain to grain contact. On a macroscopic scale, in the context of wave propagation and where the wavelength is much greater than the grain size, the aggregate effect of numerous grain to grain contacts within an elemental volume dominates. In this case, the contact orientation may be considered as randomly and homogeneously distributed over angle space, and the resulting effective frame bulk and shear stiffness is isotropic for practical purposes.

This model is practically identical to the Biot plus squirt flow (BISQ) model of Dvorkin and Nur.[20] However, there are a few differences: Dvorkin and Nur considered squirt flow as being perpendicular to the longitudinal wave direction, but here squirt flow is within the fluid film at the grain to grain contact which may have any orientation. They considered squirt flow to be governed by the same permeability as the longitudinal flow and a characteristic flow length R, but, in this case, the flow length is the radius of the contact area and the squirt flow takes place within the grain to grain contact, which is a tightly confined region with a significantly different permeability. Nevertheless, the general characteristics of the models, i.e., the low and high frequency asymptotes and a transition governed by a relaxation frequency, are similar.

The shear response of the grain to grain contact region, controlled by the elastic response of the solid contact and the viscous response of the fluid film, is modeled as a simple spring and dash-pot as shown in Fig. 4(d). The shear deformation angle θ is related to the shear S/y_o stress by

$$\frac{S}{y_o}=g_c\theta+h\frac{d\theta}{dt}, \tag{7}$$

where g_c is the shear stiffness of the solid contact and h the shear drag coefficient associated with the fluid film.

The solution for the net shear stiffness is given by

$$\frac{S}{y_o}=A_S e^{-i\omega t}, \tag{8}$$

$$\theta=A_\theta e^{-i\omega t}, \tag{9}$$

$$\frac{A_S}{A_\theta}=g_c\left(1-i\frac{\omega}{\omega_\mu}\right), \quad \text{where} \quad \omega_\mu=\frac{g_c}{h}. \tag{10}$$

At very low frequencies, the solid contact will dominate. At very high frequencies, it would appear that the viscous drag will increase monotonically, which is unphysical. In practice, there will likely be a high frequency asymptotic shear stiffness, similar to the compressive stiffness. However, the asymptotic shear stiffness is expected to occur well beyond the highest of frequencies of interest. Therefore, the above

approximation is expected to be adequate for practical purposes.

Finally, the contact model may be related to the effective frame bulk and shear moduli. Since the stiffness of water is more than an order of magnitude less than that of quartz and the other minerals that make up the solid particles, the frame strain must occur mainly at the grain to grain contacts. Given that the grain compressibility and susceptibility to shear are negligible compared to that of the contact, the above results can be scaled up to the effective frame bulk K_b and shear μ moduli,

$$K_b = K_c + \frac{K_y}{1 + i(\omega_k/\omega)}, \tag{11}$$

$$\mu = G_c\left(1 - i\frac{\omega}{\omega_\mu}\right), \tag{12}$$

where K_c and G_c are the asymptotic frame bulk and shear moduli at the low frequency limit, K_y is the difference between the asymptotic high and low frequency values of the bulk modulus, and ω_k and ω_μ are the bulk and shear relaxation angular frequencies. Let us define the relaxation frequencies, in Hz, as

$$f_k = \frac{\omega_k}{2\pi}, \tag{13}$$

$$f_\mu = \frac{\omega_\mu}{2\pi}. \tag{14}$$

This will be called the Biot–Stoll plus grain contact squirt and shear flow (BICSQS) model in recognition of the prior work by Dvorkin and Nur. The complex frame moduli in the Biot–Stoll model will be computed using the above equations. There is no increase in the number of parameters because the four constants associated with the complex frame moduli, i.e., the real and imaginary parts of the frame bulk and shear moduli, are replaced by computed values using the four independent parameters, i.e., G_c, K_y, f_k and f_μ. The remaining parameter K_c is not independent because it is related to G_c by the low frequency frame Poisson's ratio. The frequency dependent frame moduli have two advantages over the constants that they replace: (1) They are computed from a physically sound mechanical model at the grain to grain contact level, and (2) the resulting model is entirely causal. According to the Hertz–Mindlin model, the frame Poisson's ratio should be less than 0.1, but measurements reported by Bachrach, Dvorkin and Nur[21] in dry sand and glass beads indicated a value of 0.15. This is the value that will be used to compute K_c.

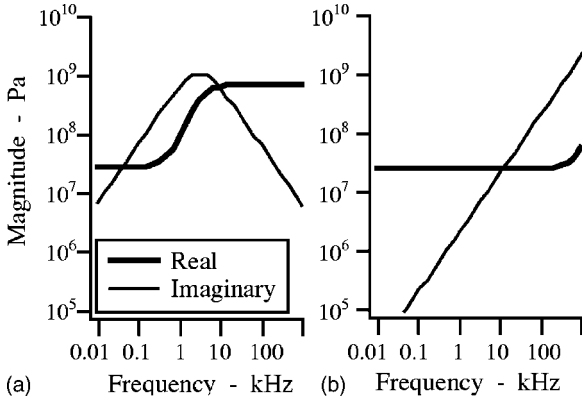


FIG. 5. (a) Complex frame bulk modulus as a function of frequency, for different values of the bulk relaxation frequency. (b) Complex frame shear modulus as a function of frequency for different values of the shear relaxation frequency.

IV. PROPERTIES OF THE BICSQS MODEL

The properties are explored as follows. Let us start with the parameter values of the baseline TY model, and replace the constant frame moduli with the above calculations. The characteristics of the sound speed and attenuation dispersion depend on the values of the contact relaxation frequencies. The following cases will be used for illustration purposes.

A. Case 1: High (infinite) bulk and shear contact relaxation frequencies

If the relaxation frequencies are well above the frequencies of interest, then the frame bulk and shear moduli retain

TABLE III. Example cases of the BICSQS model.

Parameter	Units	1	2	3	4	5
Bulk properties						
β, porosity	⋯			0.44		
ρ_r, grain density	kg/m^3			2650		
ρ_f, fluid density	kg/m^3			1000		
K_r, grain bulk modulus	GPa			36		
K_f, fluid bulk modulus	GPa			2.25		
Fluid motion						
η, fluid viscosity	kg/m-s			0.001		
κ, permeability	μ^2			17.5		
a, pore size	μ			28		
c, virtual mass coefficient	⋯			1.24		
Frame response						
μ, frame shear modulus	GPa			0.024		
K_y, frame bulk mod. difference	GPa			0.675		
f_k, bulk relaxation frequency	kHz	3	30	300	∞	∞
f_μ, shear relaxation frequency	kHz	∞	∞	∞	30	300

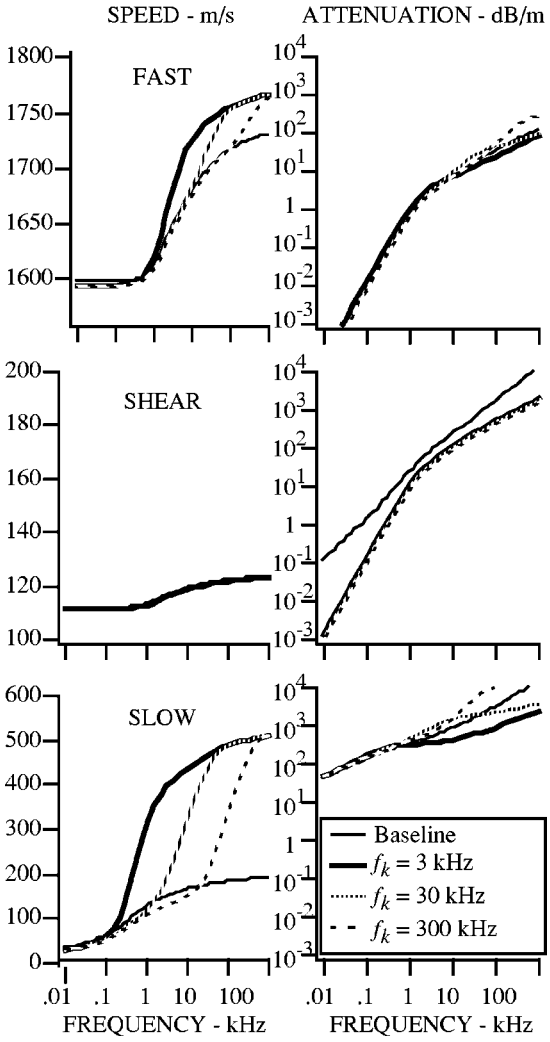


FIG. 6. Examples of wave speeds and attenuations as a function of frequency, at low bulk and high (infinite) shear contact relaxation frequencies.

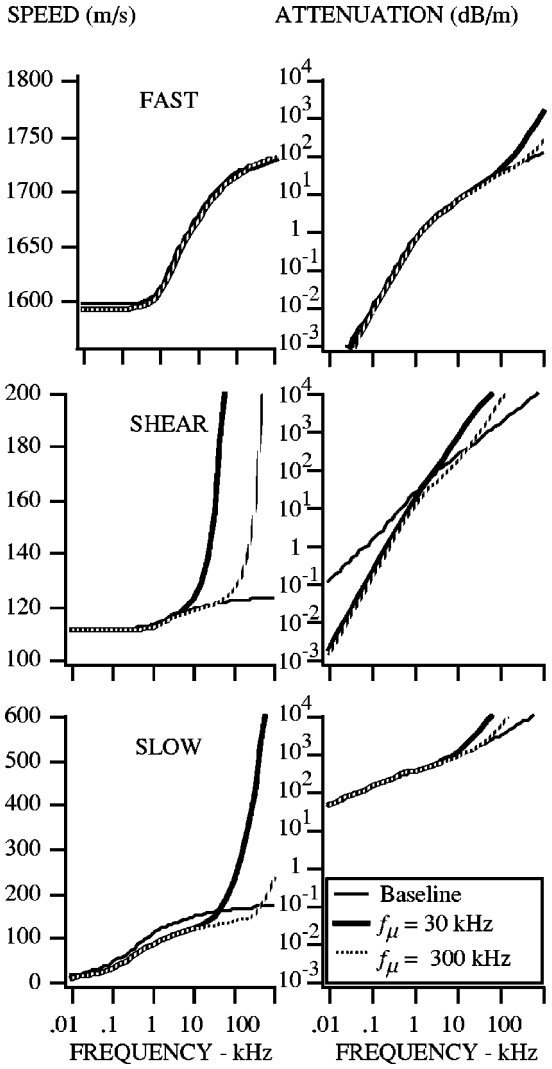


FIG. 7. Examples of wave speeds and attenuations as a function of frequency, at high (infinite) bulk and low shear contact relaxation frequencies.

their low frequency asymptotic values. This is the baseline Biot–Stoll model with constant coefficients. The attenuation goes as f^2 at low frequencies and as $f^{1/2}$ at high frequencies.

B. Case 2: Low bulk and high (infinite) shear contact relaxation frequencies

The frame bulk modulus will change with frequency, transitioning from its low frequency to its high frequency values in the vicinity of the relaxation frequency, where the imaginary part goes through a peak, as illustrated in Fig. 5(a). The frame shear modulus remains constant and real. Three different values of the bulk relaxation frequencies are illustrated, 3, 30 and 300 kHz. The parameter values are shown in columns 1, 2 and 3 of Table III. The fast, shear and slow wave speeds and attenuation are shown in Figs. 6(a)–(f). The fast wave speed goes through a larger increase in value than the baseline case. The magnitude of the increase is governed by the term K_y. The shear speed and attenuation are not significantly affected. The slow wave speed goes through a very significant increase as frequency increases beyond the relaxation frequency. Of course, in a numerical model, it is not possible to use infinity as the shear relaxation frequency, but a large finite value of 10^{10} Hz was adequate for practical purposes.

C. Case 3: High (infinite) bulk and low shear contact relaxation frequencies

The imaginary part of the shear modulus increases linearly with frequency beyond the relaxation frequency, as illustrated in Fig. 5(b). In this case the sound speed dispersion is small, as in case 1, but the attenuation goes almost as the first power of frequency. Two different values of the shear relaxation frequencies are illustrated, 30 and 300 kHz. The parameter values are shown in columns 4 and 5 of Table III. The fast, shear and slow wave speeds and attenuation are shown in Figs. 7(a)–(f). The shear wave speed increases rapidly beyond the relaxation frequency. There is also a rapid increase in the fast wave attenuation. The shear attenuation is also modified.

Therefore, depending on the values of the contact relaxation frequencies, it is possible to have a variety of frequency dependencies, with exponents ranging from a $\frac{1}{2}$ to 2. Recalling the measurements plotted in Fig. 1, it is postulated that

TABLE IV. Application of the BICSQS model to experimental data.

Parameter	Units	TY	NHMRW	TP	SH	WJTTS
Bulk properties						
β, porosity	⋯	0.44	0.36/0.37	0.35/0.55	0.37	0.37
ρ_r, grain density	kg/m^3	2650	2650	2650	2650	2690
ρ_f, fluid density	kg/m^3	1000	1000	1000	1000	1023
K_r, grain bulk modulus	GPa	36	36	36	36	36
K_f, fluid bulk modulus	GPa	2.25	2.1	2.1	1.95	2.15
Fluid motion						
η, fluid viscosity	kg/m-s	0.001	0.001	0.001	0.001	0.001
κ, permeability	μ^2	17.5	76/0.7	7.4/10.6	28	115
a, pore size	μ	28	65/6	21/20	39	57
c, virtual mass	⋯	1.24	1.34/1.32	1.35/1.15	1.32	1.35
Frame response						
μ, frame shear modulus	GPa	0.024	0.05	0.05	0.05	0.028
K_y, bulk modulus difference	GPa	0.9	0.42/1.05	0.53/0.11	0.39	0.968
f_k, bulk relaxation frequency	kHz	3.2	3.2	3.2	3.2	4.8
f_μ, shear relaxation frequency	kHz	∞	∞	400/320	320	56
Nominal fluid bulk modulus	GPa	2.3	2.14	2.14	2.14	2.395
Required gas fraction $\times 10^6$	⋯	2.8	2.2	2.2	11.3	35.4

different combinations of the bulk and shear relaxation frequencies could account for the diverse frequency dependencies observed.

V. APPLICATION TO EXTANT MEASUREMENT DATABASE

In all cases, the grain bulk modulus was assumed to be that of quartz since quartz sand was used in all of the experimental data. The grain density was assumed to be 2650 kg/m^3, except in the case of WJTTS where the grain density was specified as 2690 kg/m^3. For the laboratory experiments using fresh water, the fluid density was set to 1000 kg/m^3 and the fluid bulk modulus was expected to be in the region of 2.14 GPa, appropriate for fresh water at room temperature and atmospheric pressure.

The BICSQS model was applied to the TY experiment as follows. (1) The model was matched to the low frequency shear wave speed measurement by adjusting the value of G_c. (2) The model was matched to the low frequency sound speed by adjusting the fluid bulk modululs. Ideally, this should have happened without any adjustment, because the sound speed lower bound is simply given by the Wood equation for a suspension of the same porosity. However, it was found that the Wood equation sound speed (1585 m/s) was slightly higher than the lowest measured value (1575 m/s). There are a few possible causes. The simplest one is that the value of the pore fluid bulk modulus may be lower than the expected value of 2.3 Pa, due to the presence of microscopic gas bubbles. A slightly reduced value of 2.25 GPa allowed the model to match the low frequency sound speed. This corresponds to a gas volume fraction of only 2.8 parts per million (ppm) as shown in Table IV. No measurements were made of pore fluid gas fraction. (3) The values of K_y and f_k were adjusted to match the measured high frequency sound speed and the transition frequency. (4) The value of f_μ was adjusted to match the slope of the measured attenuation curve at high frequencies. In this case, $f_\mu = \infty$ gave the best fit. (5) The value of permeability was adjusted to match the measured absolute values of attenuation. The best-fit value of permeability was consistent with the measured value reported by TY. The pore size was computed using Eq. (1), and the virtual mass coefficient was provided by TY. The model parameters are shown in column 1 of Table IV and the results are shown in Figs. 8 and 9. The BICSQS model is unable to track the oscillations in the measured curves but it tracks the underlying trends of both the sound speed and attenuation as a function of frequency.

The NHMRW experiment involved four sand samples, but, for brevity, only the results from the samples with the smallest and largest attenuation will be modeled. The experiment did not include a measurement of shear wave speed, therefore, G_c was set to a default value of 0.05 GPa. Measurements of porosity and permeability were provided. Only an average value of the sound speed was given, therefore it was not possible to make a determination of the value of f_k, and the same value that was found for YT was assumed. The

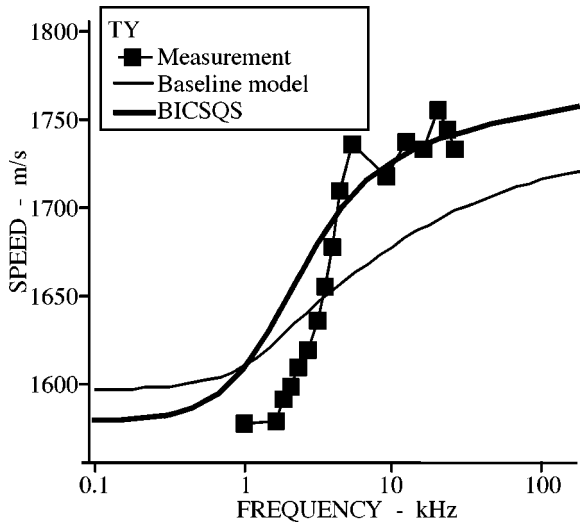


FIG. 8. Comparison of sound speed as a function of frequency in the TY experiment and the BICSQS model.

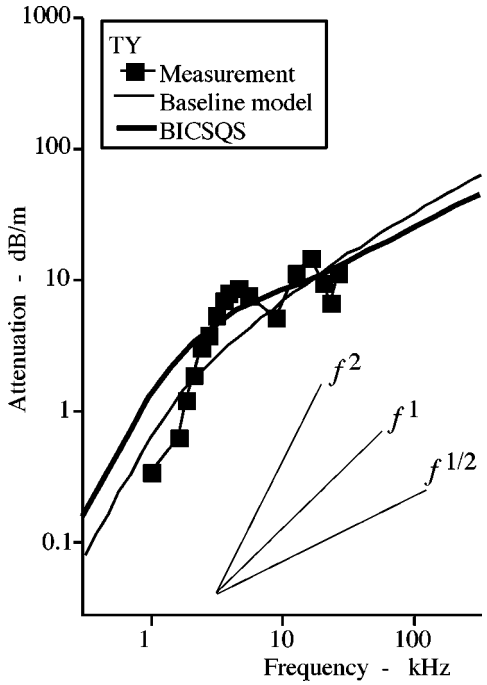


FIG. 9. Comparison of attenuation as a function of frequency in the TY experiment and the BICSQS model.

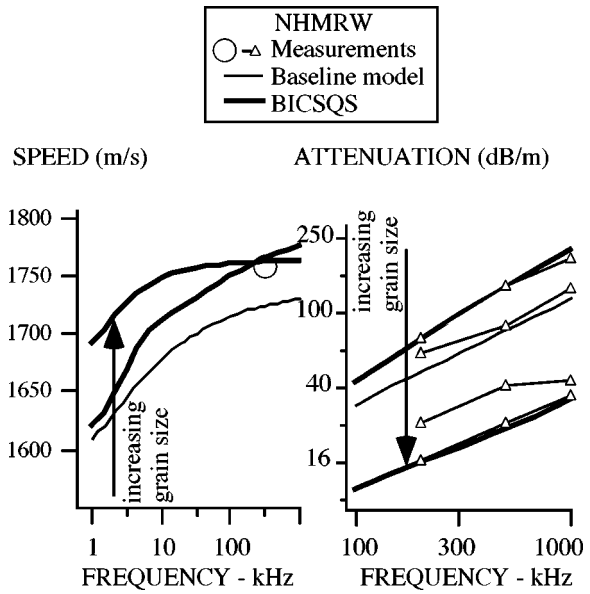


FIG. 10. Sound speed and attenuation as a function of frequency in the NHMRW experiment and the BICSQS model.

value of K_y was adjusted to fit the measured sound speed at 300 kHz—the center of the frequency band. The value of the fluid bulk modulus was not explicitly provided, but it was estimated to be 2.14 GPa, corresponding to distilled water at room temperature. A slightly lower value of 2.10 GPa was required to fit the measurements. The added mass term was computed using Eq. (18) from Ref. 22.

$$c = 1 + r\frac{(1-\beta)}{\beta}, \tag{15}$$

where the value of r, a dimensionless constant, was inverted from the measured values of c and β from the TY experiment ($r = 0.1885$). To fit the slope of the attenuation curve, again, $f_\mu = \infty$ gave the best fit. To fit the measured absolute values of attenuation, the value of permeability was adjusted. The values of permeability that gave the best fit to the highest and lowest measured attenuation curves, 76 and 0.7 μm^2, are somewhat smaller than the corresponding measured values, 242 and 6 μm^2, reported by NHMRW using a constant flow method. It is likely that the permeability for acoustic motion may be significantly different than that of a constant flow. It is also possible that the samples used in the permeability measurement were not as well compacted as the samples used in the acoustic measurements, giving rise to the higher measured values. The pore size parameter was calculated according to Eq. (1) with $k_k = 5$. The model parameters are shown in column 2 of Table IV and the results are shown in Fig. 10. The TY baseline model is also shown for comparison.

The TP experiment involved five sand samples but only the largest and smallest attenuation cases will be modeled for brevity. As in the previous data set, there was no measurement of shear wave speed and only one average sound speed measurement was given for each sand sample, therefore, the same default values of G_c and f_k are used. The value of K_y was adjusted to fit the measured sound speeds. To fit the measured slopes of the attenuation curve, the required values of f_μ were found to be 400 and 320 kHz. The permeability was adjusted to fit the absolute measured values of the attenuation. No measurements of permeability were provided by TP. The pore size parameter was calculated according to Eq. (1) with $k_k = 5$, and the added mass fraction was computed using Eq. (15). The model parameters are shown in column 3 of Table IV and the results are shown in Fig. 11.

The SH experiment contained a measurement of sound speed as a function of frequency but it was not adequately precise to show any deviation from the mean value of 1680 m/s. As in the previous cases, there was no measurement of shear wave speed and the same default values of G_c were

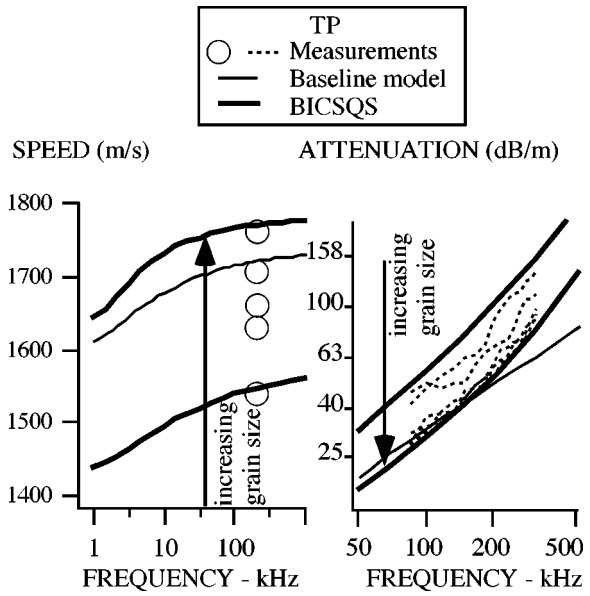


FIG. 11. Sound speed and attenuation as a function of frequency in the TP experiment and the BICSQS model.

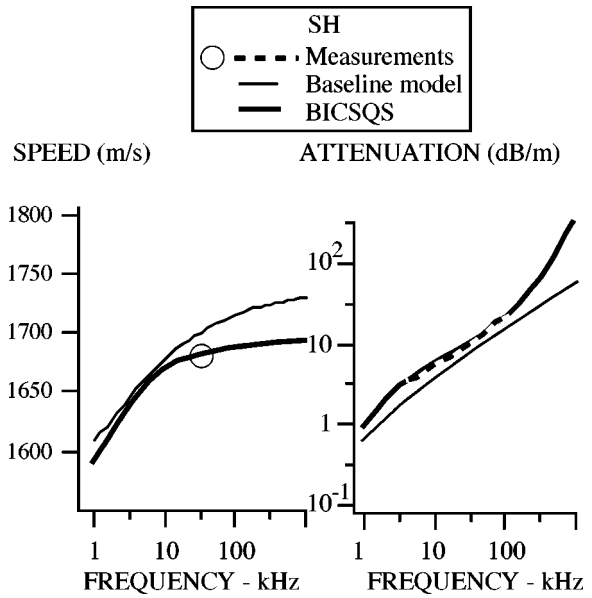


FIG. 12. Sound speed and attenuation as a function of frequency in the SH experiment and the BICSQS model.

used. The frequency range of the attenuation measurement was adequately extensive to be sensitive to the value of f_k and K_y, particularly at the low end of the band. Their values were adjusted to obtain the best fit to the shape of the curve. To fit the attenuation curve at the high end of the band, the value of f_μ was adjusted. To fit the measured mean value of sound speed, it was necessary to reduce the fluid bulk modulus, from the given value of 2.2 to 1.95 GPa, corresponding to a gas volume fraction of 11.3 ppm. Finally, the value of permeability was adjusted to fit the absolute value of attenuation. The best fit value, 28 μm^2, was reasonably close to the measured value of 39.8 μm^2. The parameter values are shown in column 4 of Table IV, and the results are in Fig. 12.

Finally, the WJTTS experiment was modeled using the same procedure as for the TY experiment. In modeling the WJTTS experiment, the grain bulk modulus was set at the value for quartz of 36 GPa, rather than the somewhat lower value adopted by WJTTS, and the porosity was set at 0.37, the value found from core sample measurements, rather than the higher value adopted by WJTTS. To match the low-frequency sound speed, again, it was necessary to reduce the given value of fluid bulk modulus, from 2.395 to 2.15 GPa, equivalent to a gas fraction of 35.4 ppm. Attempts were made to measure the pore water gas fraction, but, due to the sensitivity limits of the equipment, it could only be determined that the gas fraction did not exceed 150 ppm.[17] The values of K_y and f_k that best fit the sound speed curve are shown in column 5 of Table IV. The value of f_μ was adjusted to match the shape of the attenuation curve. The resulting value, 56 kHz, is the lowest of the five experimental data sets. Finally, to match the absolute values of the measured attenuation the required value of permeability was 115 μm^2, which is higher than the measured value of 25 μm^2 using the constant flow method. The results are shown in Figs. 13 and 14.

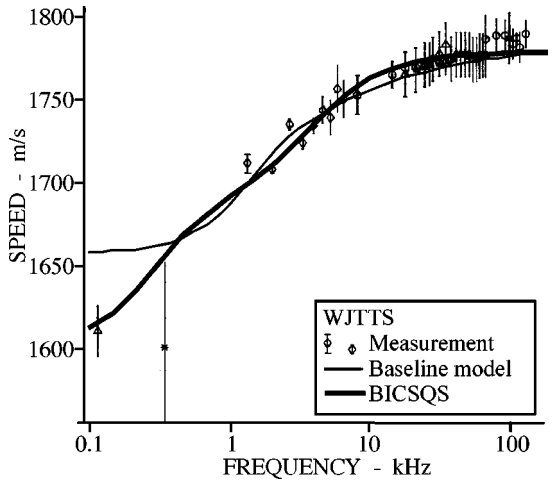


FIG. 13. Comparison of sound speed as a function of frequency in the WJTTS experiment and the BICSQS model.

VI. CONCLUSIONS AND DISCUSSIONS

In the BICSQS extension of the Biot–Stoll model, a new representation of the elastic response of the frame is proposed. In the baseline Biot–Stoll model, the frame response is described in terms of complex bulk and shear moduli. Since they are impossible to measure, they were treated as free parameters, and they were adjusted to fit measured values of the compressive and shear wave speeds and attenuations. The values were assumed to be constant. As a result, the resulting model was not causal, and values obtained at one frequency often did not apply at other frequencies.

The extension employs a physical model of the grain to grain contact, which includes squirt flow and shear drag, to compute the frame moduli. The contact is modeled as a pair of relaxation processes in compression and shear. The model

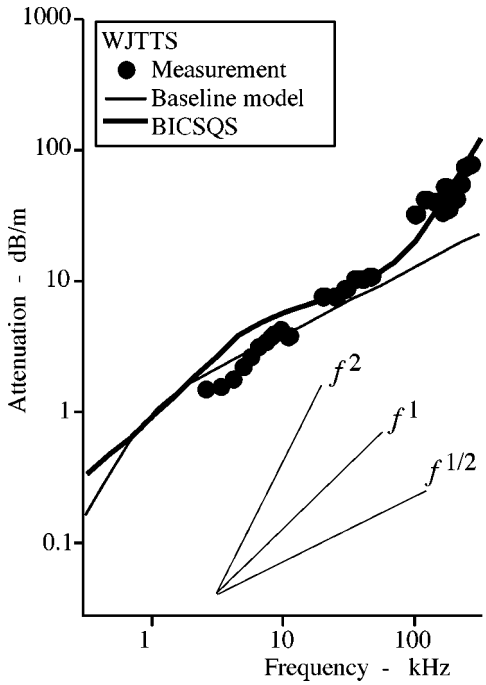


FIG. 14. Comparison of attenuation as a function of frequency in the WJTTS experiment and the BICSQS model.

is causal and based on the physics of fluid flow, and applicable over a broad range of frequencies. It differs from the commonly used approach of complex frame moduli with constant coefficients, which is noncausal, and has no physical analog. There is no net increase in the number of model parameters, since the four constants of the frame moduli are replaced by the four parameters of the new frame model.

The model is capable of matching the sound speed dispersion and the diverse frequency dependencies that are found in the extant measurement database. This was demonstrated with the five most comprehensive experimental data sets in the extant published database.

The four parameters of the BICSQS frame model include the low frequency frame shear modulus, the difference between high and low frequency asymptotic frame bulk modulus values, and the compressive and shear relaxation frequencies. The low frequency frame bulk modulus is computed from the low frequency shear modulus via the low frequency Poisson's ratio. Although the Poisson's ratio is not directly measurable, it is expected to be equal to the Poisson's ratio of the dry frame, which has been measured for several granular media.

The frame shear and bulk moduli are coupled via the dry frame Poisson's ratio, but only in the low frequency limit. At higher frequencies, the fluid film at the grain to grain contact preferentially stiffens the frame bulk modulus and effectively decouples the bulk and shear moduli.

The difference between high and low frequency asymptotic frame bulk modulus values represents the stiffening of the grain to grain contact by the fluid film. The frame bulk stiffness radically changes at the bulk relaxation frequency. The best-fit value of the frame bulk relaxation frequency was found to be in the region from 3 to 5 kHz for all five experimental data sets. The significance of this range of values has not been explored yet, but it is expected to be connected to the average dimensions of the fluid film.

The frame shear relaxation frequency varied from a high of infinity for very clean well-sorted laboratory sands, as in NHMRW, to a low of 56 kHz for *in situ* poorly sorted sand which contained a broad range of grain sizes as well as biological material as in WJTTS. The poorly sorted, but otherwise clean, sand of TP and SH had values in the region of 300–400 kHz. Clearly, a connection between the degree of sorting and the frame shear relaxation frequency is indicated.

There is some argument over whether it is appropriate to use the Biot–Stoll model to explain sound attenuation in water-saturated sand at all in the range of frequencies considered here. The excess attenuation, above the baseline Biot model, has been ascribed to scattering by WJTTS. There are two general types of scatterers: smaller and larger than the acoustic wavelength. If the attenuation is mainly due to small scatterers, such as the sand grains, it would be very significant because scattering by small scatterers is omnidirectional. If this were the dominant cause of attenuation, it should increase with grain size. However, the data from TP and NHMRW clearly show that the opposite is true. The measured attenuation decreases with increasing grain size. Therefore, although there are numerous small scatterers, they are not the dominant cause of attenuation in the range of frequencies considered. Scatterers larger than the wavelength may be characterized as volume inhomogeneities, such as spatial variations in sound speed with position. They are weak scatterers that mainly produce forward scattering. They have the effect of distorting the acoustic wave front and destroying coherence without significant loss of total acoustic energy. Undoubtedly, the latter type of scattering occurs in water-saturated sand, but it cannot be the principle cause of attenuation.

Rather than looking to scattering, one should look to permeability. It represents the susceptibility of the frame to pore fluid flow. Using clean graded laboratory sand, it was shown in NHMRW that there is a linear relationship between the permeability and attenuation. The loss mechanism is not scattering but viscous dissipation. The BICSQS model adds two more viscous absorption mechanisms to the baseline Biot model: the squirt flow and the shear drag losses. They are negligible at low frequencies. At intermediate frequencies, the squirt flow losses will dominate, peaking in the vicinity of the bulk relaxation frequency. At higher frequencies, the shear drag losses will kick in and cause the attenuation to increase more steeply with frequency than in the baseline model.

The value of permeability directly controls the absolute value of attenuation. Higher values of permeability correspond to lower values of attenuation. Larger grain sizes produce larger pore spaces, hence higher values of permeability. This explains the observed reduction in attenuation with increasing grain size within the frequency range in which these processes are dominant.

In this study, the value of permeability was adjusted to fit the offset of the measured curve of attenuation versus frequency, and checked against independently measured values where available. Permeability measurements were provided in four of the five data sets considered here. In NHMRW, it was necessary to use a value that was lower than the measured value. In WJTTS it was necessary to use values that were higher than the measured values. In TY and SH, the measured values were consistent with the model values. It appears that measured values of permeability can come within an order of magnitude of the model values. This is consistent with observations that permeability measurements may only be accurate to within an order of magnitude.[23] It is also recognized that permeability measured with constant flow methods may not be a good indicator of permeability for acoustics which involves an oscillating flow, but that is an issue that is beyond the scope of this study.

The low frequency sound speed remains an open issue. Given the nominal bulk properties of the sediment material, it was not possible to explain the measured low frequency sound speed. In this study, a choice was made to assume that the pore fluid bulk modulus was reduced, possibly by the presence of minute concentrations of gas bubbles. The concentrations indicated are up to 40 ppm by volume, which is extremely difficult to measure. Of the five experiments considered, only the SAX99 experiment included attempts to measure the pore fluid gas content, but the apparatus could only resolve 150 ppm or more, therefore a gas fraction of 40 ppm cannot be ruled out. Alternatively, there may be an, as

yet, unknown mechanism or process that depresses the low frequency sound speed. This remains an open issue for future research.

In summary, the procedure for determining the value of permeability and the four BICSQS parameters consists of the following iterative steps, given that all the other Biot parameters are known. (1) The value of G_c was adjusted to match the asymptotic low frequency shear wave speed. (2) The values of K_y and f_k were simultaneously adjusted to maximize the agreement between the model and the measured sound speed dispersion curve. (3) The value of f_μ was adjusted to match the slope of the measured attenuation curve at high frequencies. (4) The value of permeability was adjusted to match the absolute measured values of attenuation at high frequencies. Steps (1)–(4) were repeated iteratively until a stable solution was achieved. The terms ''low'' and ''high'' frequencies are referenced to the bulk wave relaxation frequency f_k, which was estimated to be in the region of 3 kHz for all the cases considered here.

Finally, no attempt was made in this study to compare the model predictions of broadband reflection loss to the measured values. This is expected to lead to further model refinements, particularly the introduction of the ''composite medium'' submodel.[24] This will be pursued in a future study.

ACKNOWLEDGMENTS

The authors would like to acknowledge our colleagues in the SAX99 program, particularly chief scientist Eric Thorsos, Applied Physics Laboratory, University of Washington (APL/UW), Michael Richardson, Naval Research Laboratory (NRL), and the scientists and students from all the participating institutions, for their cooperation and the many discussions that we have had on this and related topics. We would like to express our appreciation to Morris Stern for modifying the OASES code to accommodate the BICSQS model, and all our colleagues at the Applied Research Laboratories, University of Texas at Austin (ARL:UT) for their contributions in the data collection and theoretical analysis. This work is sponsored by the Office of Naval Research, Code 321 OA, under the management of Robert Headrick.

APPENDIX: INSIGNIFICANCE OF INERTIAL TERM AND POISEUILLE FLOW IN THE GRAIN CONTACT SQUIRT FLOW

With respect to the grain contact squirt flow, let us model the grain to grain contact as a fluid film sandwiched between two planar boundaries. The starting point is the expression for the drag force on the fluid film between two parallel planar surfaces given by Biot in Eq. (2.18) of Ref. 3. Using Biot's nomenclature, and assuming that all quantities are sinusoidal functions of time with a factor $e^{i\alpha t}$, the viscous drag force per unit area of the fluid film is given as

$$2\tau = 6\mu U_{1(\mathrm{Av})} F_1(\kappa_1)/a_1, \tag{A1}$$

where τ is the drag force on each wall, μ is the viscosity, $U_{1(\mathrm{Av})}$ is the average flow velocity of the fluid film, a_1 is half the film thickness, and $F_1(\kappa_1)$ is the frequency dependent correction representing the deviation from Poiseuille friction at high frequencies. The dimensionless term κ_1 is defined as

$$\kappa_1 = a_1(\alpha\rho/\mu)^{1/2}, \tag{A2}$$

where ρ is the fluid density.

At low frequencies, $F_1(\kappa_1)$ tends to 1, and at high frequencies both its real and imaginary parts are approximately equal to $0.234\kappa_1$. Thus, the magnitude may be approximated by

$$|F_1(\kappa_1)| = (1 + 0.110\kappa_1^2)^{1/2}. \tag{A3}$$

The inertial reaction R_i on the fluid film is simply given by the product of the mass $2a_1\rho$ and the acceleration $i\alpha U_{1(\mathrm{Av})}$,

$$R_i = 2a_1\rho i\alpha U_{1(\mathrm{Av})}. \tag{A4}$$

Taking the ratio of the magnitudes, the approximate result is

$$2|\tau|/|R_i| = 3(\mu/\rho)a_1^{-2}\alpha^{-1}\{1 + 0.110a_1^2(\alpha\rho/\mu)\}^{1/2}. \tag{A5}$$

For water, the tabulated values of ρ and μ are 1000 kg/m^3 and 0.001 kg/m-s, respectively. For a gap size of 10 μm or less, and for frequencies less than 100 kHz, the ratio is much greater than 1, indicating that the inertial reaction is insignificant compared to the viscous drag. Furthermore, the value of $|F_1(\kappa_1)|$ remains close to its low frequency asymptotic value, indicating little deviation from Poiseuille flow.

[1] E. L. Hamilton, in *Physics of Sound in Marine Sediments*, edited by L. Hampton (Plenum, New York, 1974), pp. 181–221.
[2] N. P. Chotiros, ''Inversion and sandy ocean sediments,'' in *Full Field Inversion Methods in Ocean and Seismic Acoustics*, edited by O. Diachok, A. Caiti, P. Gerstoft, and H. Schmidt (Kluwer Academic, Dordrecht, 1995).
[3] 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).
[4] R. D. Stoll, *Sediment Acoustics* (Springer-Verlag, New York, 1989).
[5] P. R. Ogushwitz, ''Applicability of the Biot Theory III: Wave speed versus depth in marine sediments,'' J. Acoust. Soc. Am. **77**, 453–464 (1985).
[6] J. G. Berryman, ''Confirmation of Biot's theory,'' Appl. Phys. Lett. **37**(4), 382–384 (1980).
[7] R. D. Stoll, ''Acoustic Waves in Ocean Sediments,'' Geophysics **42**, 715–725 (1977).
[8] E. L. Hamilton, ''Geoacoustic modeling of the sea floor,'' J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[9] A. W. Nolle, W. A. Hoyer, J. F. Mifsud, W. R. Runyan, and M. B. Ward, ''Acoustical properties of water-filled sands,'' J. Acoust. Soc. Am. **35**, 1394–1408 (1963).
[10] P. R. Thomas and N. G. Pace, ''Broadband measurements of acoustic attenuation in water-saturated sands,'' Ultrasonics **18**, 13–17 (1980).
[11] 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).
[12] A. Turgut and T. Yamamoto, ''Measurements of acoustic wave velocities and attenuation in marine sediments,'' J. Acoust. Soc. Am. **87**, 2376–2382 (1990).
[13] K. L. Williams, D. R. Jackson, E. I. Thorsos, D. Tang, and S. G. Schock, ''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**(3), 413–428 (2002).
[14] M. J. Buckingham, ''Wave propagation, stress relaxation, and grain-to-grain shearing in saturated, unconsolidated marine sediments,'' J. Acoust. Soc. Am. **108**(6), 2796–2815 (2000).
[15] R. D. Stoll and T. K. Kan, ''Reflection of acoustic waves at a water–sediment interface,'' J. Acoust. Soc. Am. **70**, 149–156 (1981).
[16] C. Carman, *Flow of Gases through Porous Media* (Academic, New York, 1956).

[17] M. D. Richardson, K. B. Briggs, L. D. Bibee, P. A. Jumars, W. B. Sawyer, D. B. Albert, R. H. Bennett, T. K. Berger, M. J. Buckingham, N. P. Chotiros, P. H. Dahl, N. T. Dewitt, P. Fleischer, R. Flood, C. F. Greenlaw, D. V. Holliday, M. H. Hulbert, M. P. Hutnak, P. D. Jackson, J. S. Jaffe, H. P. Johnson, D. L. Lavoie, A. P. Lyons, C. S. Martens, D. E. McGehee, K. D. Moore, T. H. Orsi, J. N. Piper, R. I. Ray, A. H. Reed, R. F. L. Self, J. L. Schmidt, S. G. Schock, F. Simonet, R. D. Stoll, D. Tang, D. E. Thistle, E. I. Thorsos, D. J. Walter, and R. A. Wheatcroft, ''An overview of SAX99: Environmental considerations,'' IEEE J. Ocean. Eng. **26**(1), 26–53 (2001).

[18] 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**(3), 593–601 (2002).

[19] *CRC Handbook of Chemistry and Physics*, 59th ed. (CRC, Boca Raton, FL, 1978-79), p. F-80.

[20] J. Dvorkin and A. Nur, ''Dynamic poroelasticity: A unified model with the squirt and the Biot mechanisms,'' Geophysics **58**(4), 524–533 (1993).

[21] R. Bachrach, J. Dvorkin, and A. Nur, ''Seismic velocities and Poisson's ratio of shallow unconsolidated sands,'' Geophysics **65**(2), 559–564 (2000).

[22] J. G. Berryman, ''Confirmation of Biot's theory,'' Appl. Phys. Lett. **37**(4), 382–384 (1980).

[23] D. Taylor-Smith, ''Geophysical-geotechnical predictions,'' *Acoustic Classification and Mapping of the Seabed*, Proc. Inst. Acoust. **15**(2), 1–14 (1993).

[24] N. P. Chotiros, ''An inversion for Biot parameters in water-saturated sand,'' J. Acoust. Soc. Am. **112**(5), 1853–1868 (2002).

Compressional- and shear-wave velocities and attenuation in deep-sea sediment during laboratory compaction

Klaus C. Leurer[a)]
Department of Earth and Ocean Sciences, National University of Ireland, Galaway, The Quadrangle, University Rd, Galway, Rep. of Ireland

(Received 12 November 2003; revised 22 June 2004; accepted 23 June 2004)

Velocities and attenuation of compressional (50 kHz) and shear waves (100 kHz) in two deep-sea sediment samples are investigated as a function of effective pressure up to 20 MPa with simultaneous measurement of porosity, density, and permeability. In both samples, the compressional-wave velocities show a strong increase with pressure, ranging from 1683 m/s at 1 MPa to 2128 m/s at 20 MPa in a silty clay and from 1840 m/s to 2462 m/s in a foraminiferal mud. Shear wave propagation is strongly restricted at pressures below 4 and 9 MPa in the respective sample, indicating a structural change in the sediment material at critical porosity values of 0.430 and 0.397, respectively. The compressional-wave attenuation, in terms of $1/Q_P$, varies between 0.03 and 0.08, and 0.04 and 0.07, respectively. A result previously not reported in laboratory studies is a maximum in $1/Q_P$ as a function of effective pressure observed in both samples. The shear-wave attenuation is nearly constant as a function of pressure, having average values of about 0.07. The maximum of compressional-wave attenuation occurs at the respective critical porosity for each sample. Velocities and attenuation from this experiment exhibit a systematic pattern in a Q_P/Q_S vs $(V_P/V_S)^2$ representation. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1782932]

PACS numbers: 43.30.Ma, 92.10.Vz [WMC] Pages: 2023–2030

I. INTRODUCTION

Experimental studies on the elastic properties of rocks have provided a vast amount of knowledge on sound velocity and attenuation as related to rock physical properties. Nur and Wang,[1] in an overview article on seismic-wave velocities, summarize achievements from the literature investigating dependencies of velocity on, among other parameters, porosity, pore fluid type, pore fluid saturation, and clay content in unconsolidated sediments, whereas Toksöz and Johnston[2] give an overview of identified attenuation mechanisms that can be collected under the term internal friction, among them being global fluid flow, local (squirt-) flow, and viscoelastic frame losses. In the present paper, the variation with compaction of the acoustic velocities and attenuation and the physical properties in fine-grained unconsolidated marine sediments are analyzed.

Compaction in fine-grained unconsolidated marine sediments is dominated by a reduction in porosity,[3] which in turn is caused by an increase with depth below the sea bottom of the effective pressure—the difference between overburden pressure and pore pressure—acting on the solid frame of the sediment. In acoustic studies of this type of sediment that include compaction, accurate measurements of porosity are therefore crucial. Systematic laboratory studies on the velocities of compressional and shear waves in unconsolidated marine sediments as a function of effective pressure were reported by Laughton,[4] who found a significantly larger increase in compressional-wave velocity types with pressure in the finer-grained materials than in coarser ones, because of a more pronounced porosity reduction in the former. Laughton's results support the validity of an expression found by Nafe and Drake[5] that relates compressional-wave velocity and porosity in compacting unconsolidated ocean sediments and which—by taking into account solid–frame interactions in Wood's isostress formulation[6] for sound velocity in suspensions—represents a physically more realistic relationship than, e.g., the heuristic time-average equation.[7] The attenuation of compressional waves in unconsolidated marine sediments was studied by Hamilton[8] by analyzing a large collection of *in situ* and laboratory data for correlations with porosity and grain size. The dependence on depth below the sea bottom of acoustic wave attenuation was also studied by Hamilton,[9] who found indications that the compressional-wave attenuation in fine-grained unconsolidated marine sediments passes a maximum value at a certain depth of the order of several hundred meters, an observation that contrasts the monotonic decrease with depth of attenuation in coarse-grained sediments. Field studies by Mitchell and Focke[10] have confirmed a depth-dependent attenuation maximum.

The work presented here aims at understanding the relationships between acoustic and physical properties in fine-grained unconsolidated marine sediments as they vary with effective pressure. To date, no detailed laboratory studies on these highly compactible sediments with respect to the dependence on compaction of velocities as well as attenuation of both compressional and shear waves have been reported in which porosity, density, and permeability could be observed throughout the experiment. Of particular interest is the question whether the stages the sediment passes resulting in a pressure-dependent attenuation maximum as found by

[a)]Now at: Department of Earth and Ocean Sciences, National University of Ireland, The Quadrangle Galway, Galway, Rep. of Ireland.

TABLE I. Coring location and physical parameters of the sediment samples.

Box core 23 254-2	Research Ship	Cruise 7,
Latitude: 73°03′48″N	METEOR	1 June to 28 September 1988
Longitude: 9°43′24″E		
Water depth: 2227 m		
	Sample 1	Sample 2
Sediment type	Silty clay	Foraminiferal mud
Depth below sea bottom	1.75–1.90 mbsf	5.53–5.68 mbsf
Water content	74.70%	52.70%
Wet bulk density	1560 kg/m^3	1690 kg/m^3
Porosity	0.673	0.597
Carbonate content	14.22%	18.62%
Grain size distribution		
Sand	5.03%	5.91%
Coarse silt	6.00%	13.79%
Medium silt	7.80%	13.01%
Fine silt	14.65%	18.12%
Clay	66.52%	49.17%
Composition of clay fraction		
Montmorillonite	35%	10%
Illite	38%	56%
Kaolinite	16%	22%
Chlorite	10%	12%

Hamilton[9] can be reproduced under the controlled conditions in an accurate laboratory experiment.

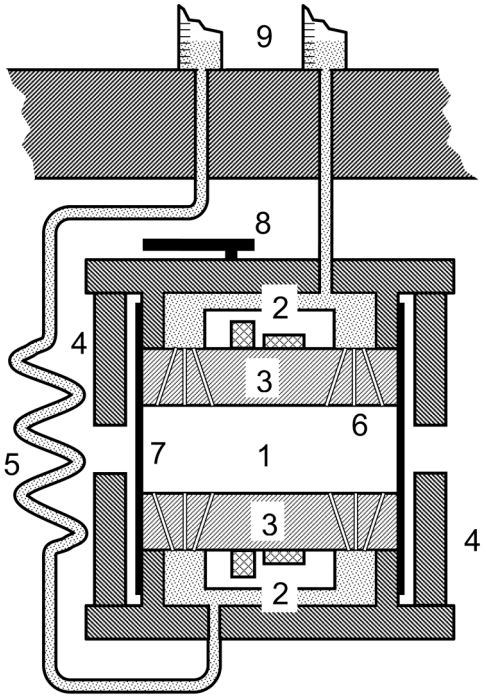


FIG. 1. Central part of the pulse transmission apparatus (schematic): (1) sample container (ø 150 mm); (2) acrylic–glass capsules with compressional- and shear-wave transducers; (3) transmitter plates; (4) distance probes; (5) flexible metal tube for drainage of pore water; (6) drainage channels (ø 1 mm); (7) jacket of reinforced rubber; (8) thermometer; (9) measuring cylinders for drained pore water.

II. EXPERIMENT

For the investigation of rocks in the laboratory, a variety of methods has been implemented, the nature of which determines the experimental outline with respect to sample size, rock type, frequency, and type of excitation. The pulse transmission method[11] is used for the experiments described here because it has a great advantage over other methods in that it allows the measurement as a function of pressure of both compressional- and shear-wave velocity in a single experiment.

A. Sediment samples

The two deep-sea sediment samples investigated in this experiment were taken from box core 23 254-2 drawn at latitude 73° 03′ 48″ N and longitude 9° 43′ 24″ E at water depth of 2227 m during Cruise 7, from 1 June to 28 September 1988, of the Research Ship METEOR.[12] The coring device was a gravity corer of 12-m length with quadratic cross section of 300-mm lateral length. The samples were carefully cut from the fresh core by stainless-steel cylinders of 150-mm diameter and 150-mm length that also served as storage containers. Initial porosity and density were determined on board ship from the remaining undisturbed part of the core material. Subsequent analyses comprised determination of carbonate content, grain size distribution, and composition of the clay fraction of the respective sediment sample. In Table I all physical parameters determined for the samples are summarized. Sample 1 is a silty clay from the depth range 1.75–1.90 mbsf, i.e., meters below the sea floor, and sample 2 is a foraminiferal mud from depth range 5.53–5.68 mbsf. The samples differ from each other by their carbonate and clay contents as well as by the composition of their clay fractions (Table I).

B. Experimental apparatus

The central part of the pulse transmission apparatus (Fig. 1) consists of a cylindrical unit that serves as the sample chamber. During measurement, this unit is immersed in the pressure vessel in which a maximum hydrostatic pressure of 20 MPa can be generated. The pressure medium is water. The piezo-ceramic transducer pairs, whose principal frequencies are 50 kHz for the compressional waves and 100 kHz for the shear waves, respectively, are mounted on transmitting plates that form the bottom and the top of the sample chamber. A flexible tube of reinforced rubber separates the lateral surface of the sample from the water in the pressure vessel. The varying sample length is measured acoustically by three pairs of distance probes that consist of a combination of piezo-ceramic transducers and aluminum cylinders.

For the experiments on the highly compactible fine-grained deep-sea sediments, the apparatus was equipped with measuring cylinders outside of the pressure vessel (Fig. 1). This allowed the measurement of the volume of pore water driven out of the sediment sample and, thus, the measurement of porosity, wet bulk density, and estimation of permeability. A possible loss of solid sample material during the drainage was prevented by thin sheets of glass fiber filters.

C. Experimental procedure

From the total recovered samples of 150-mm length and 150-mm diameter, a slice of about 60-mm length was transferred into the sample chamber. The measurements were carried out at pressure steps about 1 MPa apart until the maximum pressure of 20 MPa was reached. On each new level, the pressure was kept constant by repeated readjustment to let the sample material settle, before the acoustic measurements would then be carried out. This settling process was recorded by repeated readings of the water levels in the measuring cylinders and of the sample length via the travel times of the signals from the distance probes.

Because of the extremely large difference in time scale between natural and laboratory compaction, it is very important to apply a stability criterion to define the state in which the settling rate has decreased to a sufficiently low value. Such a criterion also ensures comparable external conditions for all measurements throughout the experiment. Based on the results of several test runs the situation could be considered stable, if, on a given pressure step, the travel times of the distance probes from two consecutive readings 30 min apart differed by less than 20 nanoseconds. This corresponded to a change in sample length of less than 30 micrometers, leading to a relative change in compressional-wave velocity of less than 0.1%.

D. Measured quantities

1. Porosity

From the initial values of porosity ϕ_0 and density ρ_0 (Table I), assuming a pore water salinity of 35‰ and density $\rho_w = 1024\ \mathrm{kg/m^3}$, the volume contents of the pore water and the sediment grains was calculated from the respective initial sample volume. Because the volume content C_s of the solid phase can be considered constant during compaction, the variation of porosity ϕ with effective pressure p is given by

$$\phi(p) = \frac{C_w(p)}{C_s + C_w(p)}, \tag{1}$$

where $C_w(p)$ is the difference $C_{w0} - C_x(p)$ between initial pore water content C_{w0} and the volume $C_x(p)$ pressed out of the sample at the respective pressure step after settling. This relative change in porosity could be measured at a fairly high accuracy of about 0.1%, because the water levels could be read to about a tenth of a milliliter, corresponding to a few hundredths of a percent of the average water content in the samples.

2. Wet bulk density

Wet bulk density ρ as a function of effective pressure was calculated using Eq. (1) and grain density $\rho_s = (\rho_0 - \rho_w\phi_0)/(1-\phi_0)$, with initial porosity ϕ_0, initial density ρ_0, and pore water density $\rho_w = 1024\ \mathrm{kg/m^3}$, from

$$\rho(p) = \rho_s[1 - \phi(p)] + \rho_w\phi(p). \tag{2}$$

3. Permeability

Knowledge of the change in sample length and the loss of pore water as a function of time and effective pressure makes possible an estimation of permeability on the respective pressure steps. Assuming laminar flow and isotropic sediment structure, the flow velocity v as the volume of water passing through the permeable surface area A per unit of time is[13]

$$v = \frac{q}{A} = \frac{\kappa \Delta p}{\eta \Delta x}, \tag{3}$$

wherein κ is permeability and q, A, Δp, Δx, and η are flow rate, cross-sectional area, difference between pore pressure and atmospheric pressure, sample length, and viscosity of the pore water, respectively.

4. Attenuation

The attenuation of the acoustic waves is expressed throughout this paper in terms of the specific dissipation function $1/Q$, the inverse of the quality factor. This measure can be converted to the attenuation coefficient α, in nepers per meter, by the relation[11]

$$\frac{1}{Q} = \frac{\alpha V}{\pi f}, \tag{4}$$

in which V is velocity and f is frequency. The inverse quality factor was determined by the spectral ratio method[11] using reference spectra from previously recorded signals obtained from measurements on a series of aluminum specimens of various lengths. Figure 2 gives two examples from sample 1 of signals and references as well as the respective spectra and the logarithms of the spectral coefficients. Using the slope m of the straight line fitted to the logarithm of the spectral ratio, the inverse of the quality factor is given by $1/Q = mV/(\pi x)$, where V is the respective velocity and x is sample length.

III. RESULTS AND DISCUSSION

Table II summarizes the results of the pulse transmission experiments on the two deep-sea sediment samples.

A. Physical properties

Porosity (Fig. 3) in both samples is marked by a significant decrease with increasing pressure, with high gradients in the lower pressure range and lower gradients near the maximum pressure of 20 MPa. Sample 1 shows a slightly wider range of variation, $0.673 < \phi < 0.339$, than sample 2 with $0.597 < \phi < 0.358$. Wet bulk density (Fig. 4) as a function of effective pressure reflects the variation of porosity with strong increase in the lower pressure range and maximum values of near $2100\ \mathrm{kg/m^3}$ at 20 MPa pressure. Both porosity and density show the pressure dependence obtained from field studies in deep-sea fine-grained sediment, as shown in Fig. 12 of Ref. 5. Permeability (Fig. 5) in both samples is marked by the low values typical of fine-grained sediments. The pressure-induced compaction causes a decrease in permeability of about 2 orders of magnitude over the total pressure range. The comparatively strong scatter in the data originates from the less-accurate measurement of flow rates rather then volumes in this experiment.

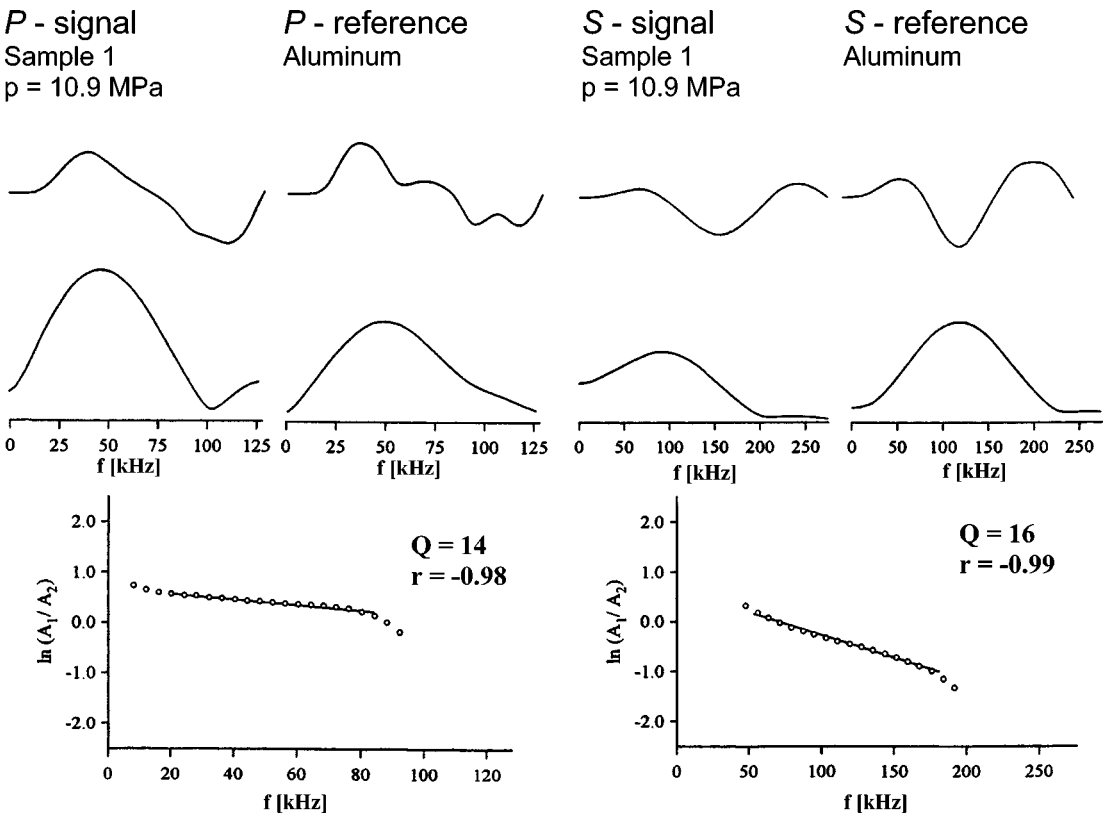


FIG. 2. Example of the received *P*-wave and *S*-wave signals from sample 1 at 10.9 MPa and the respective reference signals from aluminum specimens (top row), corresponding spectra (middle row), and the respective logarithms of the spectral ratios (bottom row).

B. Acoustic properties

The nature of the experiments and of the sample material implies that every measurement could be carried out only once, i.e., no repeated measurements were possible to reduce the experimental error. The error bars in Fig. 6 through Fig. 9 correspond therefore to the estimated error bound, or the estimated maximum absolute error.

1. Velocities

Compressional-wave velocity (Fig. 6) in both samples shows the expected strong increase in the lower pressure range and smaller gradients at higher pressures. Velocities vary from 1683 to 2128 m/s in the silty clay (sample 1) and from 1840 to 2462 m/s in the foraminiferal mud (sample 2) over the entire range of effective pressure. The shear-wave velocity (Fig. 7) could not be measured in the lower pressure range, because unambiguous detection of the shear-wave signals was possible only from critical pressure values of about 4 MPa for sample 1 and about 9 MPa for sample 2, with an *S*-wave velocity range of 450 to 765 m/s and 775 to 975 m/s, respectively. Below these critical pressure values it appears that a low degree of internal coupling in the sample material strongly restricts the propagation of shear waves; yet the comparatively high compressional-wave velocities in this lower pressure range imply that the sediment samples have finite shear rigidity. It may therefore be concluded that the change in sediment structure with pressure, i.e., the denser packing of the sediment particles, supports a more efficient transmission of the shear waves.

For a general discussion of this phenomenon, it may be noted that the significant differences between the two samples regarding the contents in montmorillonite as an expandable clay mineral appear to have an influence on the pressure range from which the shear waves are detectable in the experiment. As the water contents of the samples were determined by weighing before and after evaporation at 105 °C, the porosity values in Table I—which were used as reference values for the experiments—include the intracrystalline water in the montmorillonite. If we divide the total water content into an intraparticle and an extraparticle component, the latter can be considered as the free porosity, i.e., the volume fraction of the free pore water that corresponds to the porosity of a pack of purely solid grains. Applied to the sediment samples investigated here, using the estimation that the volume of the montorillonite approximately doubles with complete intracrystalline swelling,[14] this would yield free-porosity values of 0.36 for sample 1 and 0.38 for sample 2 at the respective critical pressures. The expanded montmorillonite crystallites cause the sediment particles to come in close contact at lower pressures in sample 1 as opposed to sample 2. This implies that the free porosity at the respective critical pressure assumes the significance of a critical porosity according to the concept of Nur *et al.*,[15] at which the sediment structure becomes fully frame supported and which typically lies in the range of 0.36 to 0.40.

2. Attenuation

The results for the compressional-wave attenuation (Fig. 8) are marked by maximum values that occur exactly in

TABLE II. Acoustic and physical properties of the sediment samples from the experiments.

Sample Nr.	Pressure (MPa)	V_P (m/s)	V_S (m/s)	$1/Q_P$	$1/Q_S$	Porosity	Permeability (m^2)	Density (kg/m^3)
Sample 1	0.5	1572	⋯	⋯	⋯	0.546	$3.18E-14$	1767
	1.1	1638	⋯	0.078	⋯	0.507	$1.23E-14$	1831
	2.1	1683	⋯	0.081	⋯	0.47	$4.8E-15$	1892
	3.1	1767	⋯	0.082	⋯	0.45	$4.7E-15$	1924
	4	1777	450	0.082	⋯	0.437	$3E-15$	1946
	5.1	1798	492	0.082	0.061	0.42	$1.7E-15$	1973
	6.2	1851	507	0.084	0.079	0.413	$1.5E-15$	1985
	7.1	1874	535	⋯	0.062	0.404	$1.6E-15$	1999
	8	1923	567	0.08	0.081	0.397	$1.5E-15$	2011
	9	1943	574	0.077	0.082	0.392	$1.3E-15$	2019
	9.9	1972	600	0.075	0.082	0.386	$1E-15$	2029
	10.9	1985	619	0.073	0.063	0.381	$6E-16$	2037
	11.9	2001	641	0.069	0.063	0.377	$6E-16$	2044
	12.6	2017	654	0.065	0.079	0.373	$6E-16$	2050
	13.6	2032	677	0.062	0.077	0.365	$7E-16$	2063
	15.3	2040	690	0.054	0.068	0.359	$5E-16$	2073
	15.8	2070	700	0.051	0.066	0.353	$1.2E-15$	2083
	17.1	2107	724	0.044	0.065	0.348	$4E-16$	2091
	18	2119	742	0.034	0.064	0.345	$5E-16$	2096
	19	2128	756	0.032	0.063	0.343	$4E-16$	2099
	19.8	2128	765	0.031	0.064	0.339	$6E-16$	2106
Sample 2	1	1840	⋯	0.036	⋯	0.501	$4.28E-14$	1848
	2	1876	⋯	0.037	⋯	0.468	$1.45E-14$	1903
	3	1987	⋯	0.041	⋯	0.448	$6.9E-15$	1936
	4	2064	⋯	0.049	⋯	0.426	$7E-15$	1972
	5	2112	⋯	0.052	⋯	0.422	⋯	1979
	6	2119	⋯	0.054	⋯	0.413	$3.8E-15$	1994
	7	2173	⋯	0.061	⋯	0.407	$3.3E-15$	2003
	8	2189	⋯	0.066	⋯	0.402	$3.1E-15$	2012
	9	2214	775	0.064	0.068	0.397	$2.9E-15$	2020
	10	2239	785	0.06	0.077	0.394	$2.5E-15$	2025
	11	2259	799	0.062	0.07	0.387	$4.6E-15$	2036
	12	2274	804	0.062	0.061	0.385	$1.3E-15$	2040
	13	2307	852	0.063	0.074	0.379	$3.7E-15$	2050
	14	2350	865	0.062	0.083	0.376	$2.4E-15$	2055
	15	2375	886	0.062	0.067	0.375	$6E-16$	2056
	16	2400	918	0.06	0.077	0.371	$2.9E-15$	2063
	17	2404	937	0.056	0.077	0.366	$3.4E-15$	2071
	18	2427	950	0.059	0.08	0.363	$1.9E-15$	2076
	19	2446	964	0.059	0.084	0.36	$1.5E-15$	2081
	20	2462	975	0.056	0.083	0.358	$1.5E-15$	2084

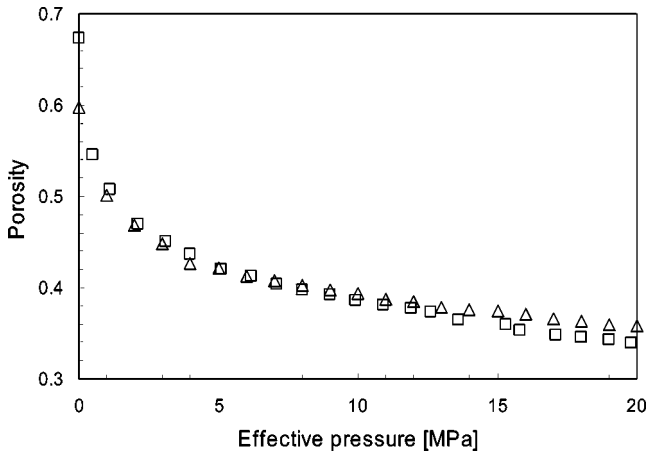


FIG. 3. Porosity of the sediment samples as a function of effective pressure (squares: sample 1; triangles: sample 2).

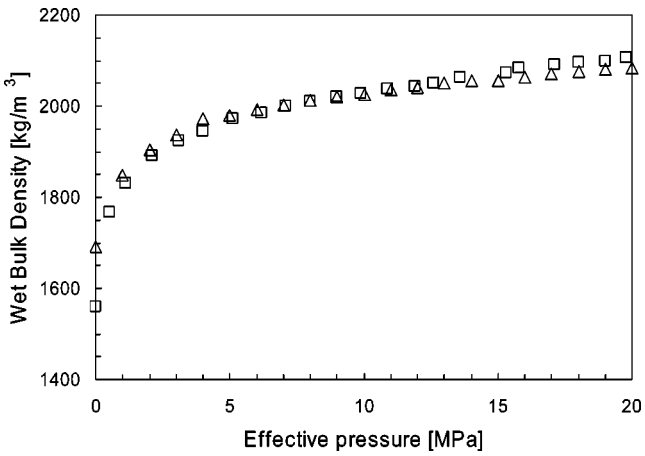


FIG. 4. Wet bulk density of the sediment samples as a function effective pressure (squares: sample 1; triangles: sample 2).

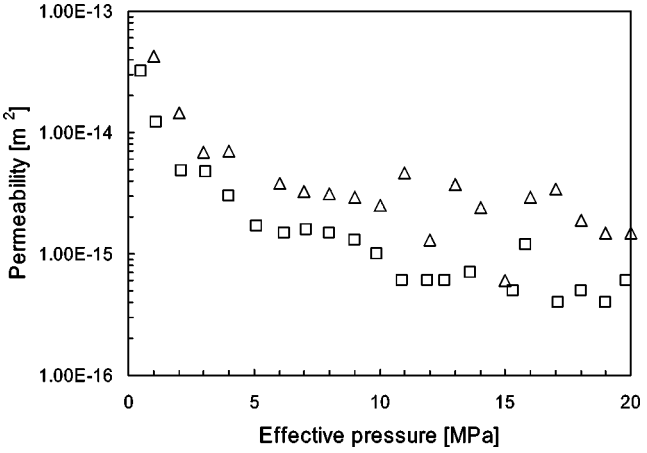


FIG. 5. Permeability of the sediment samples as a function of effective pressure (squares: sample 1; triangles: sample 2).

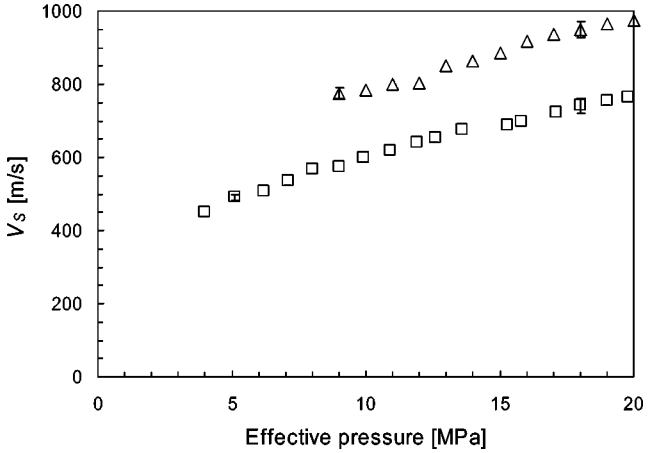


FIG. 7. Shear-wave velocity of the sediment samples as function of effective pressure (squares: sample 1; triangles: sample 2).

those critical pressure ranges in which, for the respective sample, the observation of shear waves becomes possible. The maximum values in $1/Q_P$ are near 0.08 at about 5 MPa in sample 1 and near 0.065 at about 9 MPa in sample 2. The results for the shear-wave attenuation (Fig. 9) do not allow one to decide whether or not such an attenuation maximum also exists for shear waves, as assumed by Hamilton.[16] In spite of the scatter in the data, it can be recognized that the existing attenuation data do not suggest a significant decrease in shear-wave attenuation but rather constant values with increasing compaction. The average values in $1/Q_S$ are 0.07.

The finding of a maximum compressional-wave attenuation as a function of effective pressure under the well-controlled conditions of the experiment may be regarded as a confirmation of the conclusion drawn by Hamilton[9] from his field studies, namely that the occurrence of such a maximum as a function of depth below the sea bottom may be a general phenomenon for fine-grained marine sediments but is not observed in coarse-grained ones. A depth-dependent maximum in compressional-wave attenuation was also found in field data on fine-grained marine sediments by Mitchell and Focke.[10] Although Hamilton[9] has proposed as a qualitative explanation for the attenuation maximum a combined effect of porosity reduction and pressure increase, in neither of the two studies[9,10] have the responsible attenuation mechanisms involved been specified.

C. Critical examination of the results

Because reduction in porosity is by far the most significant process during compaction of a soft, fine-grained sediment, the relationship between porosity and compressional-wave velocity and attenuation, respectively, is used here to examine to what extent the results of the experiment realistically reflect the variation of velocity and attenuation as a function of depth below the sea bottom.

1. Compressional-wave velocity versus porosity

The assumption is made that the difference between laboratory and *in situ* compaction may be considered as sufficiently small, if the relationship between compressional-wave velocity and porosity from the experiment is comparable to that of *in situ* measurements. For a corresponding quantitative test the formula of Nafe and Drake[5] was used, which is given by

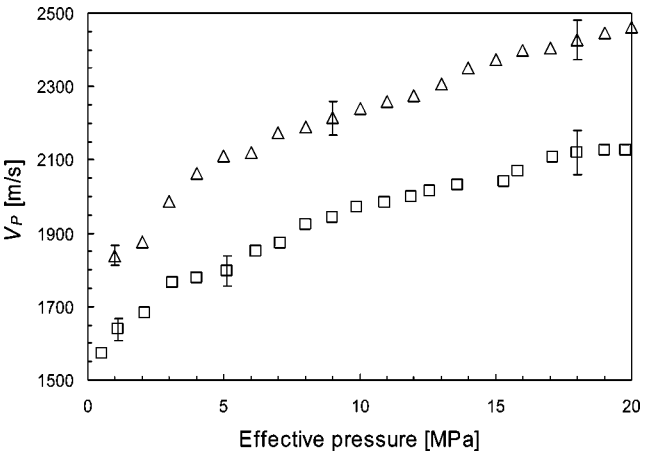


FIG. 6. Compressional wave velocity of the sediment samples as a function of effective pressure (squares: sample 1; triangles: sample 2).

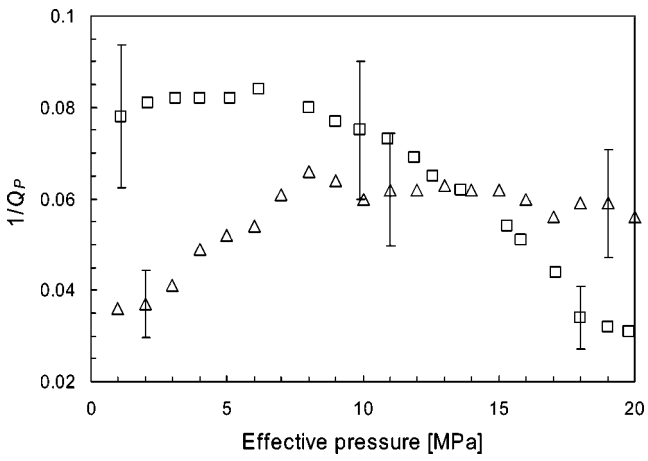


FIG. 8. Attenuation, in terms of specific dissipation, of the compressional waves of the sediment samples as a function of effective pressure (squares: sample 1; triangles: sample 2).

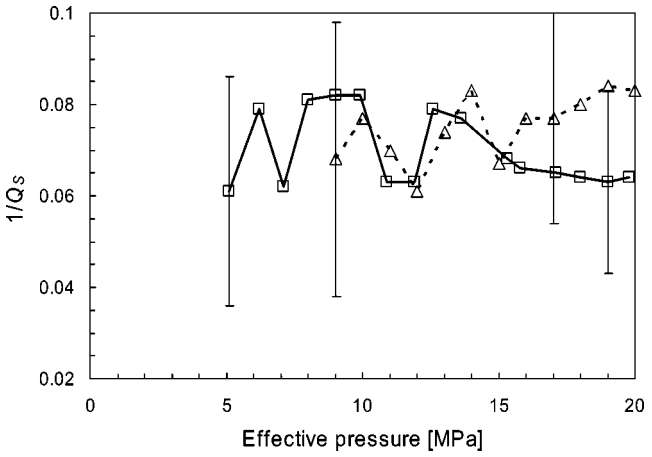


FIG. 9. Attenuation, in terms of specific dissipation, of the shear waves of the sediment samples as a function of effective pressure (squares: sample 1; triangles: sample 2).

$$V_P = \sqrt{\phi V_w^2 \left[1 + \left(\frac{\rho_w}{\rho}\right)(1-\phi)\right] + \left(\frac{\rho_m}{\rho}\right)(1-\phi)^N V_m^2}, \tag{5}$$

where V_P, V_m, V_w are the compressional wave velocities of the sediment and the grains and the sound velocity of the pore water, respectively; ρ, ρ_m, ρ_w are saturated density, grain density, and water density, respectively; ϕ is porosity. The exponent N has the significance of a material constant for a given sediment. Substituting different values for N in Eq. (5), a family of model curves is generated as shown in Fig. 10. A porosity–velocity pair (ϕ, V_P) describing an individual sediment at some stage during compaction will move along a curve corresponding to its individual exponent

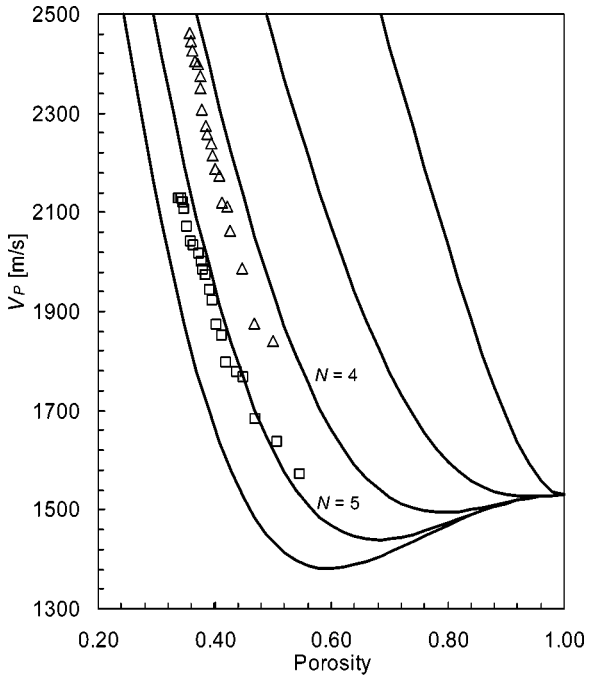


FIG. 10. Compressional-wave velocity as a function of porosity in fine-grained marine sediments. Solid lines: Model curves according to the formula of Nafe and Drake (Ref. 5), Eq. (5), with $N=\{2;3;4;5;6\}$, $V_w = 1530$ m/s, $V_s = 5000$ m/s, $\rho_w = 1024$ kg/m^3, $\rho_s = 2675$ kg/m^3 (squares: sample 1; triangles: sample 2).

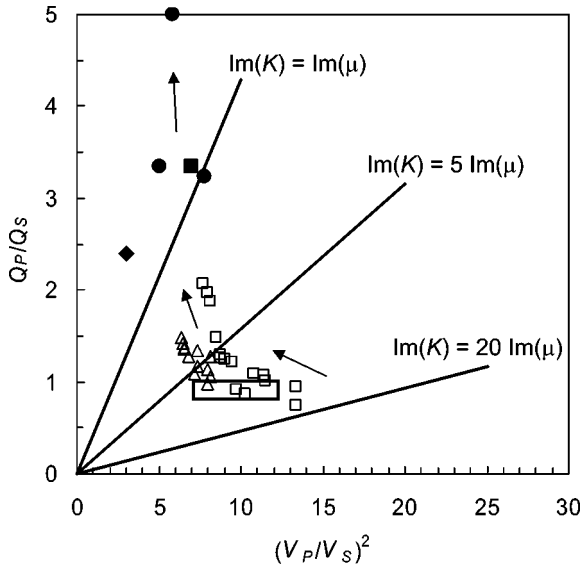


FIG. 11. Diagram for generalized interpretation of sediment acoustics (Ref. 17), modified after Winkler and Nur (Ref. 18), based on Eq. (6). The solid lines correspond to selected constant ratios Im(K)/Im(μ). (Squares: sample 1; triangles: sample 2; filled diamond: Berea sandstone (Ref. 11), $\phi = 0.22$ filled circles: Massilon sandstone, $\phi = 0.22$ (Ref. 18); filled square: Pierre shale (Ref. 19); box: boundaries of a collection of field data (Ref. 17) on saturated sand, clay, gravel. Arrows indicate the direction of increasing effective pressure in samples 1 and 2, and in the data of Winkler and Nur (Ref. 18).

N to lower porosity and higher velocity values. Figure 10 shows that the data for the compressional-wave velocity versus porosity for both samples lie fairly close to parallel to the model curves. With $N = 5.15$ for sample 1 and $N = 4.28$ for sample 2, both samples fit in the empirical range of $N \approx 4$ to $N \approx 5$, within which the majority of (ϕ, V_P) data in fine-grained marine sediments falls.[5]

2. Compressional-wave attenuation versus porosity

For the attenuation results, it is more difficult to conceive of a method to examine the comparability of laboratory and *in situ* data, because there is no simple relationship between attenuation and porosity or depth below the sea bottom. Yet, for a general test of consistency among the data and in comparison with existing results from laboratory and field measurements on the attenuation in saturated sediments, a diagram by Meissner and Theilen,[17] modified after Winkler and Nur,[18] can be used which is based on the equation

$$\frac{Q_P}{Q_S} = \frac{\mathrm{Im}(\mu)}{\mathrm{Im}(K) + \frac{4}{3}\mathrm{Im}(\mu)} \frac{V_P^2}{V_S^2}, \tag{6}$$

in which Im(μ) and Im(K) are the imaginary parts of the shear and bulk moduli, respectively. Within the scope of this diagram, in which the ratio Q_P/Q_S is plotted against $(V_P/V_S)^2$, the experimental data sets of the two sediment samples are consistent among one another and, at medium pressures, fall fairly near or within the boundaries of field data from Meissner and Theilen[17] (Fig. 11). Towards higher pressures and lower porosities, the Q_P/Q_S ratios show a strong increase at only slightly decreasing $(V_P/V_S)^2$ values. This trend points in a direction that is manifested by the

results from other experiments[11,18,19] on saturated sediments at still lower porosities (Fig. 11). This is of course possible only because the attenuation of the compressional waves from the experiments shows a strong decrease as opposed to that of the shear waves and because, at the same time, the V_P/V_S ratios decrease by only less than 10% over the latter third of the total pressure range. Furthermore, the diagram immediately reveals that the losses in compressional loading in our deep-sea sediment samples are dominant at the lower to medium compaction degrees [e.g., $\mathrm{Im}(K) > 10\,\mathrm{Im}(\mu)$ at 5 MPa for sample 1], and that shear losses rapidly gain significance with increasing pressure.

The results from the deep-sea samples in combination with the literature data give rise to the assumption that there may exist a unique pattern according to which the variation of porosity with compaction is mapped on this Q_P/Q_S vs $(V_P/V_S)^2$ representation. This in turn would support the assumption that a maximum in compressional-wave attenuation occurs as a general phenomenon in fine-grained marine sediments at a certain stage of compaction or, equivalently, at a certain depth below the ocean bottom.

IV. CONCLUSION

The experiments described in this paper represent a unique laboratory study on the acoustic properties of natural samples of highly compactible fine-grained marine sediment in that the velocities and attenuation of both compressional and shear waves were measured as a function of effective pressure under simultaneous control of porosity, density, and permeability. This was achieved by measuring one additional quantity, namely the volume of pore water pressed out of the sediment sample as a function of effective pressure.

The most significant result from the experiments is the observation of maximum values as a function of effective pressure of compressional-wave attenuation, which can be considered as a confirmation of earlier conclusions from field data from the literature. The attenuation maxima were consistently found in both samples, at different effective pressures and porosities. These porosity values can be identified as critical-porosity values marking the transition to a fully frame-supported sediment structure.

An important conclusion for underwater acoustics is that the results emphasize that compressional-wave attenuation, in terms of the specific dissipation $1/Q_P$, can be a source of valuable additional information on the mechanical properties of the sediment. While the relationship between porosity and compressional-wave velocity in compacting unconsolidated fine-grained ocean sediments is relatively well understood, but reflects predominantly the composition of the sediment in terms of the mechanical properties of solid and fluid constituents, weighted according to their volume fractions, attenuation reflects the physical processes accompanying the propagation of the compressional waves that in turn are a result of how the solid and fluid components are distributed geometrically in the sediment. Regarding the data presented here, the compressional-wave attenuation—without knowing the actual attenuation mechanism—allows identification of the critical porosity as a very important state during the entire course of compaction.

As the results from two samples represent a relatively narrow basis for general conclusions, further accurate experiments are therefore desirable. Velocity and attenuation measurements in the range of very low pressures would be of particular interest for the investigation of the earliest stages of compaction in the uppermost subsurface layers of the seafloor.

ACKNOWLEDGMENT

This study was supported by the German Research Foundation within the Joint Research Project 313.

[1] A. Nur and Z. Wang, *Geophysics Reprint Series, No. 10, Seismic and Acoustic Waves in Reservoir Rocks*, Vol. 1, Experimental Studies (Soc. Expl. Geophys., Tulsa, OK, 1989), pp. 1–9.
[2] M. N. Toksöz and D. H. Johnston, *Geophysics Reprint Series, No. 2, Seismic Wave Attenuation* (Soc. Expl. Geophys., Tulsa, OK, 1981), pp. 136–139.
[3] H. H. Rieke III and G. V. Chilingarian, *Developments in Sedimentology*, Vol. 16, Compaction of Argillaceous Sediments (Elsevier, Amsterdam, 1974), p. 107.
[4] A. S. Laughton, "Sound propagation in compacted ocean sediments," Geophysics **22**, 233–260 (1957).
[5] J. E. Nafe and C. L. Drake, "Variation with depth in shallow and deep water marine sediments of porosity, density, and the velocities of compressional and shear waves," Geophysics **22**, 523–552 (1957).
[6] A. B. Wood, *A Textbook of Sound*, 2nd ed. (Bell and Sons, London, 1941), p. 361.
[7] M. R. J. Wyllie, A. R. Gregory, and L. W. Gardner, "Elastic wave velocities in heterogeneous and porous media," Geophysics **21**, 41–70 (1956).
[8] E. L. Hamilton, "Compressional-wave attenuation in marine sediments," Geophysics **37**, 620–646 (1972).
[9] E. L. Hamilton, "Sound attenuation as a function of depth in the sea floor," J. Acoust. Soc. Am. **59**, 528–535 (1976).
[10] S. K. Mitchell and K. C. Focke, "New measurements of compressional wave attenuation in deep ocean sediments," J. Acoust. Soc. Am. **67**, 1582–1589 (1980).
[11] M. N. Toksöz, D. H. Johnston, and A. Timur, "Attenuation of seismic waves in dry and saturated rocks. I. Laboratory measurements," Geophysics **44**, 681–690 (1979).
[12] H. Hirschleber, F. Theilen, W. Balzer, B. von Bodungen, and J. Thiede, *Forschungsschiff Meteor*, Reise 7, vom 1. Juni bis 28. September 1988—Berichte der Fahrtleiter, Ber. Sonderforschungsbereich 313, Nr. 10 (University of Kiel, Kiel, 1988), pp. 1–249.
[13] T. Bourbié, O. Coussy, and B. Zinszner, *Acoustics of Porous Media* (Edition Technip, Paris, 1987), p. 31.
[14] F. Kraehenbuehl, H. F. Stoeckli, F. Brunner, G. Kahr, and M. Mueller-Vonmoos, "Study of the water–bentonite system by vapour adsorption, immersion calorimetry, and x-ray techniques. I. Micropore volumes and internal surface areas, following Dubinin's theory," Clay Miner. **22**, 1–9 (1987).
[15] A. Nur, D. Marion, and H. Yin, in *Shear Waves in Marine Sediments*, edited by J. M. Hovem, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht, 1991), Sec. 1, pp. 131–140.
[16] E. L. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[17] R. Meissner and F. Theilen, in *Proceedings of the 11th World Petroleum Congress*, 2 (Wiley, Chichester, 1983), Paper SP 3.
[18] K. W. Winkler and A. Nur, "Seismic attenuation: Effects of pore fluids and frictional sliding," Geophysics **47**, 1–14 (1982).
[19] F. J. McDonal, F. A. Angona, R. L. Mills, R. L. Sengbush, R. G. Van Nostrand, and J. E. White, "Attenuation of shear and compressional waves in Pierre shale," Geophysics **23**, 421–439 (1958).

Asymptotic accuracy of geoacoustic inversions

Michele Zanolin,[a] Ian Ingram, Aaron Thode, and Nicholas C. Makris
Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

(Received 18 April 2003; revised 17 June 2004; accepted 9 July 2004)

Criteria necessary to accurately estimate a set of unknown geoacoustic parameters from remote acoustic measurements are developed in order to aid the design of geoacoustic experiments. The approach is to have estimation error fall within a specified design threshold by adjusting controllable quantities such as experimental sample size or signal-to-noise ratio (SNR). This is done by computing conditions on sample size and SNR necessary for any estimate to have a variance that (1) asymptotically attains the Cramer–Rao lower bound (CRLB) and (2) has a CRLB that falls within the specified design error threshold. Applications to narrow band deterministic signals received with additive noise by vertical and horizontal arrays in typical continental shelf waveguides are explored. For typical low-frequency scenarios, necessary SNRs and samples sizes can often approach prohibitively large values when a few or more important geoacoustic parameters are unknown, making it difficult to attain practical design thresholds for allowable estimation error. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1787526]

PACS numbers: 43.30.Pc [WLS] Pages: 2031–2042

I. INTRODUCTION

Geoacoustic parameters of the ocean floor strongly affect sound propagation and acoustic sensing in shallow water ocean waveguides where extensive bottom interaction occurs.[1–3] A significant amount of work has been done in recent years to develop methods for estimating geoacoustic parameters and to benchmark these methods against simulated noiseless data as for example in Refs. 2 and 4–10. Much less work, however, has been done to assess the performance of geoacoustic inversions in the presence of noise.[2,11–15]

Nonlinear inversions are often required to estimate geoacoustic parameters from measured acoustic field data. Since the measured data undergo random fluctuations due to additive noise, waveguide scintillation, or source randomness, this nonlinearity often leads to estimates that are biased and exceed the Cramer–Rao lower bound (CRLB) by orders of magnitude. In these situations, exact expressions for the bias and the variance are often difficult or impractical to derive analytically.

Knowing both the CRLB and how to attain it is useful for a number of practical reasons. The mean-square error of any unbiased estimate of a deterministic parameter vector from random data cannot be less than the CRLB, which exists given mild regularity conditions on the probability density of the data.[16] This is true *regardless of the method of estimation*, and, for example, regardless of whether or not there are significant ambiguities, sometimes referred to as *sidelobes* in the estimation problem.

Parameter estimates only have practical value if their errors fall within the design thresholds specified for the given experiment. In the inversion of geoacoustic parameters, for example, design errors are often set by the needs of those who run propagation and scattering models to evaluate sonar system performance. If the CRLB for a particular experiment is always greater than the specified design error threshold, the experiment will never be able to achieve its goals and will necessarily fail. So the CRLB on its own is extremely useful as a tool in aiding experimental design in these situations. If the CRLB is less than or equal to the allowable design error, on the other hand, the practicality of the experimental design is still questionable until it is established that the parameter estimates derived from this experiment actually attain the CRLB.

Since necessary conditions for an estimate to attain the CRLB are now available and depend on controllable variables of an experiment such as signal-to-noise ratio (SNR) or sample size,[17] and the CRLB is also a function of these controllable variables, conditions are then also available to attain any specified design error. This can be done by proper adjustment of the controllable variables.

Along these lines, we follow the general estimation theory approach introduced in Ref. 17 and use it to derive conditions to accurately estimate a set of unknown geoacoustic parameters from remote acoustic field measurements. We do this by computing necessary SNRs and sample sizes for the estimates to become asymptotically unbiased, for their mean-square errors to attain the CRLB, and then for the CRLB to fall within any specified design criteria.

We note that the approach of Ref. 17 is a general consequence of estimation theory and so can be and has already been applied to obtain optimality conditions and to extract new physical insights in a number of widely divergent and physically unrelated estimation problems. These include time-delay and Doppler shift estimation,[17] source localization in an ocean waveguide,[18] and pattern recognition in 2-D images,[19] where an *optimal estimate* in this context is defined as being unbiased and having minimum variance following standard practice.[20] A basic advantage of this approach is that it is typically straightforward to implement and provides analytical insight into the mechanics of asymptotic optimality and consequently attainable accuracy for the given estima-

[a]Electronic mail: zanolin@mit.edu

tion problem. Brute force numerical calculation of estimator moments does not easily offer such insight, but is the only alternative currently available.

Our present analysis focuses on aiding experimental design by determining necessary SNRs and sample sizes to attain practical accuracies in estimating geoacoustic parameters of the seafloor from standard ocean-acoustic inverse experiments. We consider narrow-band deterministic signals received with additive ambient noise by both vertical and horizontal arrays in continental shelf waveguides. Given the large number of unknown environmental parameters in such problems, it is common practice to invert for tens or more parameters simultaneously.[1,21–23] Various combinations of geoacoustic parameters for simultaneous inversion are considered and criteria necessary for accurate inversions are presented. The conditions are found to become significantly more stringent, sometimes to the point of being prohibitive, as the number of unknown parameters to which the measured field is sensitive increases.

In Sec. II, conditions necessary for asymptotic optimality are summarized in a more explicit form than has previously appeared, and a far more condensed and efficient form of the asymptotic variance is also provided. An explicit explanation of how these necessary conditions may be used to achieve design specifications for error thresholds in a given experiment also appears in Sec. II. Analysis of illustrative problems in geoacoustic inversion appear in Sec. III. Since the data are modeled as deterministic signals measured with random ambient noise, we have not investigated the effects of model mismatch or uncertainty in sensor location, both of which may also lead to significant errors. These effects, however, will only make the necessary conditions more stringent.

II. NECESSARY CONDITIONS FOR ASYMPTOTICALLY OPTIMAL ESTIMATION AND FOR ATTAINING SPECIFIED ERROR DESIGN THRESHOLDS

Consider a set of n independent and identically distributed experimental data vectors $\mathbf{X}_i$ of dimension N obeying the probability density $p(\mathbf{X};\boldsymbol{\theta})$, where $\mathbf{X}=[\mathbf{X}_1^T,\ldots,\mathbf{X}_n^T]^T$ and $\boldsymbol{\theta}$ is an m-dimensional parameter vector. The MLE $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ maximizes the log-likelihood function $l(\mathbf{X};\boldsymbol{\theta})=\ln(p(\mathbf{X};\boldsymbol{\theta}))$ with respect to the components of $\boldsymbol{\theta}$. If the rth component of $\boldsymbol{\theta}$ is denoted by θ^r, the first log-likelihood derivative with respect to θ^r is then defined as $l_r=\partial l(\boldsymbol{\theta})/\partial\theta^r$. The elements of the expected information matrix, known as the Fisher matrix, are then given by $i_{ab}=E[l_a l_b]$, and the elements of its inverse by $i^{ab}=[\mathbf{i}^{-1}]_{ab}$, where $\mathbf{i}^{-1}$ is also known as the CRLB.

The moments of $\hat{\theta}^r$ for $r=1,...,m$ can be expressed as a series in inverse powers of the sample size n,[17,18] provided that the required derivatives of the likelihood function exist.[24] The variance can then be expressed as

$$var(\theta^r)=\frac{var_1(\theta^r)}{n}+\frac{var_2(\theta^r)}{n^2}+O\left(\frac{1}{n^3}\right), \tag{1}$$

where $O(1/n^3)$ represents integer powers higher than $1/n^2$ and $var_1(\theta^r)$ and $var_2(\theta^r)$ depend only on a single sample probability distribution. The first term on the right-hand side is the CRLB, which is the minimum variance for an unbiased estimate and also the asymptotic value of the variance in the limit as the sample size n and SNR approach infinity.

The sample size necessary for the MLE variance to asymptotically attain the CRLB is found by requiring the second-order variance to be negligible compared to the first

$$\frac{|var_2(\theta^r)/n^2|}{var_1(\theta^r)/n}\ll 1, \tag{2}$$

which implies

$$n\gg\frac{|var_2(\theta^r)|}{var_1(\theta^r)}. \tag{3}$$

Only for sample sizes satisfying this condition is it possible for the variance to be in the asymptotic regime where it continuously attains the CRLB. This follows from the fact that each term in the expansion is proportional to a unique power in $1/n$.

In a similar manner, a necessary sample size for the inversion to be asymptotically unbiased is found by requiring that the first-order bias is negligible compared to the true value of the parameter:

$$n\gg\frac{|b_1(\theta^r)|}{|\theta^r|}. \tag{4}$$

The conditions (3) and (4) provide insight into the performance of any estimate in the limit of large sample size or SNR. In fact, in this regime any estimate that satisfies these conditions must be the MLE.[20]

As noted in the Introduction, parameter estimates only have practical value if their errors fall within the design threshold specified for the given experiment. In order to attain a specified design error threshold by the present approach, the sample size n must be large enough that (I) optimality conditions (3) and (4) are satisfied and (II) the CRLB falls within the required design error threshold.

A. Statistical model for the acoustic data

We consider the field generated by a deterministic narrow band source that is received by an array of hydrophones with additive stationary ambient noise. One vector sample in the frequency domain of the measured field can be obtained from the Fourier transform of a time window of the acoustic measurements. Statistical independence of the samples requires them to have a sample spacing that is at least the coherence time of the total received field.[25] Explicitly, the jth spectral data sample $\tilde{\mathbf{X}}_j(\omega;\boldsymbol{\theta})$ for $j=1,...,n$ is given by

$$\tilde{\mathbf{X}}_j(\omega;\boldsymbol{\theta})=A(\omega)\tilde{\mathbf{g}}(\omega)+\tilde{\boldsymbol{\eta}}_j(\omega), \tag{5}$$

where $A(\omega)$ is the Fourier transform of the source amplitude, $\tilde{\mathbf{g}}(\omega)=[\tilde{g}_1(\omega;\boldsymbol{\theta}),\ldots,\tilde{g}_N(\omega;\boldsymbol{\theta})]$ is the vector of Green's functions in the frequency domain connecting the source location to the N hydrophone locations on the array, and $\tilde{\boldsymbol{\eta}}_j(\omega)=[\tilde{\eta}_{j1}(\omega),\ldots,\tilde{\eta}_{jN}(\omega)]$ is the noise spectral sample which is given by a Fourier transform of a finite time window of the noise.

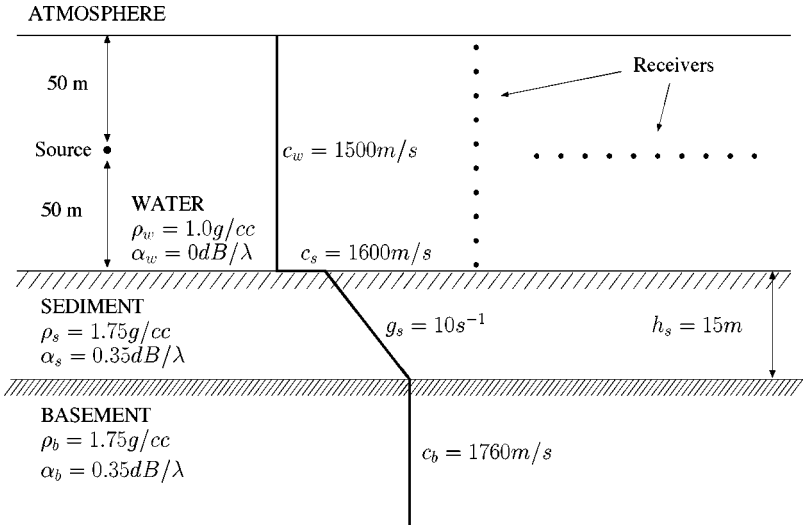


FIG. 1. Waveguide model and experimental setup. An isovelocity water column overlies a two-layer bottom: a 15-m-thick fluid sediment layer with a sound speed linearly increasing with depth stands above a basement with constant sound speed and density. A narrow band point source is located at the center of the wave guide and receiving arrays. A ten-element vertical array and horizontal arrays with 10 and 100 elements are considered. The spacing between the elements is 7.5 m and the arrays are centered in the water column.

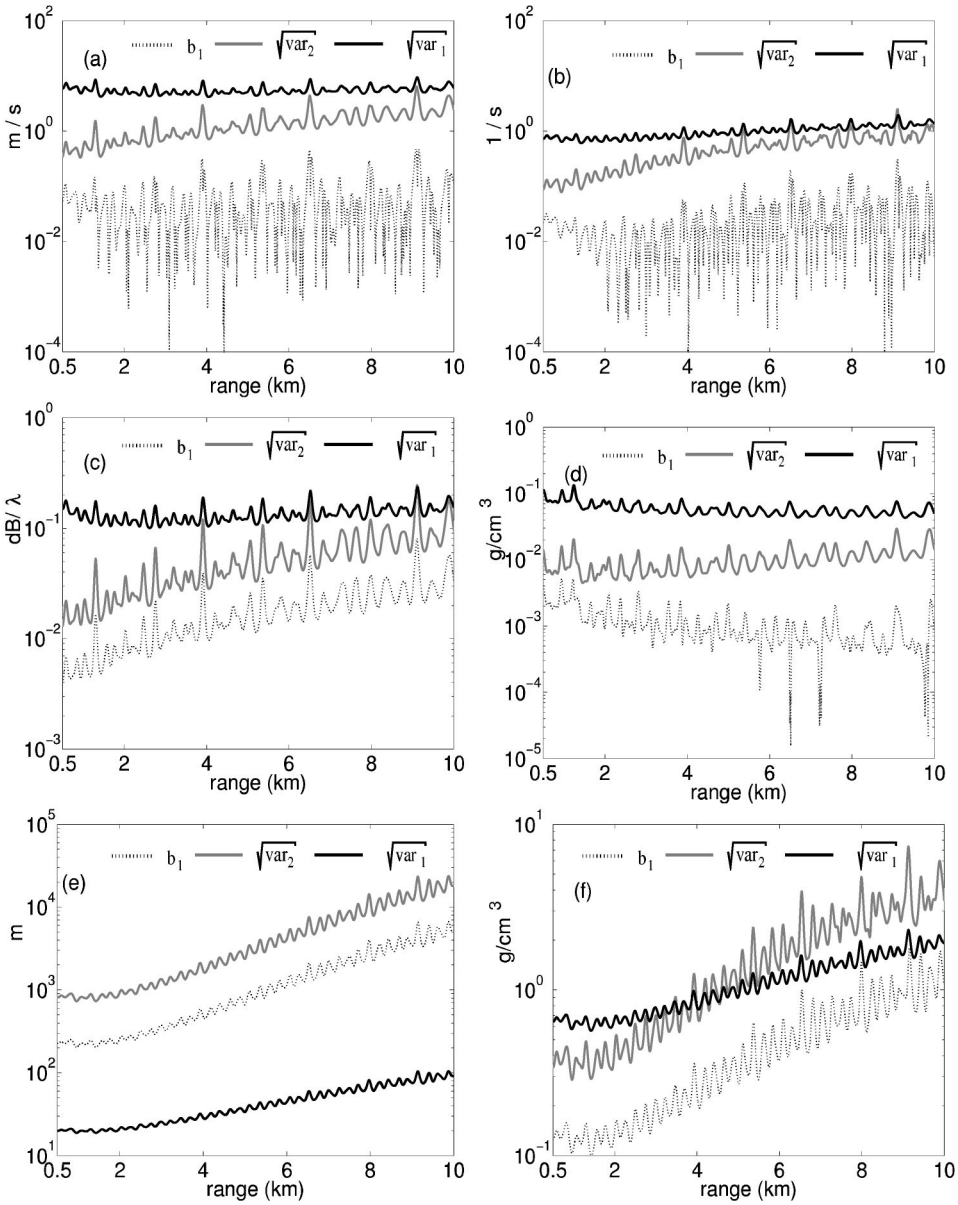


FIG. 2. $\sqrt{var_1}$ (black), $\sqrt{|var_2|}$ (gray), and b_1 (dotted) for single parameter estimates of c_s, g_s, α_s, ρ_s, h_s, and ρ_b are presented for $n=1$ as a function of range between 0.5 and 10 km for a 100-Hz source and ten-element vertical array centered at mid-depth in the watercolumn.

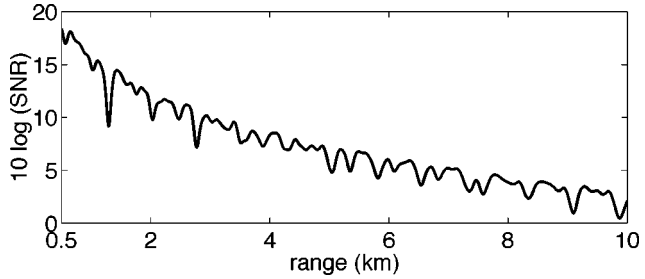


FIG. 3. SNR as functions of range between 0.5 and 10 km. Same experimental setup as Fig. 2.

The noise spectrum is well described by a circular complex Gaussian random variable (CCGR),[25,26] so that the probability density for the real measured data $\mathbf{X}_j$ with $j=1,...,n$ becomes

$$p(\mathbf{X},\boldsymbol{\theta})=(2\pi)^{-nN/2}|\mathbf{C}|^{-n/2} \times\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\mathbf{X}_j-\boldsymbol{\mu}(\boldsymbol{\theta}))^T\mathbf{C}^{-1}(\mathbf{X}_j-\boldsymbol{\mu}(\boldsymbol{\theta}))\right\}, \tag{6}$$

where $\mathbf{X}_j$ and $\boldsymbol{\mu}(\boldsymbol{\theta})$ are specified by

$$\mathbf{X}_j=\begin{bmatrix}\mathrm{Re}(\tilde{\mathbf{X}}_j(\omega;\boldsymbol{\theta}))\\ \mathrm{Im}(\tilde{\mathbf{X}}_j(\omega;\boldsymbol{\theta}))\end{bmatrix},\quad \boldsymbol{\mu}(\boldsymbol{\theta})=\begin{bmatrix}\mathrm{Re}(A(\omega)\tilde{\mathbf{g}}(\omega))\\ \mathrm{Im}(A(\omega)\tilde{\mathbf{g}}(\omega))\end{bmatrix}, \tag{7}$$

with Re(.) and Im(.) indicating the real and imaginary parts. The real covariance matrix

$$\mathbf{C}=\frac{1}{2}\begin{pmatrix}\mathrm{Re}(\tilde{\mathbf{C}}) & -\mathrm{Im}(\tilde{\mathbf{C}})\\ \mathrm{Im}(\tilde{\mathbf{C}}) & \mathrm{Re}(\tilde{\mathbf{C}})\end{pmatrix} \tag{8}$$

is specified by the spectral complex covariance matrix of the noise across the array $\tilde{\mathbf{C}}$ whose elements are given by $\tilde{\mathbf{C}}_{ln}=E[\eta_{jl}(\omega)\eta_{jn}^{*}(\omega)]=\sigma^2\delta_{ln}$, with δ_{ln} equal to 1 for $l=n$ and 0 for $l\neq n$. Note that the expectation eliminates the dependence on the sample index j. Here we assume spatially uncorrelated noise for both the horizontal and vertical apertures based on our experience with experimental data in shallow water environments. An alternative would be to use theoretical predictions based on uniformly distributed surface noise sources such as in Ref. 27.

In the present formulation, while the measured field contains parameter information, the sufficient statistic for optimal estimation in a measurement is not the measured field or its ensemble average from measured data but the entire argument of the exponential, known as the Mahalobinos distance.[28] This preserves all the relevant intersensor phase information as the ensemble average of a positive semidefinite quantity.

For this statistical model, the expressions given in Ref. 17 for the numerators of the first-order bias and the first two orders of the variance can be expressed in the much more compact form

$$b_1(\theta^r)=-\tfrac{1}{2}i^{ra}i^{bc}\boldsymbol{\gamma}_{ac}^{T}\boldsymbol{\gamma}_b, \tag{9}$$

$$var_1(\theta^i)=-i^{ii}, \tag{10}$$

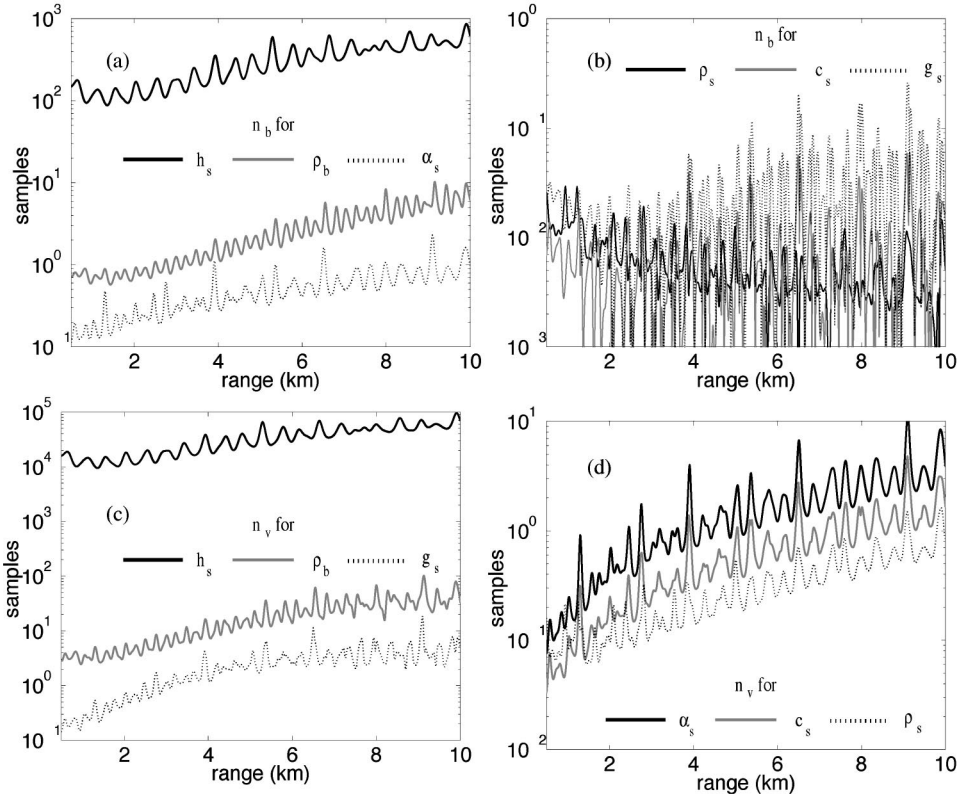


FIG. 4. n_b and n_v as functions of range between 0.5 and 10 km for single parameter. (a) n_b for h_s (black), ρ_b (gray), and α_s (dotted). (b) n_b for ρ_s (black), c_s (gray), and g_s (dotted). (c) n_v for h_s (black), ρ_b (gray), and ρ_b (dotted). (d) n_v for ρ_s (black), c_s (gray), and g_s (dotted). Same experimental setup as Fig. 2.

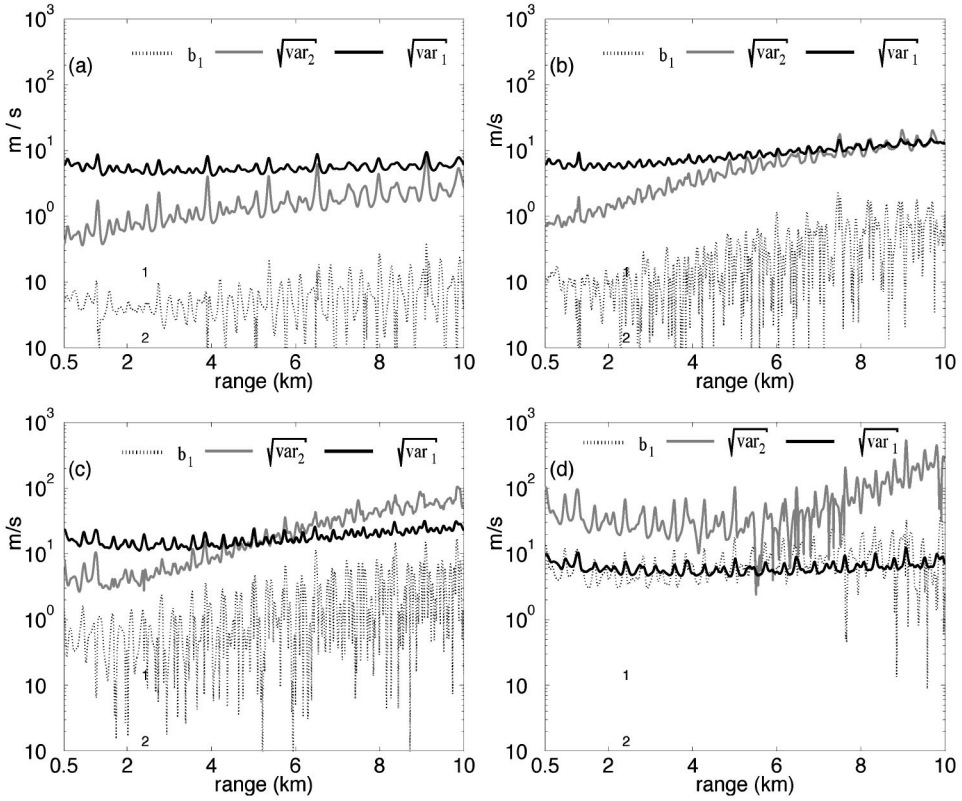


FIG. 5. $\sqrt{var_1}$ (black), $\sqrt{|var_2|}$ (gray), and b_1 (dotted) are shown as a function of range for $n=1$ between 0.5 and 10 km in inversions for c_s with one other unknown parameter. The unknown is successively (a) α_s, (b) ρ_s, (c) g_s, and (d) h_s. Same experimental setup as Fig. 2.

$$var_2(\theta^i) = i^{im} i^{in} i^{pq} \Bigg(\boldsymbol{\gamma}_{nq}^T \boldsymbol{\gamma}_{pm} - \boldsymbol{\gamma}_{mn}^T \boldsymbol{\gamma}_{pq} - \boldsymbol{\gamma}_{mpq}^T \boldsymbol{\gamma}_n + i^{zt} \Bigg(\frac{\boldsymbol{\gamma}_n^T \boldsymbol{\gamma}_{tq}}{2} \boldsymbol{\gamma}_m^T \boldsymbol{\gamma}_{zp} + (2\boldsymbol{\gamma}_{qt}^T \boldsymbol{\gamma}_n - \boldsymbol{\gamma}_{qn}^T \boldsymbol{\gamma}_t) \boldsymbol{\gamma}_{pm}^T \boldsymbol{\gamma}_z + (\boldsymbol{\gamma}_{mn}^T \boldsymbol{\gamma}_p + \boldsymbol{\gamma}_{mp}^T \boldsymbol{\gamma}_n) \boldsymbol{\gamma}_{zt}^T \boldsymbol{\gamma}_q \Bigg) \Bigg), \tag{11}$$

where $\boldsymbol{\gamma}_{c\cdots d} = [A(\omega)/\sigma]\mathbf{g}_{c\cdots d}$ and the subscripts $c\cdots d$ indicate that derivatives of the Green's function with respect to the parameters $\theta_c \cdots \theta_d$ have been taken. The Einstein summation convention is used so that if an index occurs twice in a term, once in the subscript and once in the superscript, summation over the index is implied.

The SNR for a single sample collected across the array

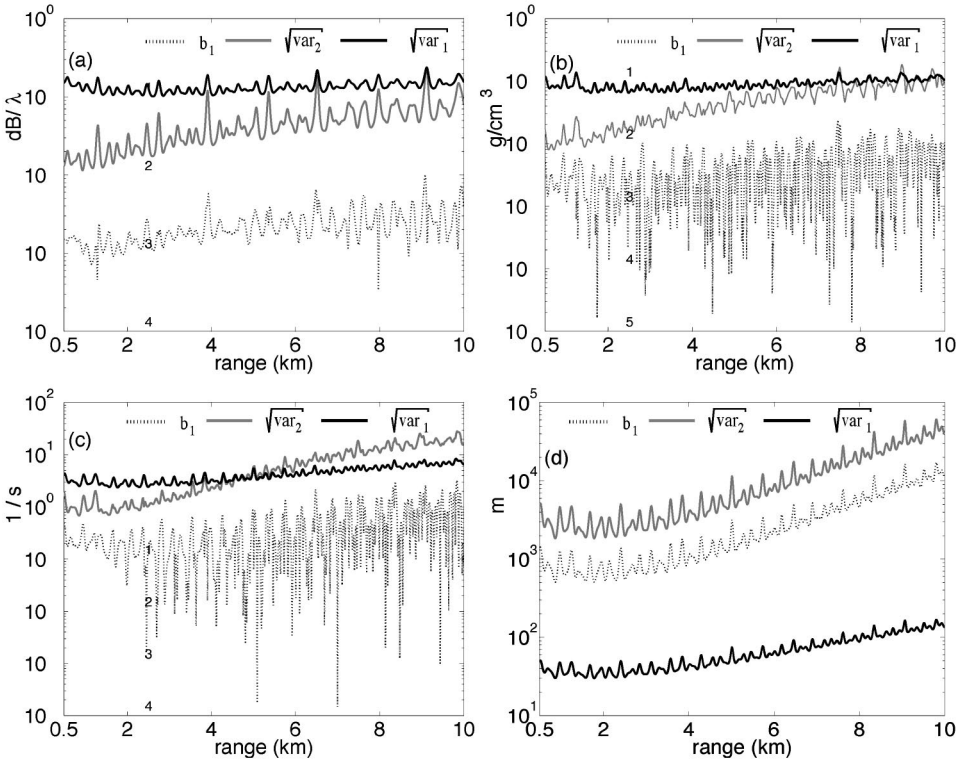


FIG. 6. $\sqrt{var_1}$ (black), $\sqrt{|var_2|}$ (gray), and b_1 (dotted) are shown as a function of range for $n=1$ between 0.5 and 10 km for successive two-parameter estimates of (a) α_s, (b) ρ_s, (c) g_s, and (d) h_s with c_s. Same experimental setup as Fig. 2.

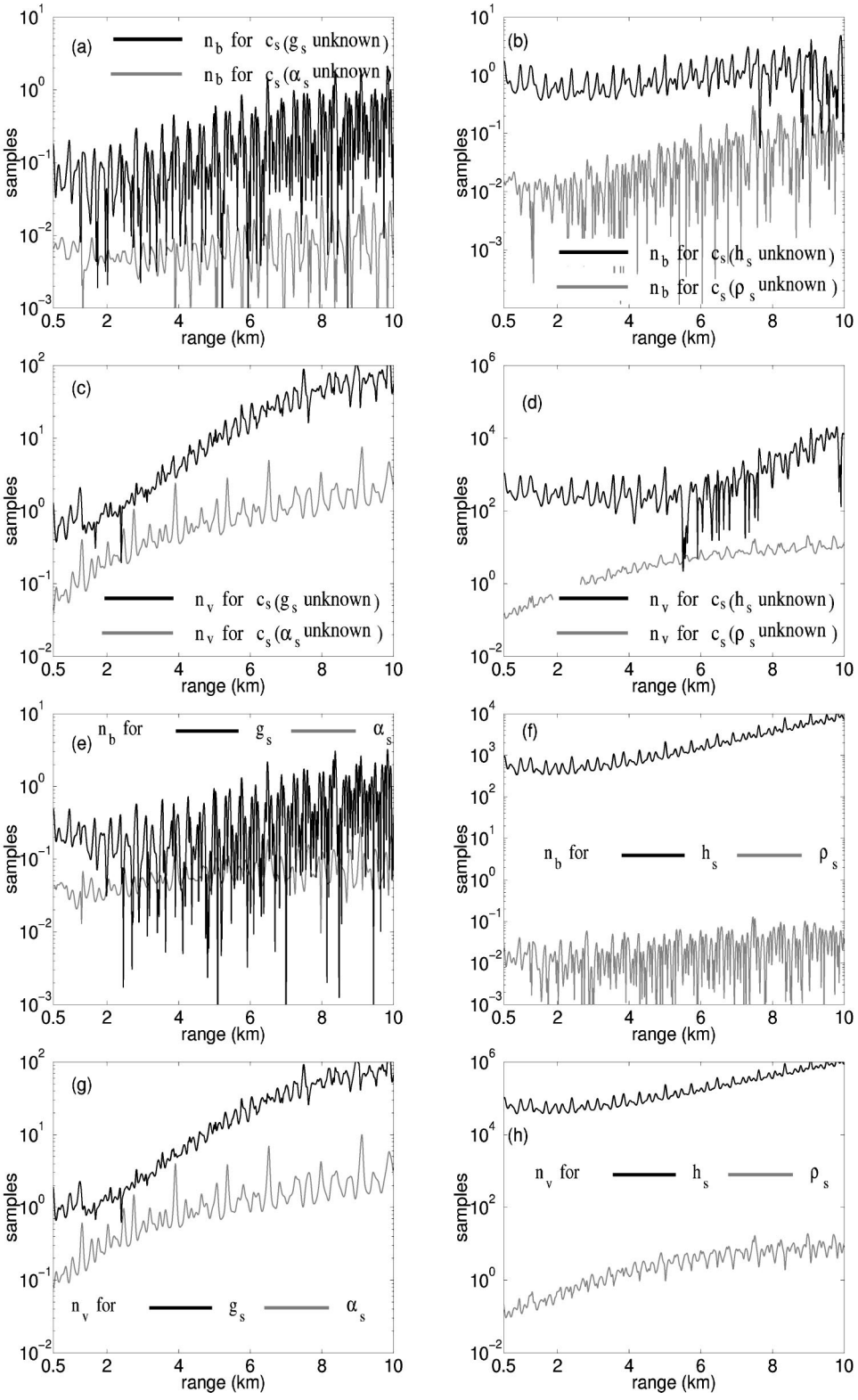


FIG. 7. For the estimation of c_s when a second parameter is unknown, n_b is presented when (a) α_s (gray) and g_s (black) are unknown (b) ρ_s (gray) and h_s (black) are unknown and n_v is presented when (c) α_s (gray) and g_s (black) are unknown and (d) ρ_s (gray) and h_s (black) are unknown. For the estimation of a sediment parameter when a c_s is unknown, n_b is presented for (e) α_s (gray) and g_s (black) and (f) ρ_s (gray) and h_s (black), and n_v is presented for (g) α_s (gray) and g_s (black) and (h) ρ_s (gray) and h_s (black).

has been defined as the ratio $\mathrm{SNR}=\boldsymbol{\mu}(\omega;\boldsymbol{\theta})\boldsymbol{\mu}(\omega;\boldsymbol{\theta})^*/\mathrm{tr}(\mathbf{C})=|\mathrm{A}(\omega)|^2\mathbf{g}(\omega;\boldsymbol{\theta})\mathbf{g}(\omega;\boldsymbol{\theta})^*/N\sigma^2$. In most geoacoustic inversion experiments performed in shallow water the SNR varies between 10 and 20 dB,[1,29,30] sometimes reaching values between 30 and 40 dB.[31] In the examples presented in this paper the SNR is set to 15 dB at a range of 1 km from the source, or, equivalently, the variance of the noise is fixed by

$$\sigma^2=\frac{1}{N}\left.\frac{|A(\omega)|^2\mathbf{g}(\omega;\boldsymbol{\theta})\mathbf{g}(\omega;\boldsymbol{\theta})^*}{10^{1.5}}\right|_{\text{range}=1\ \text{km}}. \tag{12}$$

III. ILLUSTRATIVE EXAMPLES

The conditions necessary to obtain an optimal parameter estimate in a given experimental scenario depend on a num-

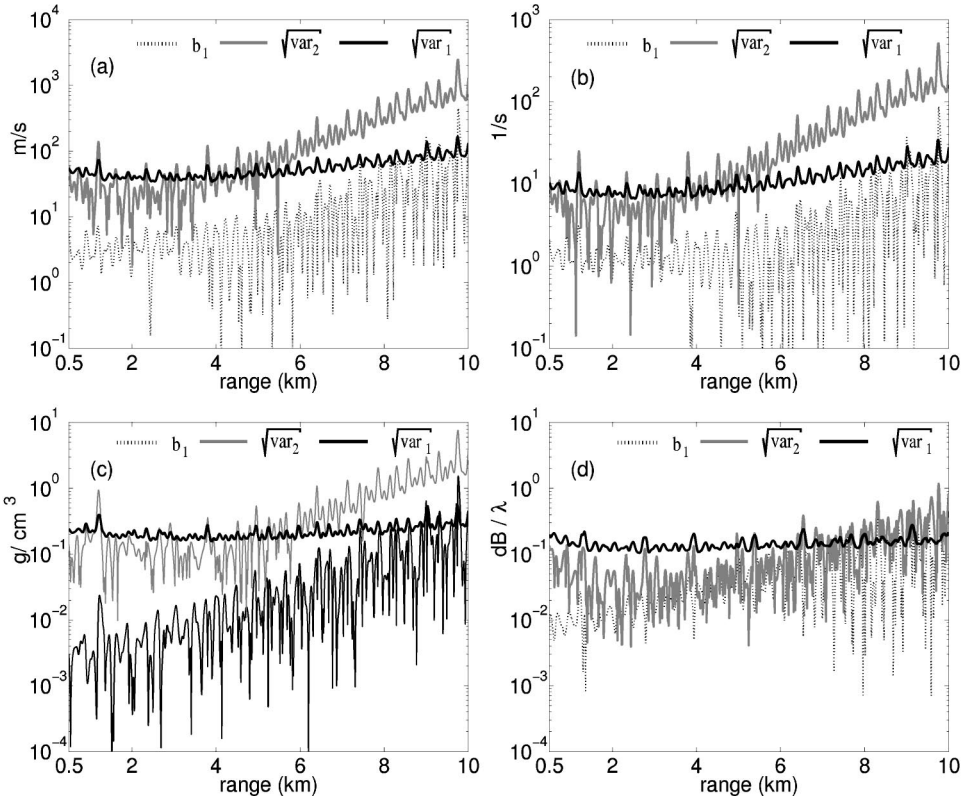


FIG. 8. Simultaneous four-parameter estimation of c_s, g_s, ρ_s, and α_s where $\sqrt{var_1}$ (black), $\sqrt{|var_2|}$ (gray), and b_1 (dotted) are presented for (a) c_s, (b) g_s, (c) ρ_s, and (d) α_s for $n=1$ as a function of range between 0.5 and 10 km. Same experimental setup as Fig. 2.

ber of variables, including the parameters involved in the inversion, the number of parameters simultaneously estimated, the frequency of the source, the range of the receivers, and the SNR. In order to isolate and illustrate these contributions, a number of simulations are performed in a waveguide representative of the continental shelf where a sediment layer overlays a bottom half-space, as shown in Fig. 1 using a modal formulation for the field as in Ref. 18. The numerical field derivatives approach used was benchmarked analytically in a Pekeris waveguide.[32] Field derivatives were also checked with three independent propagation codes including OASIS, SNAP, and a modified version of

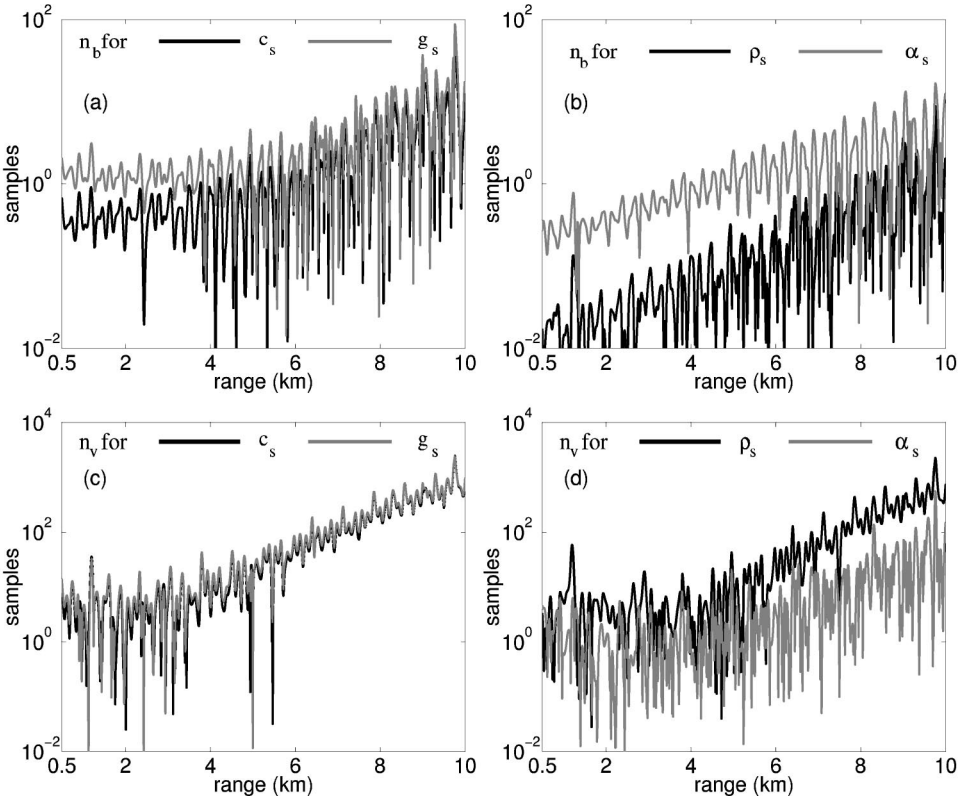


FIG. 9. Source frequency is 100 Hz. Necessary sample sizes for the simultaneous four-parameter estimation of c_s, g_s, ρ_s, and α_s: (a) n_b for c_s (black) and g_s (gray), (b) n_b for ρ_s (black) and α_s (gray), (c) n_v for c_s (black) and g_s (gray), and (d) n_v for ρ_s (black) and α_s (gray). Same experimental setup as Fig. 2.

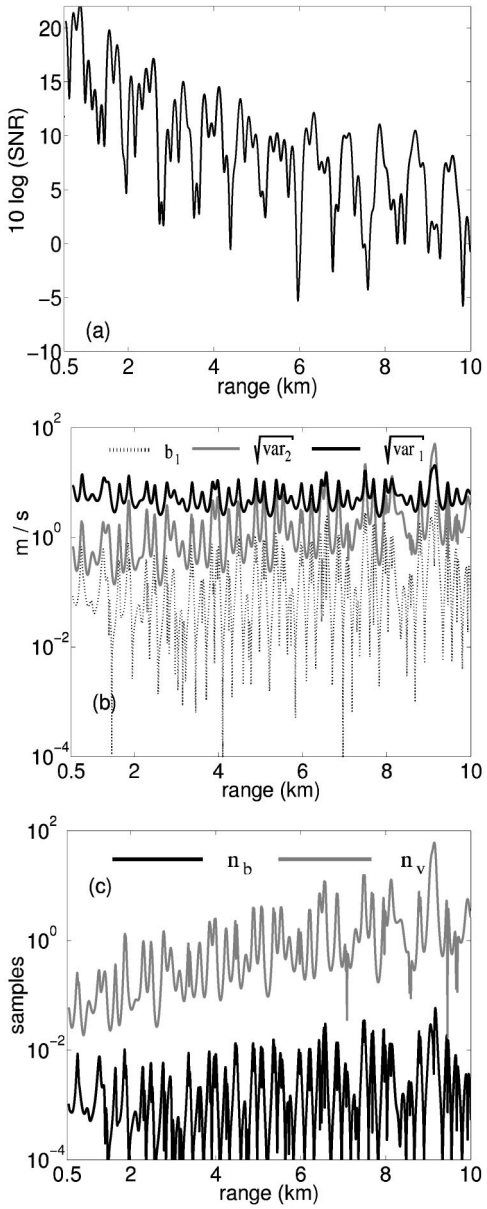


FIG. 10. Single parameter inversion of c_s using a ten-element horizontal array with 7.5-m spacing. The array is located at 50-m depth with 100-Hz source frequency. Shown for $n=1$ as a function of range between 0.5 and 10 km are (a) SNR, (b) $\sqrt{var_1}$ (black), $\sqrt{|var_2|}$ (gray), and b_1 (dotted), and (c) n_b (black) and n_v (gray).

KRAKEN. The sound speed profile in the sediment can be specified in terms of c_s and g_s as $c(z)=c_s+g_s(z-H)$, where the z axis originates at the water–atmosphere interface and is directed vertically downward.

To represent a typical experiment, in Secs. III A and III B a ten-element vertical array is centered in a water column of depth $H=100$ m with 7.5-m spacing between each element so that the shallowest element is at 16.25-m depth. The source is placed at 50-m depth. In this paper a 100-Hz deterministic monopole source is employed. Inversions performed with horizontal arrays are presented in Sec. III C to investigate the effect of array length and orientation on inversion performance.

The necessary sample sizes for the variance to attain the CRLB are computed by conservatively requiring in condition (3) that the second-order variance is ten times smaller than the CRLB for all parameters,

$$n \geqslant n_v = 10\frac{|var_2(\theta^r)|}{var_1(\theta^r)}. \tag{13}$$

Similarly, the necessary sample sizes for the inversions to be unbiased are computed by requiring in condition (4) that the first-order bias be ten times smaller than the true value of the parameter, or $n>n_b=10|b_1(\theta_j)|/|\theta_j|$ except for sound speeds where $n>n_b=200|b_1(\theta_j)|/|\theta_j|$ is used instead since these biases strongly affect the acoustic field. The conditions for an inversion to be optimal are then given by $n>n_b$ and $n>n_v$. If the computed values of n_v and n_b are less than unity, then only one sample is required and the figures can be used to determine how far the SNR can be lowered without sacrificing single-sample optimality. We especially note scenarios where n_b and n_v are large but the corresponding CRLB is small and vice-versa.

It should be noted that the illustrative examples can be used to determine SNR outside of the ranges explicitly shown due to the equivalence of n and SNR in the asymptotic expansions. For example, this means that the conditions (3) and (4) can be reformulated in terms of SNR, and that n_v and n_b are proportional to 1/SNR.

A. Single-parameter inversions

Here we investigate the requirements for estimation errors to attain specified design thresholds for single-parameter inversions. To do this we compute the sample sizes necessary for inversion optimality as well as the magnitude of the CRLB for a single sample. It is important to note that the former optimality condition need not be related to the parameter sensitivity expressed by the single-sample CRLB. This is because the optimality conditions involve higher order parameter derivatives than the CRLB.

The biases, variances, and necessary sample sizes n_v and n_b are computed as a function of source-receiver range for all eight single-parameter estimates allowable in the model. For our purposes only six of these need to be presented in Figs. 2 and 4. These are the thickness of the sediment layer h_s, the compressional wave speed at the top of the sediment layer c_s, the gradient of the compressional wave speed g_s, the attenuation in the sediment α_s, the sediment density ρ_s, and the basement density ρ_b.

The decreasing trend in inversion accuracy with range for all parameters is mostly due to the decrease in SNR shown in Fig. 3 from both spreading and attenuation loss. Stripping of higher order modes with range also plays a role in the decreased accuracy. Estimates of c_s, ρ_s, α_s, and g_s require smaller sample size to be optimal than the basement density ρ_b, and significantly smaller sample size than the thickness of the sediment layer h_s which has particularly stringent optimality conditions. Hundreds of samples are necessary for the h_s estimate to be unbiased even at relatively close ranges and thousands of samples are necessary for the variance to attain the CRLB indicating that the sediment layer thickness h_s has a highly nonlinear relationship with the acoustic measurements.

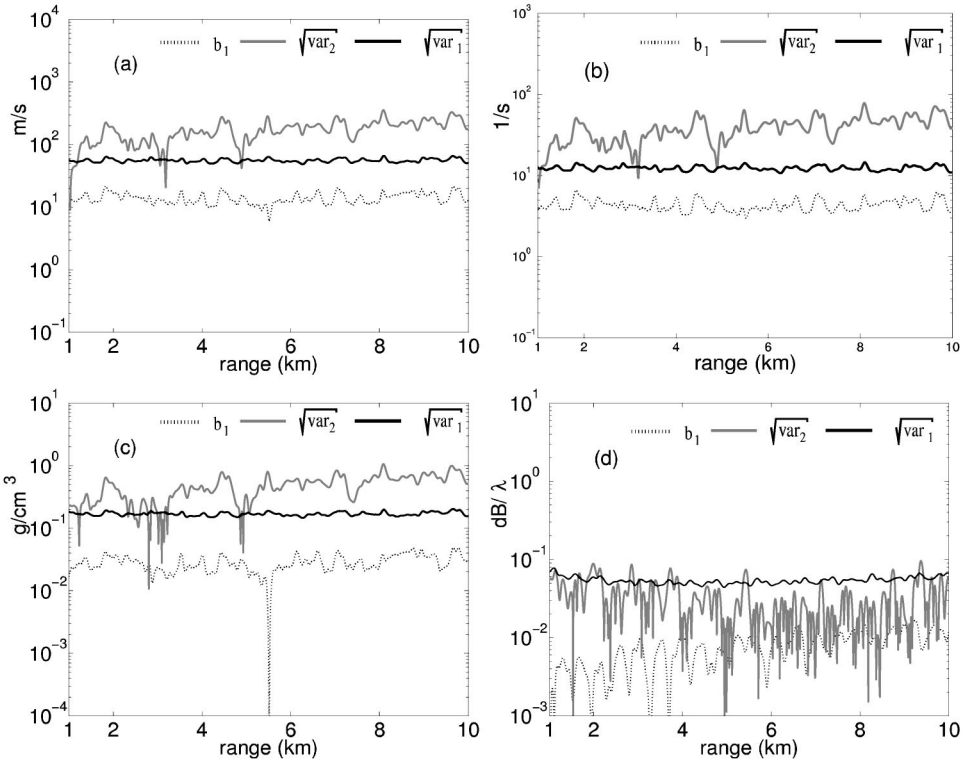


FIG. 11. Simultaneous four-parameter estimation of c_s, g_s, ρ_s, and α_s using a 100-element horizontal array with 7.5-m spacing. The array is located at 50-m depth and the source frequency is 100 Hz. $\sqrt{var1}$ (black), $\sqrt{|var_2|}$ (gray), and b_1 (dotted) are shown for $n=1$ as a function of range between 1 and 10 km for (a) c_s, (b) g_s, (c) ρ_s, and (d) α_s.

We note that while ρ_s and α_s have similar optimality conditions as c_s, the ratio of the square-root of the single-sample CRLB, which is inversly related to the sensitivity of the measurement to the parameter, to the true parameter value is on the order of at least 0.1 for ρ_s and α_s but is less than 0.01 for c_s. This highlights why the *two requirements* explicitly stated in Sec. II are necessary for a parameter estimate to attain a specified error threshold and knowledge of the CRLB alone is not enough.

The inversion of h_s, ρ_b, and the other basement parameters not explicitly presented here are significantly more difficult than the other sediment parameters because they require either prohibitlvely large sample sizes to attain optimality or because the square root of their single-sample CRLBs are large compared to the true parameter value. Sound in the water column is apparently less sensitive to basement parameters due to attenuation in the sediment for the given sediment thickness and acoustic frequency. Similar observations about this lack of sensitivity have been noted in Ref. 20 solely through CRLB analysis.

At lower frequency, penetration into the basement may be more substantial, but there may also be fewer modes. This could lead to difficulties in unambiguously inverting large parameter sets. The modal structure of the acoustic field, for example, imposes limitations on the number of bottom parameters of the given model that can be unambiguously determined with a single frequency source. To illustrate the situation, consider receivers in the water column of a Pekeris waveguide. Each mode is then described by four parameters, the real and imaginary components of the vertical wave number and of the mode's equivalent plane wave amplitude since the up- and downgoing plane wave amplitudes are the negative of each other in this case. This means that the effect of bottom properties on the acoustic field can only be expressed through $4M$ parameters, where M is the number of modes, making $4M$ an upper limit on the number of bottom parameters that can be unambiguously estimated regardless of the number of receivers in the water column. Such limitations can be potentially overcome by increasing the bandwidth.

B. Multiparameter inversions

Simulations presented in this section show that estimation performance worsens as the number of parameters simultaneously inverted increases. To see this, the quantities b_1, var_1, var_2, n_b, and n_v are plotted as a function of source–receiver range for the simultaneous estimation of two parameters, namely c_s together successively with α_s, ρ_s, g_s, and then h_s in Figs. 5–7. Each pairing affects the estimation of c_s in different ways as can be seen in Fig. 5. In fact, estimation of c_s is effectively uncoupled from that of α_s because the two-parameter estimates yield results nearly identical to those of the corresponding single parameter estimates. This can be seen by comparing the moments in Figs. 5(a) and 6(a) with the corresponding ones in Fig. 2(a), and the necessary sample sizes in Fig. 7 with the corresponding ones in Fig. 4.

The optimality conditions for an estimate of c_s, however, do become far more stringent when the estimate is made simultaneously with either the sediment density ρ_s, gradient g_s, or thickness h_s. This is consistent with intuition since c_s, ρ_s, g_s and h_s are expected to be statistically coupled since they are physically coupled in a nonlinear way through the bottom reflection coefficient and through a modal or wave number representation of the acoustic field. It is also reasonable that α_s and c_s be statistically uncoupled since the attenuation α_s leads to very slow decay in the field while the sediment sound speed c_s affects coherent modal

propagation and interference that varies far more rapidly over range (this follows because the two parameters appear in separate factors in the modal representation of the waveguide green funciton). It is interesting that thousands of samples are necessary for the variance of c_s to attain the CRLB when the sediment thickness h_s is also an unknown as can be seen by comparing Figs. 5(d) and 7(b) with Figs. 2 and 4. Simultaneous inversion for the sediment layer thickness h_s in these examples tends to induce extremely stringent optimality conditions, such as prohibitively large necessary sample sizes. This implies that sediment thickness and sediment sound speed are highly coupled for the given scenario where sediment thickness equals the acoustic wavelength. This is sensible since as the sediment thickness varies from the wavelength scale in a decreasing manner, for example, the acoustic field will become less sensitive to sediment sound speed. The couplings described in this paragraph are not apparent if only the CRLB is considered, as shown in Fig. 5

The trend of more stringent optimality conditions continues as the number of parameters to be simultaneously estimated is increased. This is shown for the four-parameter simultaneous inversion of c_s, ρ_s, g_s, and α_s in Figs. 8 and 9. The biases, variance terms, and necessary samples sizes are consistently higher than in the cases where the parameters are either inverted alone or with only one other parameter. We find the trend can become less strignent for the estimation of upper sediment layer parameters as the source frequency is increased, but the opposite is typically true for deeper parts of the bottom.

C. Horizontal array versus vertical array

Parameter estimates made from horizontal array measurements are now examined to investigate the effect of array length and orientation on inversion performance. The moments of a c_s estimate from a horizontal array of the same length and center depth as the vertical array of the previous examples are shown in Fig. 10. No improvement is found in the trend but much larger fluctuations appear upon comparing these moments with those for the vertical array in Figs. 2 and 4.

The horizontal array has much poorer angular resolution than the vertical array at the shallow horizontal grazing

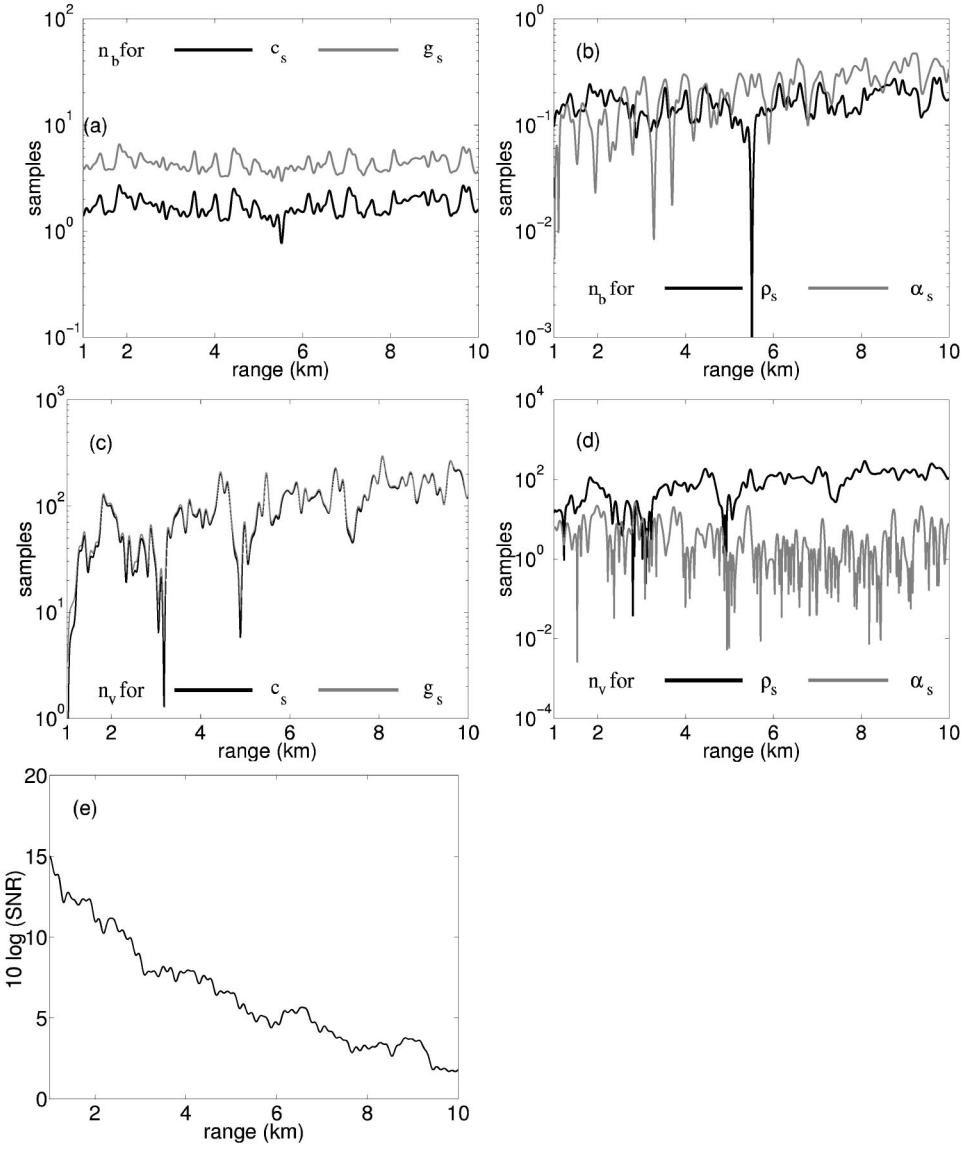


FIG. 12. Simultaneous four-parameter n_b for (a) c_s (black) and g_s (gray), (b) ρ_s (black) and α_s (gray) n_v, (c) c_s (black) and g_s (gray), (d) ρ_s (black) and α_s (gray), and (e) SNR.

angles where the dominant modes typically propagate. This makes it more susceptible to range-dependent fluctuations in SNR arising from the interference of unresolved modes. While the vertical array can resolve the shallow-angle modes with broadside angular resolution of λ/L, the horizontal array receives them at or near end-fire where the angular resolution is only $\sqrt{2\lambda/L}$.

A horizontal array of length $10L$, which can be obtained by synthetic aperture measurements, for example, would be required to have the same angular resolution for shallow grazing angles at 100 Hz as the vertical array of length L. Although this increase in array length greatly reduces the fluctuations in SNR seen in the shorter horizontal array, as shown in Fig. 12(c), it does not provide much improvement in the average trend of the MLE moments. This is illustrated for example in Figs. 11 and 12, where the results for an inversion involving c_s, ρ_s, g_s, and α_s are presented.

Comparing the performance of the long horizontal array in Figs. 11 and 12 with that of the vertical array in Figs. 8 and 9, only the inversion of the bottom attenuation shows some mild improvement. In summary, relatively short vertical arrays can outperform much longer horizontal arrays due to their higher resolving power at shallow grazing angles where dominant modes tend to propagate.

Even if the angular resolution of the receiving array is sufficient to resolve the important modes, further limitations on parameter estimates may emerge depending on the modal content of the field, as discussed previously.

IV. CONCLUSIONS

A reliable method to help attain specified accuracies in the estimation of unknown geoacoustic parameters from remote acoustic measurements is developed to aid the design of geoacoustic experiments. The approach is to compute sample sizes or SNRs necessary for estimates to (1) have variances that asymptotically attain the CRLB and (2) have CRLBs that fall within a specified design error threshold. We show both analytically and with illustrative examples that the former asymptotic condition need not be related to the parameter sensitivity expressed by the CRLB. This is because it involves parameter derivatives of higher order than the CRLB.

Applications to narrow band deterministic signals received with additive noise by vertical and horizontal arrays in typical continental shelf waveguides are explored. For typical low frequency scenarios, necessary SNRs and samples sizes often approach prohibitively large values when a few or more important geoacoustic parameters are unknown, making it difficult to attain practical design thresholds on allowable estimation error. This is found to arise because of the highly nonlinear nature of the geo-acoustic inverse problem and the strong coupling found between many of the important geo-acoustic parameters needed to characterize the acoustic field in an ocean waveguide.

[1] S. D. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water," J. Acoust. Soc. Am. **82**, 998–1017 (1987).

[2] M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean-bottom properties," J. Acoust. Soc. Am. **92**, 2770–2783 (1992).

[3] N. R. Chapman and C. E. Lindsay, "Matched Field inversion for geoacoustic model parameters in shallow water," IEEE J. Ocean. Eng. **21**(4), 347–354 (1996).

[4] S. D. Rajan, "Waveform inversion for the geoacoustic parameters of the ocean bottom," J. Acoust. Soc. Am. **91**, 3228–3241 (1992).

[5] S. E. Dosso, M. L. Yeremey, J. M. Ozard, and N. R. Chapman, "Estimation of ocean-bottom properties by matched field inversions of acoustic field data," IEEE J. Ocean. Eng. **18**(3), 232–239 (1993).

[6] C. E. Linsday and N. R. Chapman, "Matched Field Inversion for geoacoustic model parameters using adaptive simulated annealing," IEEE J. Ocean. Eng. **18**(3), 224–231 (1993).

[7] M. I. Taroudakis and M. G. Markaki, "Bottom geoacoustic inversion by matched field processing- a sensitivity study," Inverse Probl. **16**(8), 1679–1692 (2000).

[8] M. R. Fallat and S. E. Dosso, "Geoacoustic inversions via local, global and hybrid algorithms," J. Acoust. Soc. Am. **105**, 3219–3230 (1999).

[9] D. P. Knobles, R. A. Koch, E. K. Westwood, and T. Udagawa, "The inversion of Ocean Waveguide Parameters using a Non linear least square approach," J. Comput. Acoust. **6**(1&2), 83–97 (1998).

[10] M. I. Taroudakis and M. G. Markaki, "Bottom geoacoustic inversion by broad band Matched Field Processing," J. Comput. Acoust. **6**(1&2), 167–183 (1998).

[11] S. E. Dosso and M. J. Wilmut, "Quantifying data information content in geo acoustic inversions," IEEE J. Ocean. Eng. **27**(2), 296–304 (2002).

[12] H. Schmidt and A. Baggeroer, "Physics-imposed resolution and robustness issues in seismo acoustic parameter inversion," in *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. Diaschok *et al.* (Kluwer Academic, Dordrecht, 1995), pp. 85–90.

[13] P. Daly and A. Baggeroer, "Cramer-Rao bounds for shallow water environmental parameter estimation," *OCEANS 97*, MTS IEEE Conference Proceedings (1997), Vol. 1, pp. 430–435.

[14] V. V. Borodin and G. R. Minasian, "Statistical approach to ocean acoustic tomography, Cramer-Rao Lower Bounds for accuracy of sound-speed field reconstruction," in *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. Diaschok *et al.* (Kluwer Academic, Dordrecht, 1995), pp. 91–95.

[15] S. E. Dosso and P. L. Nielsen, "Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow water data," J. Acoust. Soc. Am. **111**, 143 (2002).

[16] C. R. Rao, *Linear Statistical Inference and its Applications* (John Wiley & Sons, NY, 1973) p. 326.

[17] E. Naftali and N. C. Makris, "Necessary conditions for a maximum likelihood estimate to become asymptotically unbiased and attain the Cramer-Rao Lower Bound. I. General approach with an application to time-delay and Doppler shift estimation," J. Acoust. Soc. Am. **110**, 1917–1930 (2001).

[18] A. Thode, M. Zanolin, E. Naftali, I. Ingram, P. Ratilal, and N. Makris, "Necessary conditions for a maximum likelihood estimate to become asymptotically unbiased and attain the Cramer-Rao lower bound. II. Range and depth localization of a sound source in an ocean waveguide," J. Acoust. Soc. Am. **112**, 1890–1910 (2002).

[19] M. Betke and N. Makris, "Recognition, Resolution and Complexity of Objects Subject to Affine Transformation," Int. J. Comput. Vis. **44**(1), 5–40 (2001).

[20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice–Hall, Englewood Cliffs, NJ, 1993), p. 19.

[21] D. P. Knobles, R. A. Koch, L. A. Thompson, K. C. Focke, and P. E. Eisman, "Broadband sound propagation in shallow water and geoacoustic inversion," J. Acoust. Soc. Am. **113**, 205–222 (2003).

[22] C. Siedenburg, N. Lehtomaki, J. Arvelo, K. Rao, and H. Schmidt, "Iterative Full Field Inversion Using Simulated Annealing," in *Full Field Inversion Methods in Ocean and Seismo-Acoustics*, edited by O. Diaschok *et al.* (Kluwer Academic, Dordrecht, 1995), pp. 121–126.

[23] S. D. Rajan, J. A. Doutt, and W. M. Carey, "Inversion for the compressional Wave Speed Profile of the Bottom from Synthetic aperture Experiments Conducted in the Hudson Canyon Area," IEEE J. Ocean. Eng. **23**(3), 174–187 (1998).

[24] L. R. Shenton and K. O. Bowman, *Maximum Likelihood Estimation in Small Samples* (Griffin, New York, 1977), p. 7.

[25] N. C. Makris, "The effect of saturated transmission scintillation on ocean acoustic intensity measurements," J. Acoust. Soc. Am. **100**, 769–783 (1996).

[26] N. C. Makris, ''Parameter resolution bounds that depend on sample size,'' J. Acoust. Soc. Am. **99**(5), 2851–2861 (1996).
[27] W. A. Kuperman and F. Ingenito, ''Spatial correlation of surface generated noise in a startified ocean,'' J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[28] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (Wiley, New York, 1971), p. 75.
[29] M. Siderius and P. L. Nielsen, ''Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array,'' J. Acoust. Soc. Am. **112**, 1523–1535 (2002).
[30] P. Ratilal, P. Gerstoft, and J. T. Goh, ''Subspace approach to inversion by genetic algorithms involving multiple frequencies,'' J. Comput. Acoust. **6**(1&2), 99–115 (1998).
[31] S. E. Dosso and P. L. Nielsen, ''Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow water data,'' J. Acoust. Soc. Am. **111**, 143–158 (2002).
[32] I. Ingram, ''Necessary conditions for Geo-Acoustic Parameter Inversions to become Asymptotically Unbiased and Attain the Cramer-Rao Lower Bound,'' Master thesis, Massachusetts Institute of Technology, 2002.

Bayesian model selection applied to self-noise geoacoustic inversion

David J. Battle,[a] Peter Gerstoft, William S. Hodgkiss, and W. A. Kuperman
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

Peter L. Nielsen
NATO Undersea Research Centre, 19138 La Spezia, Italy

(Received 15 March 2004; revised 25 June 2004; accepted 28 June 2004)

Self-noise geoacoustic inversion involves the estimation of bottom parameters such as sound speeds and densities by analyzing towed-array signals whose origin is the tow platform itself. As well as forming inputs to more detailed assessments of seabed geology, these parameters enable performance predictions for sonar systems operating in shallow-water environments. In this paper, Gibbs sampling is used to obtain joint and marginal posterior probability distributions for seabed parameters. The advantages of viewing parameter estimation problems from such a probabilistic perspective include better quantified uncertainties for inverted parameters as well as the ability to compute Bayesian evidence for a range of competing geoacoustic models in order to judge which model explains the data most efficiently.
[DOI: 10.1121/1.1785671]

PACS numbers: 43.30.Pc, 43.60.Pt [AIT] Pages: 2043–2056

I. INTRODUCTION

In self-noise geoacoustic inversion, plant and hydrodynamic noise generated by the tow ship as a by-product of its normal operation is used to interrogate the ocean environment. Because of its inherent mobility, reduced complexity, and low environmental impact, self-noise inversion using a towed array is a very promising modality of geoacoustic exploration.[1]

In our previous paper,[2] data acquired during the joint NATO/Marine Physical Laboratory experiment—MAPEX2000—was analyzed using maximum-likelihood (ML) methods to evaluate the feasibility of self-noise inversion from towed-array data. The major conclusion drawn from this preliminary work was that matched-field processing (MFP), in conjunction with global search procedures such as genetic algorithms (GA), was sufficiently sensitive in the near field to permit robust first-order inversion of parameters such as *p*-wave velocity for a range-independent bottom environment known to be reasonably well characterized as a fluid half-space. This was despite low to moderate signal-to-noise ratios (SNR) during the experiment and considerable uncertainty in relation to several important geometric parameters, such as water depth, source range, and array shape.

In this paper, we again direct our attention to the near-field inversion problem using towed-array data, applying a different paradigm to its solution and addressing two important questions: First, can we better quantify the sensitivity limits of near-field inversion? Second, is there a consistent way of ranking our success in modeling an unknown environment using multiple parametrizations? To address the first question, we focus on characterizing the posterior probability densities (PPDs) associated with ensembles of parameter samples generated by a procedure known as Gibbs sampling. Assuming that the model is correct, the PPD then summarizes our complete state of knowledge about the estimated parameters including their mean (expected) values, maximum *a posteriori* (MAP) values, and variances. In answer to the second question, we find as a consequence of knowing the PPD that Bayesian probability theory embodies a natural ranking for competing models, known as *evidence*. In Sec. II, we briefly reiterate a few salient aspects of Bayesian probability theory relevant to the current analysis. For completeness, we restate Bayes' rule and underline the interpretation of likelihoods, priors, and evidence in the context of inverse problems. In Sec. III, we describe in some detail our implementation of the Gibbs sampler used to obtain the results in later sections. While our development parallels that of Dosso[3–5] as recently applied to the analysis of synthetic and experimental geoacoustic data, we offer our own insights into the workings of the algorithm and some further suggestions for improving its efficiency. In Sec. IV we revisit the MAPEX2000 experiment and various aspects of the signal processing and modeling requirements of near-field inversion. Whereas in Ref. 2 we concentrated on demonstrating inversion consistency throughout an extensive portion of the dataset, in Sec. V we focus on the *methodology* of Bayesian model selection—analyzing a single 10-s frame and deriving PPDs corresponding to a variety of geoacoustic models with increasing levels of complexity. In addition to presenting marginal distributions for the inverted parameters, we evaluate the Bayesian evidence for each model using both analytic Gaussian approximations and a more accurate method known as reverse importance sampling. Our conclusions are given in Sec. VI.

[a]Now with the Ocean Engineering Dept., Massachusetts Institute of Technology, Cambridge, MA. Electronic mail: dbattle@mit.edu

II. BAYESIAN INFERENCE

The single aspect that most popularly distinguishes Bayesian probability and its application in inverse problems—Bayesian *inference*—from its so-called *orthodox* or *frequentist* alternatives is the use of prior (before-data) probabilities to modulate posterior (after-data) probabilities according to Bayes' rule

$$p(\mathbf{m}|\mathbf{d},\mathcal{M}_k)=\frac{p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)p(\mathbf{m}|\mathcal{M}_k)}{p(\mathbf{d}|\mathcal{M}_k)}, \tag{1}$$

where $\mathbf{m}$ is a vector of model parameters $m_i\,(i=1,...,P)$, $\mathbf{d}$ is a vector of measured data $d_n\,(n=1,...,N)$, and the *models* $\mathcal{M}_k\,(k=1,...,M)$ embody various parametrizations considered plausible in explaining $\mathbf{d}$. $p(a|b)$ denotes conditional probability, meaning the probability of some outcome a given a previous outcome b. In words, Eq. (1) reads

$$\text{posterior}=\frac{\text{likelihood}\times\text{prior}}{\text{evidence}}.$$

Specializing to the geoacoustic problem at hand, we regard the m_i as representing acoustic quantities that we wish to infer from data $\mathbf{d}_f$ acquired at discrete frequencies $\omega_f\,(f=1,...,F)$ from an array of N hydrophones. The likelihood $p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)$ quantifies the error in matching the $\mathbf{d}_f$ with *replica* fields generated by a continuous wave (cw) propagation model, while the prior $p(\mathbf{m}|\mathcal{M}_k)$ defines and weights plausible ranges for the parameters m_i. The evidence appearing in the denominator can, at first, be thought of as an overall normalization for the PPD, which is independent of the parameter vector $\mathbf{m}$.

A. First level of inference

As defined by MacKay,[6] the first of the two levels of Bayesian inference is concerned with parameter estimation problems in which each proposed model is assumed to be correct, i.e., all conceivable possibilities are encompassed within its prior parameter space. This assumption is responsible for the aforementioned normalization

$$\int p(\mathbf{m}|\mathbf{d},\mathcal{M}_k)d\mathbf{m}=1. \tag{2}$$

Aside from this technicality, it is the *shape* of the PPD, and hence only the likelihood and prior terms, that influence the first level of inference.

It is often a criticism of Bayesian probability that, although it offers clear guidance on how to use prior probabilities through Eq. (1), it is neutral as to how they are derived in the first place. As is the case with inference in general, priors have an inescapable subjective element, which highlights the fact that all inference is subjective to a point. In this work, flat priors have been assumed for all parameters, with lower bounds l_i and upper bounds u_i, such that the normalized prior density for each parameter m_i is given by

$$p(m_i|\mathcal{M}_k)=\frac{1}{(u_i-l_i)}=\frac{1}{w_i}. \tag{3}$$

In such cases, provided the bulk of the posterior probability is bounded within the assumed w_i, the prior is said to be *diffuse* and does not significantly influence the shape of the PPD. Even diffuse priors, however, can impact the *relative* probabilities of proposed models, thus influencing the second level of inference, as described next.

B. Second level of inference

Whereas at the first level of inference only the likelihood and (usually to a lesser extent) the prior are important, the second level involves the denominator, or *evidence* term, which is the normalizing constant for the PPD, *viz.*

$$p(\mathbf{d}|\mathcal{M}_k)=\int p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)p(\mathbf{m}|\mathcal{M}_k)d\mathbf{m}. \tag{4}$$

At the second level of inference, the intention is to rank each model $\mathcal{M}_k$ in terms of its ability to explain the data. As is well known, such a ranking cannot be based solely on the likelihood, as arbitrarily complex models can match the data arbitrarily closely. Paralleling Eq. (1), a posterior probability can also be attributed to each $\mathcal{M}_k$, independent of its parameters. However, in this case, normalization is not meaningful, leaving the proportionality

$$p(\mathcal{M}_k|\mathbf{d})\propto p(\mathbf{d}|\mathcal{M}_k)p(\mathcal{M}_k). \tag{5}$$

The reason that Eq. (5) cannot be normalized is simply that the universe of models $\mathcal{M}_k$ is subjective and infinite. Usually, the best that can be done is to enumerate M models for comparison, and—assuming no initial preference—the resulting prior probabilities are given by

$$p(\mathcal{M}_k)=1/M. \tag{6}$$

It then follows, for the purposes of model comparison, that the posterior probability of each model is simply proportional to its evidence, which for reasons that will become apparent later in this section is sometimes referred to as the *marginal* or integrated likelihood. According to Bayesian theory, the model associated with the highest numerical evidence as given by Eq. (4) is to be preferred.

What may not be immediately apparent from the preceding discussion is how Bayesian evidence automatically discriminates against models that are overly complex, thereby expressing a preference for *parsimonious* parametrizations.[7] This preference is also known as Occam's razor[6] and becomes more visible after assuming a *hypothetical* Gaussian form for the parameter likelihood around the maximum likelihood (ML) point $\hat{\mathbf{m}}$ [with an ML value $p(\mathbf{d}|\hat{\mathbf{m}},\mathcal{M}_k)$ and a parameter covariance matrix $\mathbf{C}_m$]

$$p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)=p(\mathbf{d}|\hat{\mathbf{m}},\mathcal{M}_k)\exp\left(-\frac{\Delta\mathbf{m}^T\mathbf{C}_m^{-1}\Delta\mathbf{m}}{2}\right). \tag{7}$$

In Secs. V B and V C, it will be shown that Eq. (7) is often *not* a good approximation to real likelihood functions, which require numerical integration (Sec. III F). However, such approximations are common in probability theory,[8,6] principally because the integral or *hypervolume* of a Gaussian likelihood function in P dimensions has the analytic form

$$\int p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)d\mathbf{m}=p(\mathbf{d}|\hat{\mathbf{m}},\mathcal{M}_k)(2\pi)^{P/2}\sqrt{\det\mathbf{C}_m}. \tag{8}$$

Making the further simplifying assumptions that the prior is flat and that the integral in Eq. (4) is not strongly affected by including the tails of $p(\mathbf{d}|\mathbf{m},\mathcal{M}_k)$ which lie outside the prior ranges u_i and l_i, Eq. (4) can be rewritten

$$p(\mathbf{d}|\mathcal{M}_k) = p(\mathbf{d}|\hat{\mathbf{m}},\mathcal{M}_k)(2\pi)^{P/2}\sqrt{\det \mathbf{C}_m}\,p(\mathbf{m}|\mathcal{M}_k)$$
$$= \underbrace{p(\mathbf{d}|\hat{\mathbf{m}},\mathcal{M}_k)}_{\text{likelihood}} \times \underbrace{(2\pi)^{P/2}\sqrt{\det \mathbf{C}_m}\prod_{i=1}^{P} w_i}_{\text{Occam factor}}, \quad \text{evidence} \tag{9}$$

from which it can be seen that the maximum likelihood achieved by a particular model does, as expected, weigh positively toward its ranking. Weighing against the likelihood, however, is the so-called *Occam factor*,[6,7,9] which can be interpreted as an integral of the prior probability density weighted according to the distribution of the likelihood. In other words, the Occam factor is a measure of the concentration of prior probability within the high-likelihood region of a parameter space, and it displays the following general properties:[6–9]

(i) The Occam factor penalizes models that incorporate large numbers of free parameters through the growth of the prior hypervolume $\Pi_{i=1}^{P} w_i$.
(ii) For the same reason, the Occam factor penalizes models with wide *a priori* parameter bounds w_i.
(iii) The Occam factor penalizes models that have to be finely tuned to fit the data, as these have concentrated PPDs and correspondingly small posterior volumes.

In geophysical theory, the first to describe Bayesian evidence in model selection was Jeffreys in 1939.[10] Only very recently, however, new applications have been reported.[7,9,11] The reasons for this no doubt stem from the complexity of real-world problems and the difficulties in deriving and manipulating the necessary probability distributions. Although Eq. (4) is a simple prescription, such integrations require specialized numerical methods that have only recently become practical.

Whereas alternative concepts such as minimum descriptive lengths (MDL), Akaike information criteria (AIC), and likelihood ratio tests have also been applied to model selection,[12] proponents of Bayesian inference point to the fact that the concept of evidence flows naturally and consistently from the basic *desiderata* of probability theory. In fact, most other methods can be viewed as being closely related to, or approximations of the full Bayesian framework described here.[6] Recent analyses have indicated that Bayesian model selection, which has higher computational demands than other approximate methods, is capable of superior performance, particularly in cases of low SNR and/or small amounts of data.[13]

C. Marginal inference

As discussed above, the calculation of (usually nonanalytic) PPDs is central to numerical applications of the first and second levels of Bayesian inference. When summarizing *a posteriori* knowledge of parameter values, however, it is desirable to treat each m_i (or perhaps pairs of m_i in the 2D marginal probability distribution case) separately, while integrating over the range of influence of the other parameters $m_{i\neq j}$, giving *marginal* probability distributions of the form

$$p(m_j|\mathbf{d},\mathcal{M}_k) = \int p(m_j, m_{i\neq j}|\mathbf{d},\mathcal{M}_k)\,dm_{i\neq j}. \tag{10}$$

Parameter estimation based on the properties of marginal distributions is known as marginal inference, and will be applied extensively in Sec. V, in which both one- and two-dimensional marginal distributions are computed from samples drawn from the joint parameter PPD.

III. GIBBS SAMPLING

The basic inputs to Bayesian analysis are probability distributions, and in practice, these can be difficult to estimate given the dimensionality of real-world problems. Gibbs sampling is an iterative Markov chain Monte Carlo (MCMC) procedure designed to sample from joint posterior distributions using only samples from conditional distributions. In the case of large-scale problems with many parameters possessing unknown correlations, joint distributions are usually not available, whereas conditional distributions often are.

The origins of the Gibbs sampler date back to Hastings[14] in statistical analysis and subsequently Geman and Geman,[15] who applied the idea to large-scale image reconstruction problems. Comprehensive accounts of the theory and application of Gibbs sampling to inverse problems—particularly those in geophysics—as well as references to the original literature include Gelfand and Smith,[16] Smith and Roberts,[17] Sen and Stoffa,[18] and Mosegaard and Sambridge.[11] In relation to ocean-acoustic problems of the type of interest here, Dosso[3–5] recently analyzed synthetic and experimental data and concluded that Gibbs sampling is a powerful and robust means of estimating geoacoustic parameters and their associated errors.

In this section, we briefly discuss issues relevant to the integration of high-dimensional PPDs in geoacoustic inversion. These include the Metropolis–Hastings approach to sample generation, definition of a matched-field energy function, Gibbs sampler initialization, coordinate rotation, convergence criteria, and finally, one approach to the numerical estimation of Bayesian evidence.

A. Sample generation

In essence, Gibbs sampling consists of generating successive samples from P conditional distributions, such that at the completion of the qth iteration, the parameter vector $(m_1^q, m_2^q, m_3^q, \ldots, m_{P-1}^q, m_P^q)$ can be considered to have been drawn from the joint PPD. For clarity, it will be implicitly assumed that posterior probabilities are conditional on the data $\mathbf{d}$ and model $\mathcal{M}_k$ from this point. Starting with the initial parameter vector $(m_1^0, m_2^0, m_3^0, \ldots, m_{P-1}^0, m_P^0)$ and denoting the conditional distribution of parameter m_1 with respect to parameters $m_{2\rightarrow P}$ in the first iteration as $p(m_1|m_2^0, m_3^0, \ldots, m_{P-1}^0, m_P^0)$, the sampling proceeds with each successive conditional distribution immediately incor-

porating the previously selected parameter value. Using the "←" symbol to denote the drawing of a sample, the first iteration can be illustrated as[8]

$$m_1^1 \leftarrow p(m_1|m_2^0, m_3^0, \ldots, m_{P-1}^0, m_P^0)$$
$$m_2^1 \leftarrow p(m_2|m_3^0, m_4^0, \ldots, m_P^0, m_1^1)$$
$$m_3^1 \leftarrow p(m_3|m_4^0, m_5^0, \ldots, m_1^1, m_2^1)$$
$$\vdots \quad \vdots \quad \vdots$$
$$m_P^1 \leftarrow p(m_P|m_1^1, m_2^1, \ldots, m_{P-2}^1, m_{P-1}^1),$$

with the qth iteration given by

$$m_1^q \leftarrow p(m_1|m_2^{q-1}, m_3^{q-1}, \ldots, m_{P-1}^{q-1}, m_P^{q-1})$$
$$m_2^q \leftarrow p(m_2|m_3^{q-1}, m_4^{q-1}, \ldots, m_P^{q-1}, m_1^q)$$
$$m_3^q \leftarrow p(m_3|m_4^{q-1}, m_5^{q-1}, \ldots, m_1^q, m_2^q)$$
$$\vdots \quad \vdots \quad \vdots$$
$$m_P^q \leftarrow p(m_P|m_1^q, m_2^q, \ldots, m_{P-2}^q, m_{P-1}^q).$$

In the Metropolis–Hastings variant of the Gibbs sampler used here, samples are drawn from each conditional distribution according to the Metropolis rule, wherein uniformly distributed perturbations, resulting in modified parameter vectors $\mathbf{m}'$, are accepted with probability

$$p_{\text{accept}} = \begin{cases} \dfrac{p(\mathbf{m}')}{p(\mathbf{m})} & \text{for } p(\mathbf{m}') \leqslant p(\mathbf{m}), \\ 1 & \text{for } p(\mathbf{m}') > p(\mathbf{m}). \end{cases} \tag{11}$$

As in Eq. (1), the $p(\mathbf{m})$ are posterior probabilities which, given simplifying Gaussian assumptions regarding the noise and modeling errors, take the exponential form

$$p(\mathbf{m}) = \exp[-E(\mathbf{m})], \tag{12}$$

where $E(\mathbf{m})$ is an *energy function* to be discussed shortly. The above combination of sequential perturbation, acceptance, and rejection establishes a Markov chain whose sampling density can be shown to converge to the Gibbs–Boltzmann distribution from thermodynamics[11,18]

$$p(\mathbf{m}) = \frac{1}{Z} \exp\left[-\frac{E(\mathbf{m})}{T}\right], \tag{13}$$

with the normalization or *partition* function Z ignored, and a particular choice of unity for the temperature T. This is the same equilibrium distribution associated with simulated annealing (SA) algorithms, and hence Gibbs sampling has the interpretation of being SA conducted at a constant temperature of $T=1$.[19,20]

When evaluating the multidimensional integrals [Eq. (10)] with importance sampling, the variance of the integral estimates is reduced if the generating distribution is proportional to the integrand, and this is actually obtained with a Gibbs sampler.[3,8,18] Generally, when evaluating the multidimensional integrals [Eq. (10)] with importance sampling, the sampling distribution should be concentrated in regions of the parameter space where the PPD is most significant. Ideally, if the PPD is perfectly mirrored by the sampling distribution, then the error in estimating the PPD and its associated marginal integrals tends to zero—a situation never achieved in practice, but approximated with a given amount of computation such that Gibbs sampling is usually far more attractive than either less adaptive or more exhaustive sampling techniques.[3,8,18] While other methods of importance sampling have also been applied to PPD estimation, including, for example, GA in ocean acoustics,[21–26] the great advantage of Gibbs sampling is that the sampling distribution is *known* to be given by Eq. (13). In the case of GA and related methods, the sampling distribution is usually not known and consequently the results may be difficult to interpret.[4]

B. Energy function

In Ref. 2, parameter estimation was carried out by directly maximizing normalized Bartlett power objective functions of the form

$$B_f(\mathbf{m}) = \left[\frac{\mathbf{w}_f^\dagger(\mathbf{m})\hat{\mathbf{R}}_f\mathbf{w}_f(\mathbf{m})}{\text{tr}[\hat{\mathbf{R}}_f]\|\mathbf{w}_f(\mathbf{m})\|^2}\right], \tag{14}$$

where the $\mathbf{w}_f$ were replica vectors calculated from a continuous wave acoustic propagation model and the $\hat{\mathbf{R}}_f$ ($f=1,\ldots,F$) were cross-spectral density matrices (CSDMs) averaged from N_s snapshot vectors $\mathbf{d}_{n,f}$ of the array data at frequencies ω_f according to

$$\hat{\mathbf{R}}_f = \frac{1}{N_s}\sum_{n=1}^{N_s} \mathbf{d}_{n,f}\mathbf{d}_{n,f}^\dagger. \tag{15}$$

To formulate the energy function $E(\mathbf{m})$ required by Eq. (12), Eq. (14) can be transformed into the *mismatch* or error function[27]

$$\phi_f(\mathbf{m}) = 1 - B_f(\mathbf{m}), \tag{16}$$

following which[3,25]

$$E(\mathbf{m}) = \sum_{f=1}^{F} \frac{\phi_f(\mathbf{m})|\mathbf{d}_f|^2}{\hat{\sigma}_f^2}, \tag{17}$$

where $|\mathbf{d}_f|^2$ is the total power seen by the array in frequency band f and $\hat{\sigma}_f^2$ is a corresponding estimate of the variance associated with the combination of model mismatch, ocean, and instrumental noise.

To estimate $\hat{\sigma}_f^2$, two simplifying assumptions have been made here. First, considering the single frame of data to be discussed in Sec. IV, and the three frequencies selected for inversion, it is reasonable to take the SNR as approximately constant across frequency. Second, the spatial distribution of noise along the array has been assumed constant despite the short range of the experiment which, in reality, would have resulted in higher SNRs at the end of the array nearer the source. Following the maximum-likelihood arguments of Gerstoft and Mecklenbräuker in relation to the noise variance expected after averaging N_s snapshots, as in Eq. (15), the average variance can be estimated as[3,25,27]

$$\hat{\sigma}_{\text{ave}}^2 = \frac{\phi_{\text{ave}}(\hat{\mathbf{m}})|\mathbf{d}_{\text{ave}}|^2}{N_e}, \tag{18}$$

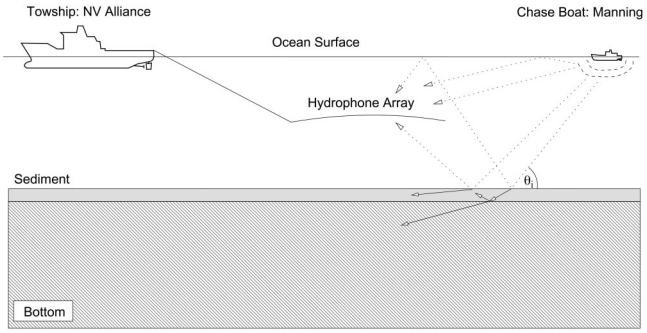


FIG. 1. The configuration of the ALLIANCE and the MANNING during the MAPEX2000 self-noise experiment.

where $\phi_{\text{ave}}(\hat{\mathbf{m}})$ is the average normalized Bartlett mismatch obtained using a global search (at the first level of inference), $|\mathbf{d}_{\text{ave}}|^2$ is the average measured power, and N_e is an estimate of the number of degrees of freedom associated with the noise.

As discussed by Gerstoft and Mecklenbräuker[25,27] and Dosso,[3,4] the selection of appropriate values for N_e (and hence $\hat{\sigma}^2_{\text{ave}}$) in geoacoustic inversion is problematic due to the fact that the structure of the true model and hence also that of the true signal are generally unknowable. Because of this, *modeling* errors actually tend to dominate even at moderate SNR values—in most cases swamping the impact of noise completely. Faced with this difficulty, we adopted a heuristic approach and chose $N_e=5$ (Refs. 28, 25, 3, 4) noting that: (1) This was approximately the number (six) of independent snapshots used, and (2) from experience, acceptance ratios (proportions of accepted Metropolis moves) on the order of 25% to 50% are typically indicative of Gibbs sampler algorithms performing correctly.[11,13] Too large an estimate of N_e leads to artificially narrow PPDs *relative to the prior bounds* through the exponentiation in Eq. (12), and correspondingly poor acceptance ratios.

To summarize, the final energy function used in our Gibbs sampler had the form

$$E(\mathbf{m}) = S\sum_{f=1}^{F} \phi_f(\mathbf{m}), \tag{19}$$

with $S = N_e / \phi_{\text{ave}}(\hat{\mathbf{m}})$. Given that the highest average Bartlett power (correlation coefficient squared) obtained here was approximately 0.8, the overall energy scaling factor S was

$$S = \frac{N_e}{\phi_{\text{ave}}(\hat{\mathbf{m}})} \approx \frac{5}{0.2} = 25. \tag{20}$$

C. Initialization

As Gibbs samplers essentially sample around the modes of PPDs, they must be correctly initialized to regions of high posterior probability. In many reported applications, an initialization or *burn-in* phase has preceded to the equilibrium ($T=1$) sampling phase such that an initial quenching from some high temperature takes place.[8,3,4] In this work, we found that SA initialization did perform satisfactorily provided the starting temperature and annealing schedule were correctly chosen—usually by trial and error. However, substituting a genetic algorithm (GA) in place of SA ultimately proved faster in locating the main concentrations of probability. This was probably due to the insensitivity of GA to parameter correlations that can seriously affect the acceptance ratios obtained during conventional SA.

D. Coordinate rotation

As noted by Collins and Fishman,[29] algorithms that generate univariate parameter perturbations, such as SA and Gibbs sampling, are particularly susceptible to poor acceptance ratios when sampling strongly correlated parameter spaces. In ocean acoustics, such parameter correlations are frequently encountered,[30] and coordinate rotation, whereby the parameter covariance matrix **C**, or alternatively, a covariance matrix computed from field derivatives[29] is effectively orthogonalized, is a standard solution that will not be discussed here in detail. Following Dosso,[3] our Gibbs sampler

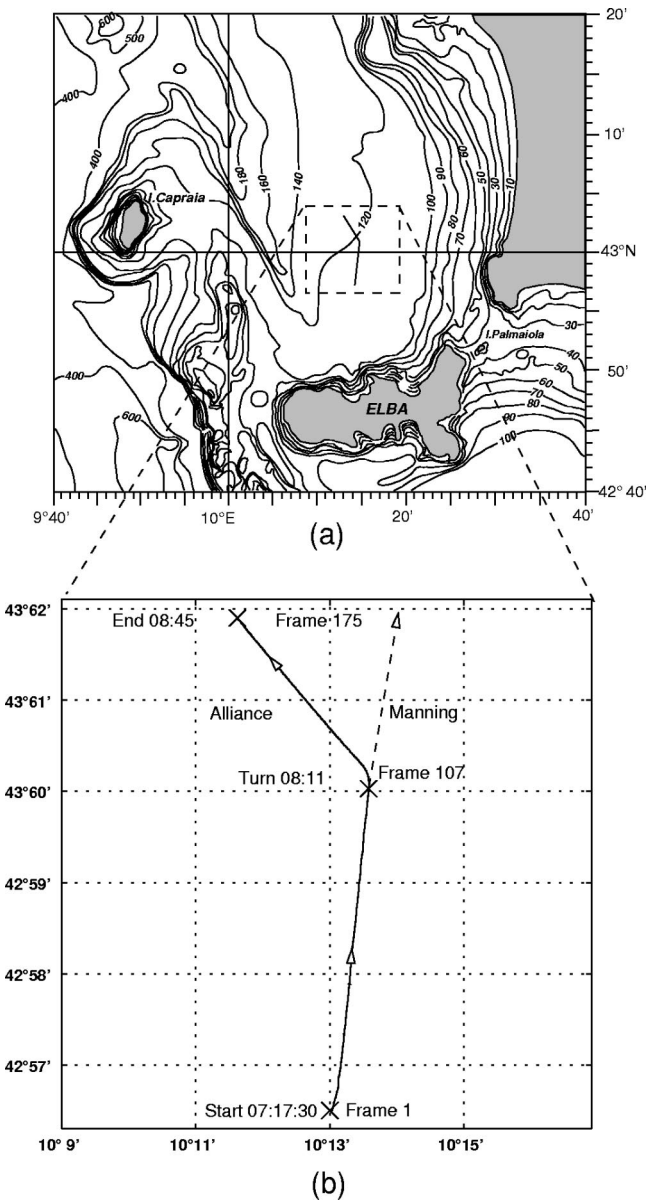


FIG. 2. The tracks of the NRV ALLIANCE and MANNING during the MAPEX2000 self-noise experiment. All times are UTC. Each frame represents a 30-s increment.

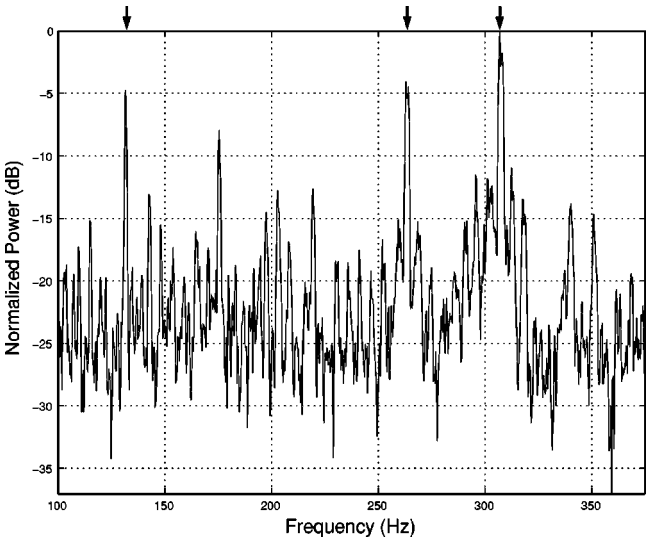


FIG. 3. Normalized power spectral density computed from time-series data beamformed in the approximate direct-path direction of the support ship MANNING. The three frequencies used for inversion were 131.47, 262.94, and 306.88 Hz, as indicated by the top markers.

originally included a covariance estimation phase conducted at $T=1$ following which the parameter covariance was estimated from sampled vectors $\mathbf{m}$ as

$$\mathbf{C}_m \approx \langle \mathbf{m}\mathbf{m}^\dagger \rangle - \langle \mathbf{m} \rangle \langle \mathbf{m} \rangle^\dagger. \tag{21}$$

Collins and Fishman[29] originally suggested using the covariance matrix of the derivatives of the field. Instead of sampling the covariance matrix $\mathbf{C}_m$, the Cramer–Rao lower bound matrix $\mathbf{C}_{\mathrm{CRLB}}$ which contains information about parameter coupling, can be used to estimate the required parameter rotations. In this work the Cramer–Rao formulation due to Baggeroer *et al.* extended to the multifrequency case[31] is used

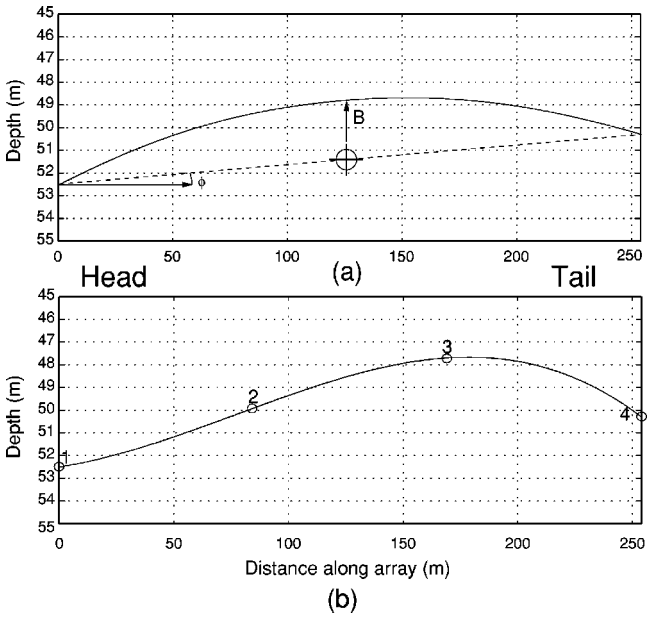


FIG. 4. Towed-array shape models. (a) Simple parabolic model with depth, bow and tilt. (b) Cubic spline model fitting a variable number of depth points—in this case z_{a1}, z_{a2}, z_{a3}, and z_{a4}.

$$[\mathbf{C}_{\mathrm{CRLB}}^{-1}]_{ij} = \sum_{f=1}^{F} \mathrm{Tr}\left[\mathbf{K}_f^{-1} \frac{\partial \mathbf{K}_f}{\partial m_j} \mathbf{K}_f^{-1} \frac{\partial \mathbf{K}_f}{\partial m_i}\right], \tag{22}$$

where the $\mathbf{K}_f$ are outer products of replica fields $\mathbf{w}_f$ calculated at the ML point $\hat{\mathbf{m}}$ and loaded uniformly on their diagonals to approximate the estimated SNR of the data. That Eq. (22) can be used in this way *independently of the data* is less surprising when it is considered that coordinate rotation is simply another kind of importance sampling designed to concentrate samples in regions of high posterior probability. While it is conceivable that a proposed ocean model might be sufficiently far from reality for Eq. (22) to fail as the basis of such a sampling distribution, in all cases reported here we observed substantial improvements in Gibbs sampler acceptance ratios when $\mathbf{C}$ was estimated semianalytically using Eq. (22) (up to 50%) as opposed to Eq. (21) (≈25%).

E. Convergence criteria

As we were mainly interested in *marginal* PPDs, we based our Gibbs sampler convergence criteria on the maximum fractional change undergone by any 50-bin parameter histogram in 100 cycles through the Gibbs sampler. While not as rigorous as the dual-population convergence criteria used by others,[3] we checked for convergence every 100 steps and found that histogram changes below 10% were a reliable sign of convergence. All results presented in Sec. IV achieved this criterion.

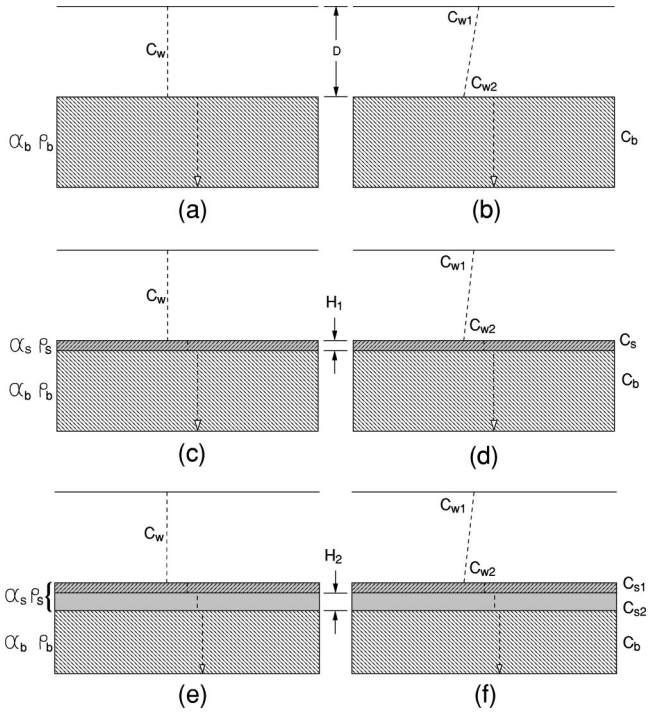


FIG. 5. Six acoustic waveguide models with varying degrees of complexity and their associated parameters. (a) Iso-velocity water over half-space. (b) Sound-speed gradient over half-space. (c) Iso-velocity water over sediment and half-space. (d) Sound-speed gradient over sediment and half-space. (e) Iso-velocity water over two-layer sediment and half-space. (f) Sound-speed gradient over two-layer sediment and half-space. Note the distinct surface and bottom water column velocities C_{w1} and C_{w2} in models (b), (d), and (f). Also note that while the two sediment layer velocities C_{s1} and C_{s1} are distinct in models (e) and (f), the densities and attenuations of these layers (ρ_s and α_s) have been assumed equal for simplicity.

TABLE I. Parameters and inversion bounds for the first case involving a parabolic array and Pekeris environment.

Parameter	Symbol	Unit	Min	Max
Water sound speed	C_w	m/s	1510	1530
Water depth	D	m	115	125
Bottom sound speed	C_b	m/s	1450	1650
Bottom density	ρ_b	g/cm^3	1	3
Bottom attenuation	α_b	dB/λ	0	2
Source depth	Z_s	m	0	5
Source range	R_s	m	275	325
Array depth	Z_a	m	45	55
Array tilt	ϕ	degrees	−2	2
Array bow	B	m	−5	5

F. Reverse importance sampling

In Sec. II B, Bayesian evidence was defined in terms of an integral that, given a Gaussian approximation to the parameter likelihood and flat priors, had a simple analytic form. Unfortunately, as will be apparent in Sec. V, such approximations can result in errors of many orders of magnitude and sometimes compromise the outcome of Bayesian model selection. More accurate calculation of evidence calls for a numerical integration approach.

Ó Ruanaidh and Fitzgerald[8] describe such an approach that is closely related to the concept of importance sampling, but works in reverse to extract the probabilistic evidence of a model from samples $\mathbf{m}_q$ already available from its joint PPD.

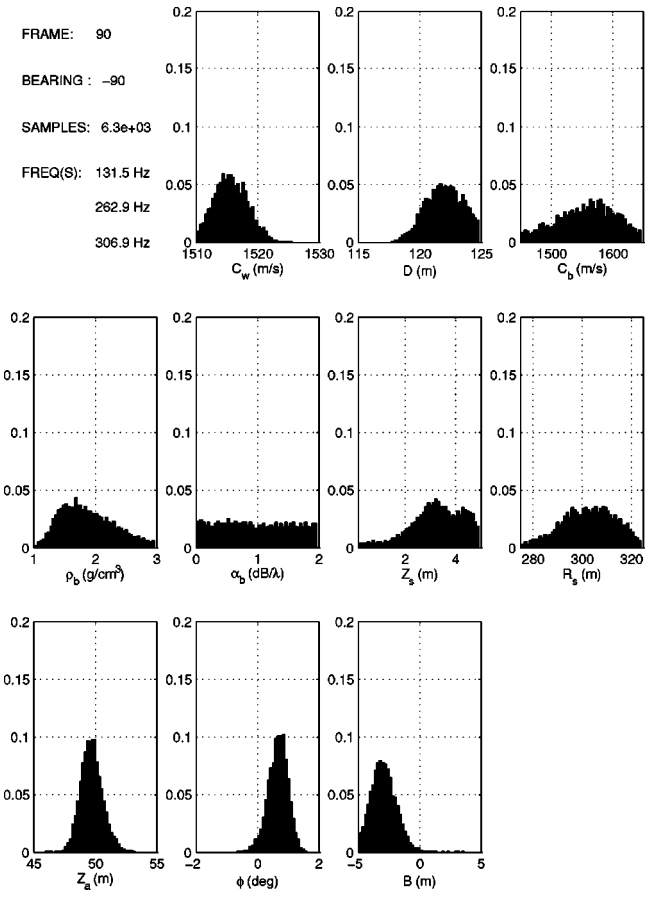


FIG. 6. Marginal PPD histograms for frame 90. Environment model (a), parabolic array and MANNING source.

In the present case, these samples will have been generated at the first level of inference, i.e., in estimating the PPD of the parameters via Gibbs sampling. The task then is to invoke a Gaussian approximation, not for the integral itself, but to be *reverse sampled* by the existing ensemble, thereby calibrating the unknown evidence integral against that of the Gaussian, which is known. Taking $g(\mathbf{m})$ as a normalized Gaussian

$$\int g(\mathbf{m})d\mathbf{m}=1, \tag{23}$$

and $\mathcal{E}$ as the unknown evidence integral

$$\mathcal{E}=\int f(\mathbf{m})d\mathbf{m}, \tag{24}$$

$\mathcal{E}$ can be shown to be approximated by[8]

$$\frac{1}{\mathcal{E}}\approx\frac{1}{Q}\sum_{q=1}^{Q}\frac{g(\mathbf{m}_q)}{f(\mathbf{m}_q)}. \tag{25}$$

Calculating Bayesian evidence by reverse importance sampling (RIS) is straightforward, requiring negligible further computation once the PPD samples $\mathbf{m}_q$ and their functional values $f(\mathbf{m}_q)$ have been generated by Gibbs sampling.

IV. SELF-NOISE INVERSION

A. MAPEX2000

The 90-min experiment discussed here was conducted by the NATO Undersea Research Center and the Marine Physical Laboratory as part of MAPEX2000, which was specifically directed at validating a range of array processing and geoacoustic inversion techniques. The data were acquired north of the island of Elba, off the Italian west coast on the 29th of November, 2000 between the times of 07:17:30 and 08:45:00 UTC.[2] The array used consisted of 128 hydrophones evenly spaced at 2 m. Half-wavelength sampling therefore occurred at approximately 375 Hz. Depth control with this array proved less accurate than hoped, although this turned out to be of little consequence, as the array shape was included in the vector $\mathbf{m}$ along with the other parameters to be optimized.

The self-noise inversion dataset comprised 175 frames sampled at 30-s intervals. Only the first 10 s of each interval, or 60 000 samples at a sampling rate of 6000 samples per second, were recorded. One novel aspect of this self-noise experiment was that two research vessels were involved. From about frame 60 to frame 107, the NRV ALLIANCE, with its array towed approximately 330 m behind, was followed by a smaller vessel, the MANNING, at a range of approximately 900 m. The horizontal distance of the MANNING from the tail of the array was therefore approximately 300 m.

As in our earlier analysis,[2] we have made the assumption here that the source ship was well approximated as a point source over the range of frequencies considered. While realizing that this may not generally be the case with larger vessels or at higher frequencies, the point-source assumption seems to have been borne out at least in the case of the MANNING by consistently sharp matched-field processor (MFP) peaks.

TABLE II. Gibbs sampler inversion results for the first case involving a parabolic array and Pekeris environment.

Parameter	Unit	MAP	Mean	Std. dev.
C_w	m/s	1515	1516	3
D	m	121.5	122.1	1.5
C_b	m/s	1620	1560	47
ρ_b	g/cm^3	2.0	1.9	0.4
α_b	dB/λ	2.0	1.0	0.6
Z_s	m	4.5	3.3	1.0
R_s	m	295	303	7
Z_a	m	50.2	49.9	0.9
ϕ	degrees	0.4	0.7	0.3
B	m	−2.3	−2.8	1.0

At frame 107, ALLIANCE started a 45° turn, while MANNING continued on track. Then, at frame 122, MANNING left station and departed the area at increased speed without turning. In this analysis, we consider only the 10 s of data acquired in frame 90, during which the MANNING was in station approximately 300 m behind the Alliance, as illustrated in Fig. 1. We do not consider the ALLIANCE, which posed much lower a source level for the purposes of reliable self-noise inversion.[2]

Figure 2 shows the track of the two ships during the experiment, and gives an indication of the local bathymetry. Over the entire track, the depth sounder recorded a steady increase in depth from approximately 116 to 124 m. Expendable bathy-thermograph casts showed an almost iso-velocity (winter) water column with a velocity of approximately 1520 m/s. The speed of the ALLIANCE was a steady 4 knots and the prevailing sea state was 1.

In the area of the experiment, the seafloor has been fairly well characterized over the last 30 years,[23,32,26,33] and is known to be flat and covered with a thin layer of clay and sand–clay sediments. Below the thin sediment layer is known to exist a reasonable approximation to an acoustic half-space with the approximate parameters $c_b \approx 1600$ m/s, $\rho_b \approx 1.8$ g/cm^3, and $\alpha_b \approx 0.15$ dB/λ.[26] At finer levels of detail, Holland and Osler[34] have noted the existence of some randomly placed thin, high-speed layers within sediment core samples.

In no results to date has there been evidence of significant shear wave propagation in the seafloor in the North Elba area, and hence only compressional wave parameters were considered here. In any case, the assumption of a shear-wave velocity c_s within the plausible range for the bottom type in question ($c_s \ll c_w$) would lead mainly to an increased bottom loss,[35] to which the current near-field inversion technique would be relatively insensitive.

B. Preprocessing

From each 60 000-point frame, 6 Kaiser–Bessel windowed, 16 384-point snapshots were averaged with 50% overlap in order to estimate both the power-spectral densities (PSDs) and cross-spectral density matrices (CSDMs). Elementary frequency tracking was used to follow small variations in the tonal frequencies emitted by the MANNING. The three center frequencies selected for inversions were 131.47, 262.94, and 306.88 Hz, as indicated by the markers at the top of the normalized PSD in Fig. 3, which was computed from time-series data beamformed in the approximate direct-path direction of the MANNING. In this paper, all angles are given with respect to the vertical down direction, with 90° pointing toward the towship ALLIANCE and −90° pointing toward the support ship MANNING.

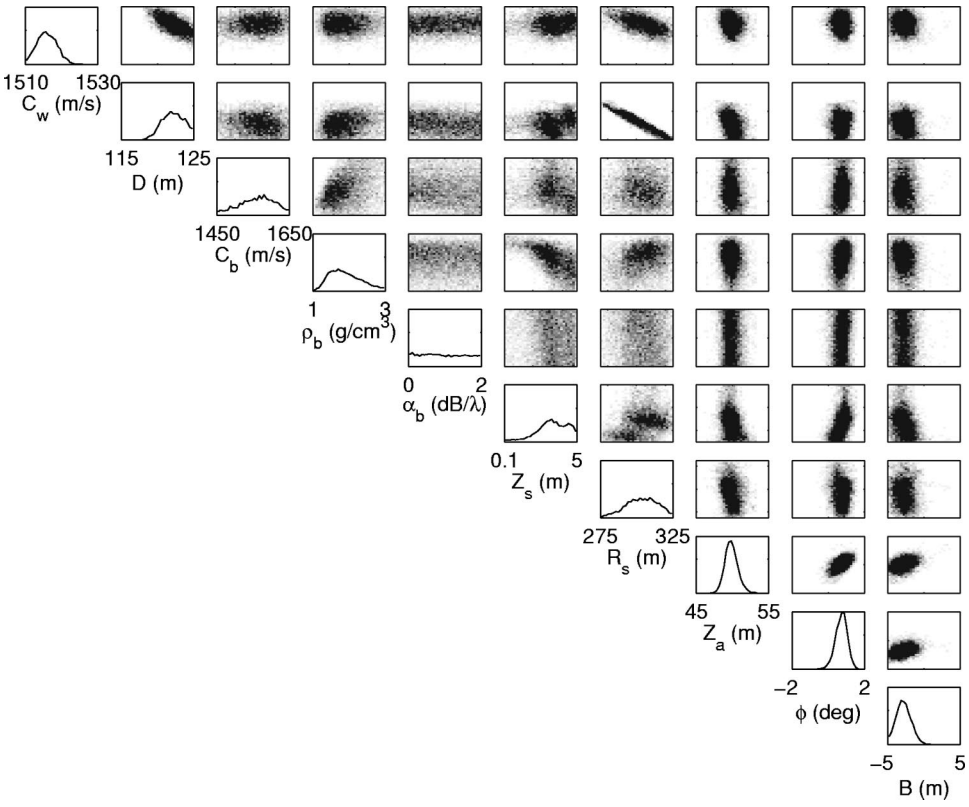


FIG. 7. 1D and 2D Marginal PPDs for frame 90. Environment (a), parabolic array and MANNING source (PPDs below the diagonal are symmetric with those above).

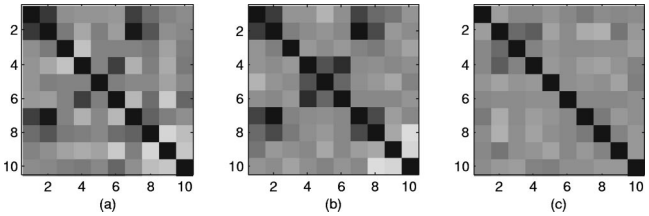


FIG. 8. Parameter correlation matrices for the Pekeris environment and parabolic array (parameter order as per Tables I and II): (a) Gibbs sampler estimate; (b) Estimate via Eq. (22); (c) Estimate via Eq. (21) after sampling in rotated coordinates. The parameter indices now relate to a new set of orthogonal coordinates.

Given the level of interference originating from nearby shipping traffic, it was found useful to incorporate additional spatial filtering of the snapshot data. Preprocessing therefore entailed temporal windowing and 2D FFT transformation of each snapshot, frequency masking to exclude certain ranges of spatial frequencies, and then inverse 1D FFT transformation back into phone-frequency space. In this analysis, the spatial frequency cutoff was set to include arrival angles within 60° of endfire to capture most of the expected multipath arrivals from the MANNING. To retain the normalization of Eq. (14), identical filtering was applied to the replica vectors $\mathbf{w}_f$.

C. Propagation modeling

Accurately modeling the response of acoustic environments is a principal aspect of geoacoustic inversion. In the present case of self-noise inversion, particular attention is required to aspects of short-range propagation that differentiate it from long-range waveguide propagation—namely the leaky or *virtual modes* that result from steep angles of bottom incidence.[1] In view of this, two approaches to propagation modeling capable of accuracy in the near field were used, the first of which—wave number integration—has already been detailed in Ref. 2. The second approach, based on the complex effective depth (CED) ideas of Zhang and Tindle,[36] was motivated by the modeling requirements of Gibbs sampling, which can easily run to the order of 10^5 models for a single inversion.

The first great advantage of CED models is that they are based on normal modes, which are characteristic of the environment and independent of source–receiver geometry. By keeping environmental parameters at the beginning of the Gibbs chain, it is possible to improve efficiency by reusing the mode functions to evaluate the array field for a wide variety of geometric variations. Second, by replacing the single-interface reflection coefficient of Zhang and Tindle with an invariant embedding scheme,[37] the capability of our CED code was extended to arbitrary bottom layering at little additional cost. The disadvantage of our existing CED code is that it cannot handle water column sound-speed gradients, which were instead modeled by wave number integration in this work. Modifications of our CED code along the lines described by Westwood *et al.*[38] are expected to remedy this problem, enabling Gibbs sampling to be applied efficiently to more complex environments for which wave number integration is currently slower by a factor of about 20.

TABLE III. Results for the array model selection. The three-point spline is seen to give the highest evidence. Note that the table displays log evidence and the probability of a model is based on just the evidence; thus, the small differences for each model are important.

		Log evidence	
Model	Min E	RIS	Gaussian
Parabola	20.1	−14.4	−14.3
Spline 3	**19.4**	**−14.2**	**−13.8**
Spline 4	19.1	−14.4	−13.8
Spline 5	18.9	−14.8	−14.1
Spline 6	17.9	−15.3	−14.4
Spline 7	17.7	−15.5	−14.6
Spline 8	17.3	−15.7	−15.5

D. Array modeling

From the standpoint of real data analysis, it is important to allow for geometric distortion of the array from its ideal horizontal and straight configuration. Whereas in Ref. 2 we modeled the array as a parabolic curve with bow B meters and tilt ϕ degrees, in this work we extended the description to a cubic spline passing through multiple equally spaced points. Each point has freedom in depth, allowing complicated array shapes to be modeled. Figure 4 is an illustration of both models, as it is our intention in Sec. V to compare each with respect to Bayesian evidence, thereby quantifying the level of complexity most appropriate.

V. RESULTS

In this section, we present inversion results obtained for frame 90 of the MAPEX2000 dataset using the MANNING as a source, and a variety of geoacoustic and array parametrizations. Strictly speaking, the selection of the array and environmental models should have been made concurrently; however, we undertook these analyses separately on the assumption that the array parameters were well determined by the data and hence hierarchically separable from the less

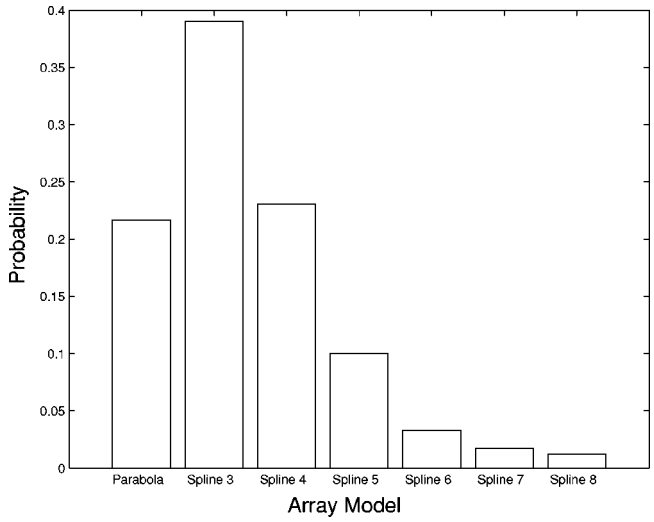


FIG. 9. Normalized probabilities for the towed-array models listed in Table III based on Bayesian evidence.

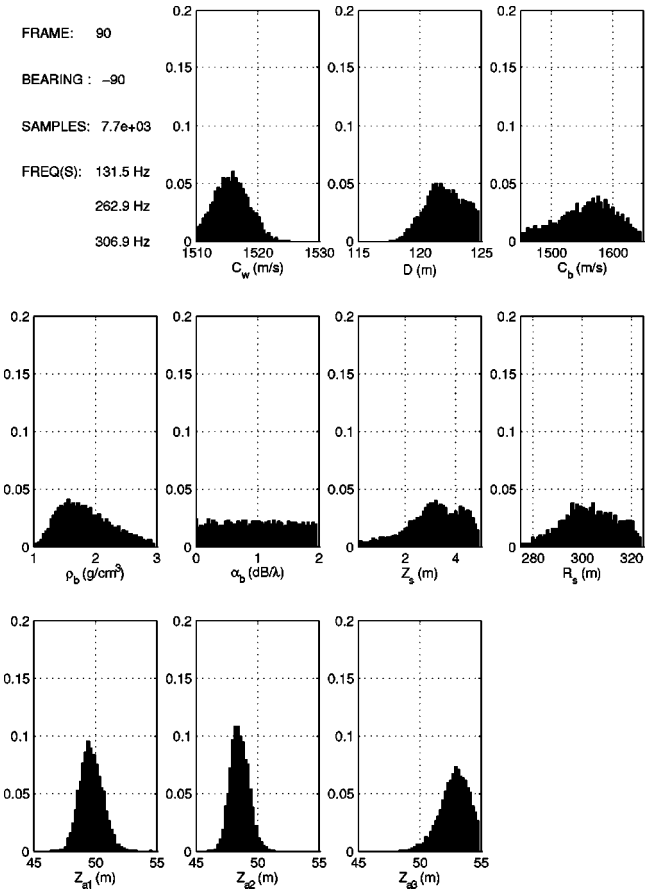


FIG. 10. Marginal PPD histograms for frame 90. Environment model (a), three-point array and MANNING source.

strongly determined environmental parameters. MacKay[6] remarks that this approach can often emerge as a reasonable approximation.

The structure of this section, then, is first in Sec. V A to revisit the baseline case of the parabolic array and Pekeris waveguide environment. Second, in Sec. V B, the behavior of Bayesian evidence with respect to array model complexity is analyzed assuming the Pekeris environment. Finally, in Sec. V C we fix the array model at the most probable configuration determined previously and compare the six environmental models illustrated in Fig. 5 with respect to their Bayesian evidence.

A. Parabolic array and Pekeris environment

The first result assumes the same simple Pekeris environment and parabolic array model as previously employed in Ref. 2, except that in this case, we have retained the bottom density ρ_b and attenuation α_b parameters previously neglected due to their perceived insensitivity. Table I details the search bounds used in the first inversion, which converged after 63 000 CED forward models (each at three frequencies) in approximately 1 h.

In Fig. 6, histograms generated by Gibbs sampling proportional to the marginal PPDs of the ten inverted parameters have been smoothly resolved into 50 bins, indicating good convergence. Table II summarizes the results of the first inversion in terms of MAP and mean parameter estimates in addition to standard deviations computed from the ensemble of samples. In most respects, the results obtained here agree well with those obtained earlier using a GA global search.[2] In particular, the MAP estimate of the bottom velocity C_p of 1580 m/s is identical to that obtained previously. In this case, however, at least the relative parameter sensitivities are clearly evident from the widths and curvatures of the marginal PPDs.

It is noteworthy that the bottom density ρ_b in Fig. 6 appears to have been well resolved. In fact, ρ_b appears as well resolved here as the bottom velocity C_b. As the bottom reflected arrivals are relatively steep, the magnitude of these arrivals is related to the impedance contrast at the bottom boundary. Physically, the estimated density of 1.9 ± 0.4 g/cm^3 agrees well with the value of 1.8 g/cm^3 commonly assumed for sediment densities in the North Elba area.[32,26,33]

Typifying the characteristics of a poorly resolved parameter, the variance of the *p*-wave bottom attenuation α_b is a large proportion of its prior range, and the marginal PPD is correspondingly flat. This confirms earlier observations that bottom attenuation is not a sensitive parameter in near-field inversion.

In Fig. 7, 1D and 2D marginal probability plots indicate relatively homogeneous PPDs for most parameter pairs, except in the case of water depth D and source range R_s, which shows a very strong positive correlation of the type commonly found in shallow-water acoustics.[21,30] This correlation is also apparent in the parameter covariance matrices, estimated via Eqs. (21) and (22) and plotted as *correlation* matrices ρ_{ij} in Figs. 8(a) and (b), respectively, such that

$$\rho_{ij} = C_{ij} / \sqrt{C_{ii} C_{jj}}. \tag{26}$$

As discussed in Sec. III D, the estimates obtained using the two methods should agree given sufficient samples and an accurate enough ML estimate. Visually, Figs. 8(a) and (b) bear a close resemblance, with high positive correlations between water depth D (parameter 2) and source range (parameter 7), and some negative correlations elsewhere. In subsequent sampling, these correlations were effectively removed through coordinate rotation as described previously, with the result being a more diagonally dominant correlation matrix for the orthogonalized parameter set, as illustrated in Fig. 8(c). Estimating **C** using Eq. (21) typically required ten thousand or so forward models to achieve 5% relative convergence, whereas using Eq. (22) took almost negligible time.

B. Array model selection

In this section, we isolate the issue of array selection from that of environment selection by holding the latter constant. Taking the simple Pekeris waveguide as a baseline environment, Table III summarizes the outcomes of Bayesian evidence calculations for seven array models, including the original parabolic array with three parameters, and six additional cubic spline arrays with up to eight parameters (equally spaced depth points). The second column of the table gives the minimum error energies achieved with each array, illustrating the point in Sec. II B that arbitrarily complex models can achieve arbitrarily good fits to the data. The third column gives the logarithms (to base 10) of the evi-

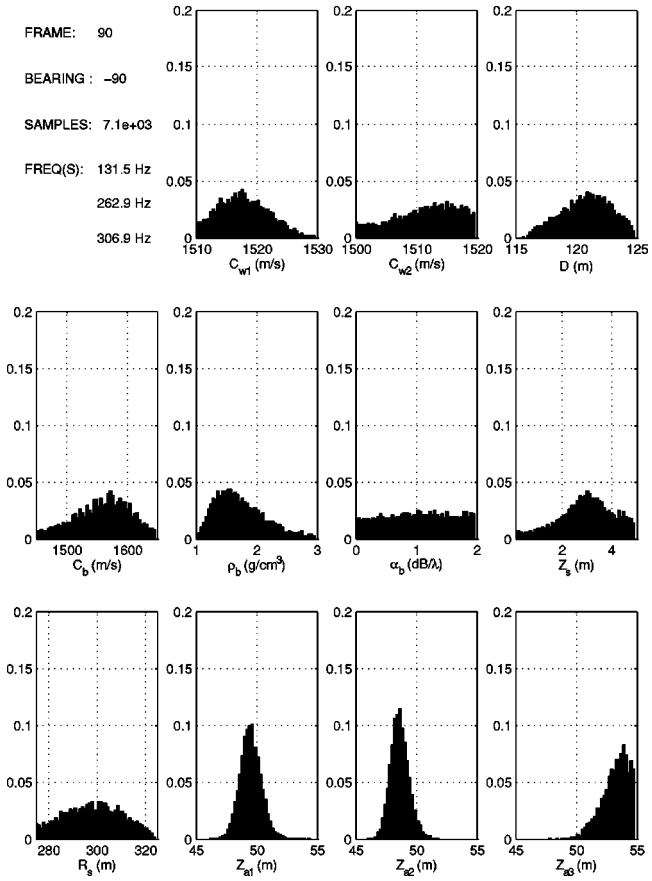


FIG. 11. Marginal PPD histograms for frame 90. Environment model (b), three-point array and MANNING source.

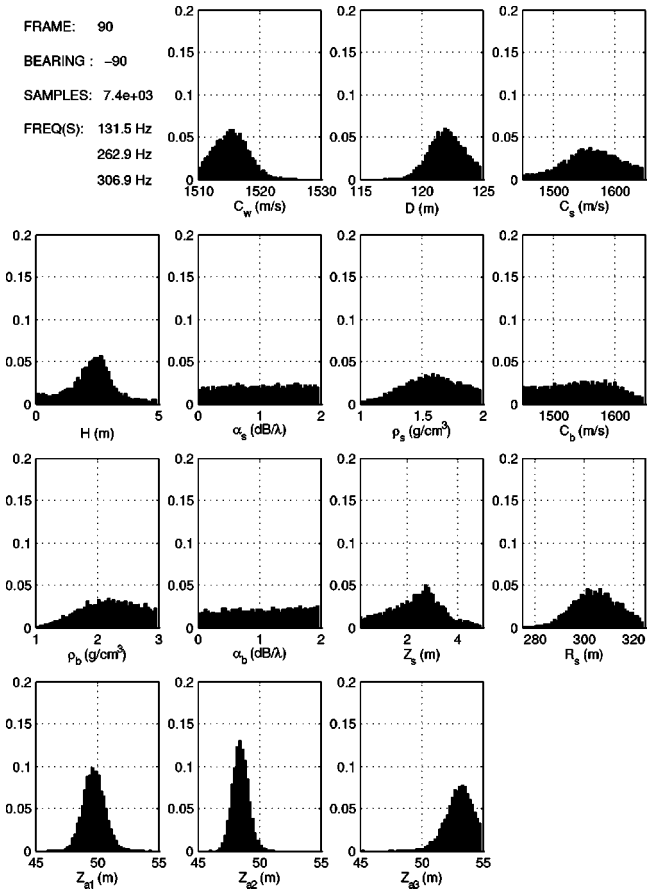


FIG. 12. Marginal PPD histograms for frame 90. Environment model (c), three-point array and MANNING source.

dence values as calculated by reverse importance sampling—the most accurate method used here. These log-evidence values can be compared with those in the fourth column, which were calculated by a fast approximate method based on the Gaussian assumption of Eq. (9). As discussed in Sec. II B, in the absence of prior preferences toward particular models, that with the highest evidence (or log evidence) is preferred, and the array model with the highest evidence in this case is the three-point spline. While array models of higher complexity achieved lower energies, the Occam factor implicit in the evidence formulation automatically discriminated against them, leading to a monotonic decrease in probability, as illustrated by the normalized probability plot of Fig. 9.

Interestingly, both the parabola and three-point spline possess the same number of parameters, though the geometric mapping of these parameters into the misfit function was different for each case. While an argument could be made that the particular prior bounds used weighed unreasonably against the parabolic array (or vice versa), these bounds were believed to be reasonably representative of the state of prior uncertainty. For this reason, the three-point spline array was selected here as the basis for subsequent inversions.

C. Environment model selection

Having established that the three-point spline had the highest Bayesian evidence of the seven array models evaluated, we now turn to evaluating evidence values for various environmental models. Figure 5 depicts schematically the six ocean-acoustic environments considered, including the simple Pekeris waveguide used as the baseline case above. Again, the choice of models was subjective—in this case based on previous geoacoustic inversions in the North Elba region[23,32,26,33] and also recent high-resolution surveys,[34,39] which have indicated near-surface stratification that might reasonably be detected by the present self-noise technique. The addition of sound-speed gradients to the water column, while seemingly unnecessary given the almost iso-velocity profile measured at sparse locations during the experiment, was of interest here because the impact of such profiles needs to be better understood in relation to near-field inversion.

Marginal PPDs generated by Gibbs sampling for each of the six environments—assuming the three-point spline array previously selected—appear in Figs. 10 to 15. In the case of model (a) in Fig. 10, it can be observed that the change in array parametrization has not significantly affected either the shapes of the PPDs or the parameter estimates obtained previously using a parabolic array. Similarly to Fig. 6, the source depth Z_s is observed to be multimodal, but becomes less so following the addition of a water column sound-speed gradient. For model (b) in Fig. 11, the velocity estimate for the lower water column was 1512 ± 6 m/s relative to the surface velocity estimate of 1518 ± 4 m/s, suggesting a significant deviation from the iso-velocity assumption made in Ref. 2.

Model (c) is again an iso-velocity model, with the bulk water sound-speed estimate C_w being 1516 ± 3 m/s. How-

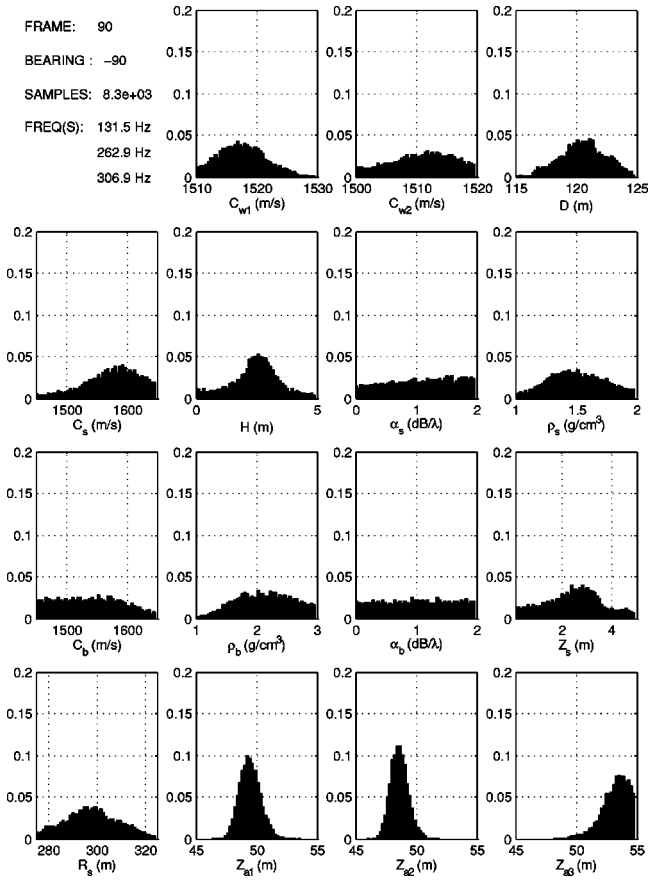


FIG. 13. Marginal PPD histograms for frame 90. Environment model (d), three-point array and MANNING source.

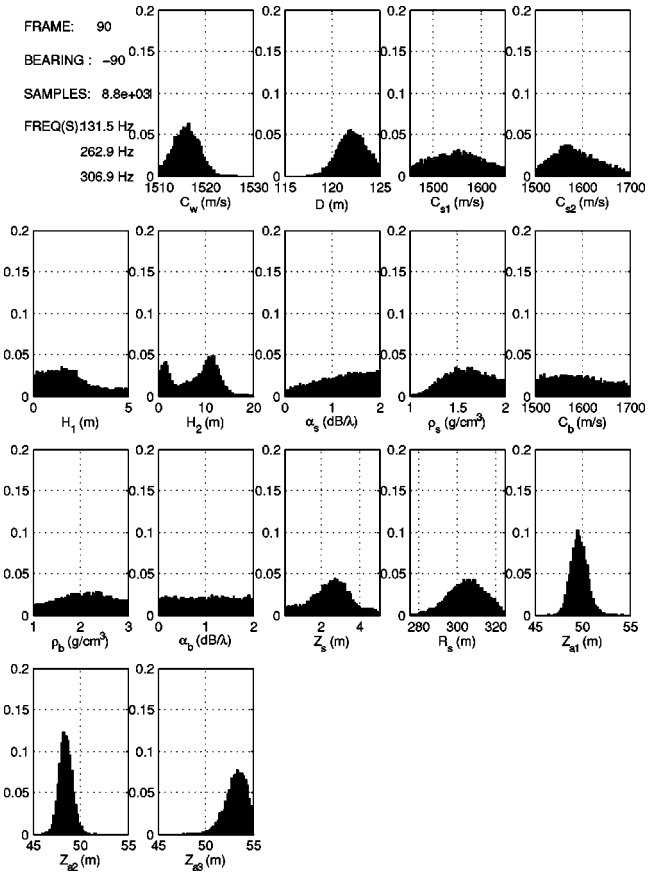


FIG. 14. Marginal PPD histograms for frame 90. Environment model (e), three-point array and MANNING source.

ever, in this case, the possibility of a sediment layer with its own density and *p*-wave velocity has been included. As is apparent in Fig. 12, the existence of such a sediment layer down to 2.3 ± 1.0 m is suggested by the well-resolved peak in the sediment thickness PPD. This feature was subsequently confirmed to be common to most of the MAPEX2000 dataset between frames 90 and 117, and its estimated depth correlates very well with bottom structures proposed by Gingras and Gerstoft[26] for the North Elba seabed, as does the estimated sediment density of 1.6 ± 0.2 g/cm^3. With the reintroduction of a water column sound-speed gradient, this peak remains unchanged; however, the sediment sound-speed estimate increases from 1565 to 1576 m/s while the sediment density estimate decreases to 1.5 ± 0.2 g/cm^3. It should also be noted that in comparison with the sediment parameters, those of the half-space below are poorly resolved, with the only meaningful observation being that the half-space density is likely to be in excess of 2 g/cm^3.

Figures 14 and 15 illustrate the effect of introducing still further structure to the bottom acoustic model in the form of an additional iso-velocity fluid sediment layer. The prior bound of 20 m for the thickness of this layer was intended to reflect what was believed to be the overall sensitivity limit of the MAPEX2000 inversion geometry over the frequency range used—though this itself is a function of the actual bottom structure. Whereas the PPD of the H_1 layer thickness parameter becomes substantially flatter after adding a second layer, the H_2 PPD appears distinctly multimodal, with a sec-

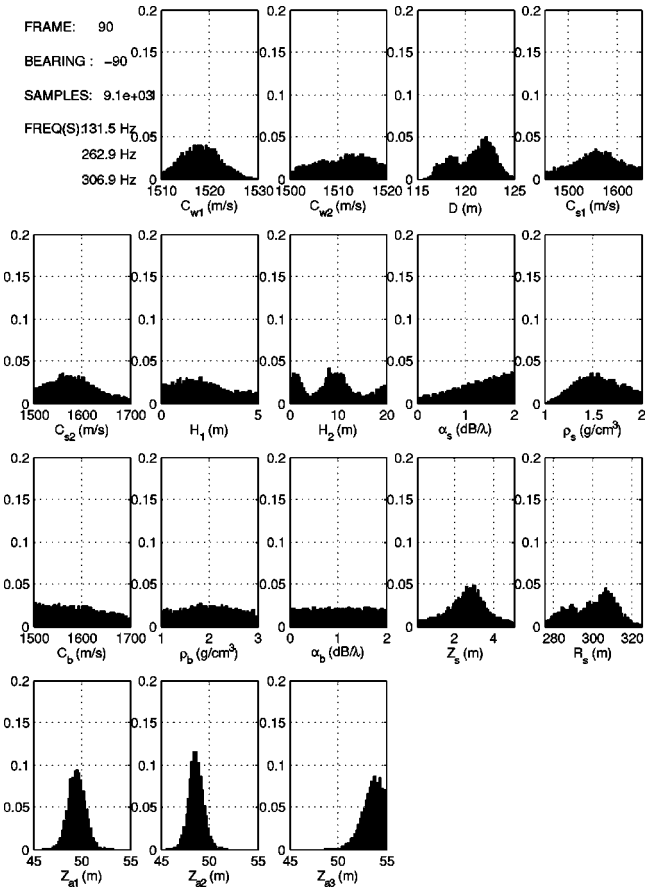


FIG. 15. Marginal PPD histograms for frame 90. Environment model (f), three-point array and MANNING source.

TABLE IV. Results for environment model selection. Model (d) is seen to give the highest evidence. Note that the table displays log evidence and the probability of a model is based on just the evidence; thus, the small differences for each model are important.

Model	Min E	Log evidence	
		RIS	Gaussian
(a)	19.4	−14.2	−13.8
(b)	19.1	−14.1	−13.7
(c)	19.7	−15.2	−14.7
(**d**)	**14.7**	**−13.6**	**−12.7**
(e)	15.0	−13.9	−13.1
(f)	14.9	−13.7	−13.1

ond peak at approximately 10 m. This is possibly on account of two actual sediment layers in the bottom, or also possibly just a resonant condition for the bottom interface reflectivity.

As done in Sec. V B for array model selection, the quantitative results of environment model selection are summarized in Table IV and Fig. 16. Unlike the previous case, it is difficult to rank competing environmental models *a priori* in relation to their parametric complexity, and hence the evidence does not behave monotonically either side of the optimal selection—model (d). The MAP and mean parameter estimates for this model, along with their associated standard deviations, represent the final output of the two levels of Bayesian inference employed in this paper, and are summarized in Table V.

Also unlike the previous case, the environment model found to have the highest Bayesian evidence was also that which achieved the best overall fit. Given the hierarchical relationship between models (d) and (f), it is at first surprising that the latter could not achieve a lower energy, though it is possible that the initial GA search simply failed to find the globally optimal solution. This is testimony to why it is more desirable to characterize PPDs in inverse problems rather than just search for the best fit. Further, it can be noted from Eq. (25) that the reverse importance sampling procedure used for accurate evidence estimation—unlike the Gaussian approximation of Eq. (9)—does not depend on the maximum

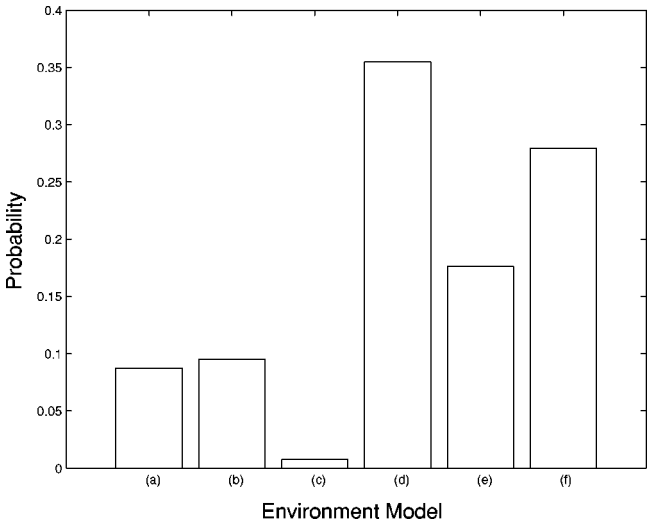


FIG. 16. Normalized probabilities for the environment models listed in Table IV based on Bayesian evidence.

TABLE V. Results of final geoacoustic inversion.

Parameter	Unit	MAP	Mean	Std. dev.
C_{w1}	m/s	1518	1518	4
C_{w2}	m/s	1512	1511	5
D	m	121	121	1.8
C_s	m/s	1598	1577	45
H	m	2.7	2.5	1
α_s	dB/λ	2.0	1.1	0.6
ρ_s	g/cm^3	1.5	1.5	0.2
C_b	m/s	1539	1540	51
ρ_b	g/cm^3	2.5	2.2	0.5
α_b	dB/λ	1.4	1.0	0.6
Z_s	m	2.0	2.5	1
R_s	m	301	300	10
z_{a1}	m	49.6	49.7	0.9
z_{a2}	m	48.7	48.8	0.7
z_{a3}	m	53.8	53.4	1.0

likelihood estimate $p(\mathbf{d}|\mathcal{M}_k)$. The evidence values in the second column of Table IV, therefore, are unbiased by suboptimal ML solutions, unlike those in the third column—although in any case, both methods identified model (d) as the most probable.

In relation to Fig. 16, there appears to be a consistent demarcation between environments with iso-velocity water columns as opposed to sound-speed gradients—the left and right columns of Fig. 5. The latter are always more probable than the former, with the difference between models (c) and (d) being almost anomalous. This suggests that even at the short ranges necessary for self-noise inversion, it is worthwhile modeling the water column as possessing at least a first-order sound-speed gradient.

Finally, on the basis of histogram plots such as those in Figs. 10 to 15, it is not easy to see the reasons for particular model preferences, but in numerous reruns, the essential pattern of Fig. 16 was fairly consistently reproduced. In any case, we stated in the Introduction that the focus of this paper was on the *methodology* of Bayesian model selection rather than exhaustive demonstration, which would need consistency checks across a more extensive portion of our data set than the single frame analyzed here, and also possibly a wider range of realistic ocean models.

VI. CONCLUSIONS

A Bayesian approach to solving self-noise inversion problems has been presented. This approach provides consistent methods for both parameter estimation at the first level of inference, and model selection at the second level of inference.

To implement the calculations necessary—namely PPD estimation, marginalization, and integration—we used a Metropolis–Hastings variant of the popular Gibbs sampler algorithm, which, combined with fast acoustic modeling, made the tasks of parameter estimation and model selection tractable. Variations on conventional Gibbs sampler design, such as an initial fast GA optimization, and coordinate rotation according to semianalytic covariance estimates proved successful in accelerating overall algorithm performance.

The second level of inference is concerned with model selection, i.e., the problem of selecting a model that best explains a data set at a given signal to noise ratio. Bayesian model selection, as discussed here, is based on calculating evidence, which is the integral of the product of likelihood and prior probability over all model parameters. This approach penalizes models that are more complex than warranted by the data by virtue of the Occam factor.

ACKNOWLEDGMENTS

This work was supported by NAVSEA/ASTO under Contract N00014-01-D-0043-D003 and by ONR under Grant N00014-03-1-0393.

[1] W. A. Kuperman, M. F. Werby, K. E. Gilbert, and G. J. Tango, ''Beam forming on bottom-interacting tow-ship noise,'' IEEE J. Ocean. Eng. **OE-10**(3), 290–298 (1985).
[2] D. J. Battle, P. Gerstoft, W. A. Kuperman, W. S. Hodgkiss, and M. Siderius, ''Geoacoustic inversion of tow-ship noise via near-field-matched-field processing,'' IEEE J. Ocean. Eng. **OE-28**(3), 454–467 (2003).
[3] S. E. Dosso, ''Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach,'' J. Acoust. Soc. Am. **111**, 129–142 (2002).
[4] S. E. Dosso and P. L. Nielsen, ''Quantifying uncertainty in geoacoustic inversion. II. Application to broadband, shallow-water data,'' J. Acoust. Soc. Am. **111**, 143–159 (2002).
[5] S. E. Dosso, ''Environmental uncertainty in environmental source localization,'' Inverse Probl. **19**, 419–431 (2003).
[6] D. J. C. MacKay, ''Bayesian interpolation,'' Neural Comput. **4**, 415–447 (1992).
[7] A. Malinverno, ''Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem,'' Geophys. J. Int. **151**, 675–688 (2002).
[8] J. J. K. Ó Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing* (Springer, New York, 1996).
[9] A. Malinverno, ''A Bayesian criterion for simplicity in inverse problem parametrization,'' Geophys. J. Int. **140**, 267–285 (2000).
[10] H. Jeffreys, *Theory of Probability* (Oxford University Press, Oxford, 1939).
[11] K. Mosegaard and M. Sambridge, ''Monte Carlo analysis of inverse problems,'' Inverse Probl. **18**, 29–54 (2002).
[12] C. F. Mecklenbräuker, P. Gerstoft, J. F. Bohme, and P. Chung, ''Hypothesis testing for geoacoustic environmental models using likelihood ratio,'' J. Acoust. Soc. Am. **105**, 1738–1748 (1999).
[13] C. Andrieu and A. Doucet, ''Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC,'' IEEE Trans. Signal Process. **47**(10), 2667–2676 (1999).
[14] W. K. Hastings, ''Monte Carlo sampling methods using Markov chains and their applications,'' Biometrika **87**, 97–109 (1970).
[15] S. Geman and D. Geman, ''Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images,'' IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984).
[16] A. E. Gelfand and A. F. M. Smith, ''Sampling-based approaches to calculating marginal densities,'' J. Am. Stat. Assoc. **85**, 398–409 (1990).
[17] A. F. M. Smith and G. O. Roberts, ''Bayesian computation via the gibbs sampler and related Markov chain Monte Carlo methods,'' J. R. Stat. Soc. Ser. B. Methodol. **55**, 3–23 (1993).
[18] M. K. Sen and P. L. Stoffa, ''Bayesian inference, Gibbs sampler and uncertainty estimation in geophysical inversion,'' Geophys. Prospect. **44**, 313–350 (1996).
[19] W. A. Kuperman, M. D. Colins, J. S. Perkins, and N. R. Davis, ''Optimal time-domain beamforming with simulated annealing including application of *a priori* information,'' J. Acoust. Soc. Am. **88**(4), 1802–1810 (1990).
[20] M. D. Collins and W. A. Kuperman, ''Focalization: Environmental focusing and source localization,'' J. Acoust. Soc. Am. **90**, 1410–1422 (1991).
[21] P. Gerstoft, ''Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions,'' J. Acoust. Soc. Am. **95**, 770–782 (1994).
[22] P. Gerstoft, '' Inversion of acoustic data using a combination of genetic algorithms and the Gauss-Newton approach,'' J. Acoust. Soc. Am. **97**, 2181–2191 (1995).
[23] P. Gerstoft and D. F. Gingras, ''Parameter estimation using multifrequency range-dependent acoustic data in shallow water,'' J. Acoust. Soc. Am. **99**, 2839–2850 (1996).
[24] J. Hermand and P. Gerstoft, ''Inversion of broadband multitone acoustic data from the Yellow Shark summer experiments,'' IEEE J. Ocean. Eng. **OE-21**(4), 324–346 (1996).
[25] P. Gerstoft and C. F. Mecklenbräuker, ''Ocean acoustic inversion with estimation of *a posteriori* probability distributions,'' J. Acoust. Soc. Am. **104**, 808–819 (1998).
[26] D. F. Gingras and P. Gerstoft, ''Inversion for geometric and geoacoustic parameters in shallow water: Experimental results,'' J. Acoust. Soc. Am. **97**, 3589–3598 (1995).
[27] C. F. Mecklenbräuker and P. Gerstoft, ''Objective functions for ocean acoustic inversion derived by likelihood methods,'' J. Comput. Acoust. **8**(2), 259–270 (2000).
[28] P. Gerstoft, SAGA Users guide 2.0, an inversion software package, SACLANT Undersea Research Centre, SM-333, 1997. http://www.mpl.ucsd.edu/people/gerstoft/saga
[29] M. D. Collins and L. Fishman, ''Efficient navigation of parameter landscapes,'' J. Acoust. Soc. Am. **98**, 1637–1644 (1995).
[30] G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey, ''Mirages in shallow water matched field processing,'' J. Acoust. Soc. Am. **105**, 3245–3265 (1999).
[31] A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, ''Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem,'' J. Acoust. Soc. Am. **83**, 571–587 (1988).
[32] F. B. Jensen, ''Comparison of transmission loss data for different shallow water areas with theoretical results provided by a three-fluid normal-mode propagation model,'' in *Sound Propagation in Shallow Water*, edited by O. F. Hastrup and O. V. Olesen, CP-14 (SACLANT ASW Research Centre, La Spezia Italy, 1974), pp. 79–92.
[33] D. D. Ellis and P. Gerstoft, ''Using inversion techniques to extract bottom scattering strengths and sound speeds from shallow-water reverberation data,'' in Proc. ECUA96, 1996, pp. 557–562.
[34] C. W. Holland and J. Osler, ''High-resolution geoacoustic inversion in shallow water: A joint time- and frequency-domain technique,'' J. Acoust. Soc. Am. **107**, 1263–1279 (2000).
[35] D. D. Ellis and D. M. F. Chapman, ''A simple shallow water propagation model including shear wave effects,'' J. Acoust. Soc. Am. **78**, 2087–2095 (1985).
[36] Z. Y. Zhang and C. T. Tindle, ''Complex effective depth of the ocean waveguide,'' J. Acoust. Soc. Am. **93**, 205–213 (1993).
[37] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994).
[38] E. K. Westwood, C. T. Tindle, and N. R. Chapman, ''A normal mode model for acoustoelastic ocean environments,'' J. Acoust. Soc. Am. **100**, 3631–3645 (1996).
[39] C. W. Holland, ''Regional extension of geoacoustic data,'' in 5th *Europ. Conf. Underwater Acoustics*, edited by Zakharia, Chevret, and Dubail (European Commission, Luxembourg) 2000, pp. 793–796.

Testing low/very low frequency acoustic sources for basin-wide propagation in the Indian Ocean

Donna K. Blackman[a] and Catherine de Groot-Hedlin
Scripps Institution of Oceanography, La Jolla, California 92093

Phil Harben
Lawrence Livermore National Labs, Livermore, California 94550

Allan Sauter and John A. Orcutt
Scripps Institution of Oceanography, La Jolla, California 92093

(Received 29 August 2003; revised 2 June 2004; accepted 9 July 2004)

Low/very low frequency acoustic signals were transmitted to distant receivers in the Indian Ocean. The aim was to test methods for characterizing the hydroacoustic capability of the International Monitoring System (IMS) that discriminates for nuclear tests in the region. Several acoustic sources were deployed between Seychelles and Fremantle, Australia, and the IMS receivers comprised a network of hydrophones off Diego Garcia and Australia. Two of the three acoustic sources tested produced basin-scale propagation of impulsive signals. Single glass spheres imploded within the sound channel produced a clear signal at frequencies above ~40 Hz, at ranges of hundreds to a thousand kilometers. Five-sphere glass implosions were recorded at ranges up to 4400 km. Near-sea surface shots from a large airgun array were recorded in several cases at ranges of hundreds to thousands of kilometers, the frequency of the highest signal-to-noise ratio arrivals varied within the 5–100 Hz band. High background noise level was a key factor at IMS stations that did not detect the airgun signals in the 5–15 Hz band. In a few cases, details of bathymetric features that are not well represented in the digital elevation model contributed to unexpected variation in relative signal levels between IMS stations. [DOI: 10.1121/1.1786711]

PACS numbers: 43.30.Qd [WLS] Pages: 2057–2066

I. INTRODUCTION

Hydroacoustic monitoring for nuclear explosions in the Indian Ocean is a goal for which detailed testing is still under way. The signal from a large acoustic source in the Indian Ocean basin would almost certainly be detected by the monitoring hydrophones, if not directly then via reflections off topographic slopes behind blocked areas (Pulli *et al.*, 1999; Hanson *et al.*, 2002). The challenge is to accurately determine the location of the source and to discriminate between natural, legitimate manmade, or clandestine nuclear explosion causes for the signal, so that appropriate response is undertaken (Hanson *et al.*, 2001).

Determination of a hydroacoustic source location depends on models of both sound propagation within the ocean and seafloor topography, which can attenuate the signal as it travels along the path between source and receiver. Differences in oceanographic properties (temperature, salinity) result in variable sound speed structure and this affects the travel time of acoustic signals. For basin-wide propagation, the low velocity ''sound channel'' is responsible for minimizing losses to allow detection of very distant sources. Signals that propagate along paths where topographic features protrude into the sound channel, or where the source or receiver is near a continental shelf or slope, lose energy due to partitioning of phases at the seafloor interface. Diffraction and scattering at topographic features can alter the path and contribute to complexity in the wave form that is recorded at a receiver. Therefore, models of propagation and loss employed in event location and characterization of an unknown source depend on databases of oceanic properties and topography. The required resolution of these databases, particularly the topography, is one of the things that remain to be determined for the Indian Ocean International Monitoring System (IMS). In addition, while recent modeling techniques are able to incorporate most aspects of underwater sound propagation and loss, there is a trade-off between the level of modeling complexity and the resources required for the calculation. For real-time monitoring, one would like to minimize the computational effort while retaining sufficient accuracy to obtain results within the desired limits. One way to assess which factors control variability in the acoustic propagation/losses for one region versus another is to transmit a known source signal from various locations within an ocean basin. Differences in the recorded signal between stations can guide plans for additional mapping to enhance the database or efforts to refine numerical modeling approach.

A fundamental question arises when undertaking to quantitatively assess the resolution of databases employed in very low/low frequency (VLF/LF) sound propagation models within an ocean basin—what acoustic sources can provide the information needed while ensuring a feasible series of experiments? Human and marine mammal safety is a key concern and regulation of international shipping and handling of explosives is another difficulty. We deployed two types of sources that were nonexplosive. Small implosives

[a]Author to whom correspondence should be addressed; electronic mail: dblackman@ucsd.edu

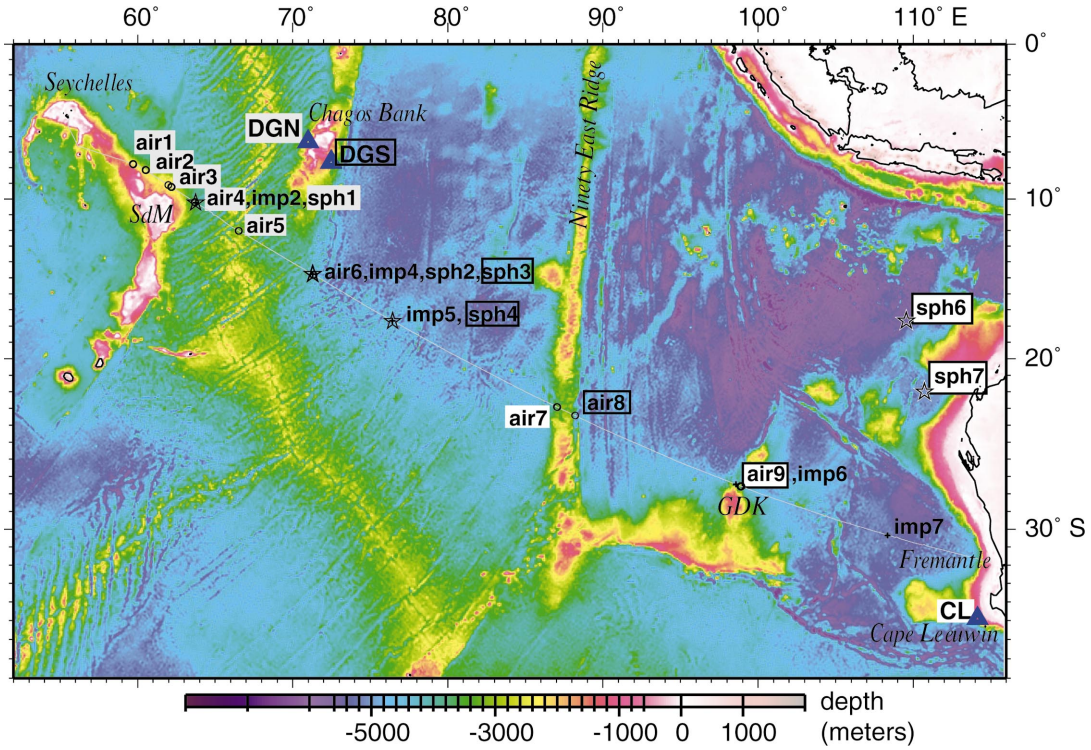


FIG. 1. Topographic model of Smith and Sandwell (1997) in the Indian Ocean region is overlain by 2001 R/V EWING shiptrack (gray line) and the acoustic source locations for our experiment. Yellow-red colors illustrate shallow bathymetric features that attenuate signals propagating in the sound channel across the basin. IMS hydrophone stations are shown by blue triangles. Circles show where the airgun array shooting (*air*) was conducted. Stars show glass sphere (*sph*) implosion stations and five-sphere implosions are highlighted in white. Crosses are where the MPL/SIO imploding cylinder (*imp*) was fired. Gray-fill boxes indicate sources observed at IMS hydrophones at Diego Garcia, north (DGN), black-outline boxes indicate sources observed at Diego Garcia, south, (DGS), and white-fill boxes show sources observed at Cape Leeuwin (CL). SdM indicates Saya de Malha bank; GDK is Gulden Draak Knoll.

were fired as single-shots in the water column at depths of 300–1200 m. The other source type was a large airgun array, such as is commonly employed for marine seismic reflection surveys conducted by oil industry and academic researchers. This system was fired repeatedly near the sea surface (9–12 m depth) at selected source sites.

Our work in the Indian Ocean took advantage of available shiptime in 2001 and a sparse network of hydrophones that are part of the IMS overseen by the Comprehensive Test Ban Treaty Office of the United Nations. The research ship R/V EWING departed Mahe Island, Seychelles, and steamed essentially a great circle path to Fremantle, Australia. During the 18-day cruise, the Ewing's airgun array was used nine times and implosive sources were deployed twelve times, of which nine successfully fired (Fig. 1). During the following cruise off northwest Australia, two additional, larger, implosive sources were successfully fired.

The results reported here are initial steps toward the overall goal of fully characterizing the influence on acoustic propagation throughout the Indian Ocean Basin of topographic blockage, diffraction, and scattering, and oceanographic complexities such as the high temperature and salinity gradients near the Antarctic Convergence. We have determined the range over which small and moderate sized implosive signals (Secs. III A and III B) will propagate with sufficiently low loss to be detected at distant hydrophones (Sec. III C). In addition, we discuss the relationship between near-source seafloor topography (Secs. IV B and V) and the efficiency of coupling acoustic energy from near-sea surface sources (Sec. IV A) into the sound channel for subsequent basin-scale propagation. We assess the accuracy with which the recorded airgun array signals can be used to calculate the location of near-sea surface sources (Sec. IV C). Finally, the differences between expected losses and the level of the observed signal-to-noise ratio for different source-receiver paths are noted (Sec. V) and possible causes for the difference are discussed.

II. RECEIVERS

A. IMS hydrophones

The hydrophone stations of the International Monitoring System that were operating in the Indian Ocean in late 2001 are located near Diego Garcia and off Cape Leeuwin, southwestern Australia (Fig. 1). The two stations in the vicinity of Diego Garcia—DGN and DGS—are each comprised of 3 hydrophones deployed 1–2 km apart. DGN hydrophones are deployed at water depths of 1180–1250 m on the western slope of the Chagos Bank, on which the island of Diego Garcia is located. The instruments for DGS are at water depths of 1350–1415 m on the eastern slope of Chagos Bank. The Cape Leeuwin (CL) hydrophones are deployed at water depths of 1045–1065 m. Digital data are recorded at 250 samples/s and the instrument response dictates useful data for our analyses in the 2–100 Hz band.

The background noise levels are different at the three IMS hydrophone stations (Fig. 2). During our October cruise, the southwest monsoon ensured steady head winds

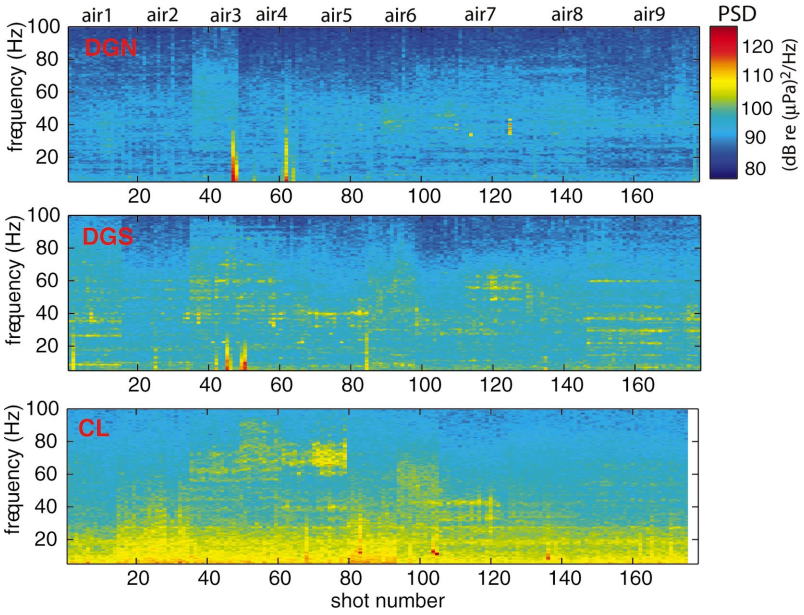


FIG. 2. Noise levels at the IMS stations during the times of airgun array shooting. Spectragrams show 10 s window that immediately precedes each shot using data taken from a representative sensor at each station. Note that time is not continuous along the horizontal axis (see Table I and Fig. 1). The shooting sites are labeled at top.

and seas from the east that, in combination with swell traveling from the Southern Ocean, resulted in higher noise at DGS than at DGN. Cape Leeuwin background noise levels varied the most but were consistently higher than DGS during our study. This reflects storms and generally high sea states in the Southern Ocean. A series of distant earthquakes briefly increased the background levels on several days at Diego Garcia (for example, during *air3* shots 45–48, jd 281).

B. Near-source hydrophone

A calibrated hydrophone was deployed from the ship's deck to document the near-source signal for each implosion and a few representative airgun array shots. Source location and time, accurate to a second, were obtained from shipboard GPS data. More detailed timing was obtained from the near-source hydrophone data. The arrival times of the direct and sea surface reflected signals recorded by the hydrophone were picked with a precision of 0.001 s and the time difference between the two phases was used to determine the depth of the hydrophone, using a local water velocity profile. The depth of the implosive source was determined from the wire-out reading on the winch and was usually accurate to within a meter or two. Spectral estimates are too noisy to provide independent confirmation of implosion depth at meter-scale resolution. The average local velocity for the depth interval between the hydrophone and the source was used to compute the absolute source time by differencing from the direct arrival time at the hydrophone. The signal level was corrected back to a 1 m distance from the source assuming linear decrease of pressure with radial distance from a point source. The overall uncertainty of the implosive source time, considering all these factors, is typically ± 0.01 s.

III. IMPLOSIONS

A. LLNL imploding glass spheres

Imploding sources have a relatively long history of use (Isaacs and Maxwell, 1952; Orr and Shoenberg, 1976). Recent work shows that they can be made reliable and can be designed to produce a signal that is adequate for long-range calibration (Sauter and Dorman, 2003; Harben *et al.*, 2000). Two versions of glass sphere imploding sources were employed for our experiment—a single-sphere and a five-sphere device (Harben *et al.*, 2000; Harben and Boro, 2001). The single 22 l glass sphere implodes when a piston driven smashing system triggers at a predetermined depth through failure of a rupture disk. The five-sphere device uses the same piston driven smashing system to break a central sphere. The surrounding four spheres fail as a direct result of the first one. The single sphere produces a signal [Fig. 3(a)] with an initial, relatively low frequency rarefactional pulse that corresponds to the inrush of water caused by failure of the glass container. Convergence at the sphere center is the instant of shock wave formation, giving rise to the spike that

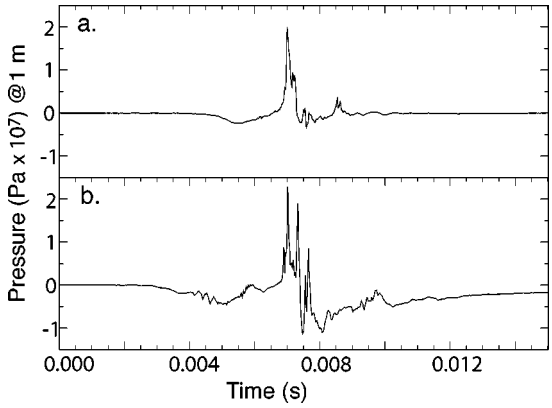


FIG. 3. Near-source recording of glass sphere implosions at 680 m depth, corrected to 1 m range from the shots. (a) Single glass sphere implosion; (b) five-sphere implosion.

dominates the record; a small bubble pulse follows. The record in Fig. 3(b) shows a composite collapse of five spheres during the initial rarefactional pulse of the wavetrain. Short duration features riding on the longer rarefactional pulse are consistent for both of the five-sphere source wave forms we recorded but the reasons for these features are not clear. The shock wave from the first implosion is visible, overlain, and followed by the implosion shock waves from the other spheres. The number and pattern of these spikes were significantly different for the two five-sphere source records. The time differences are consistent with failure caused by fracture propagation from sphere to sphere but are not consistent with a shock wave driven by collapse of adjacent spheres from the first implosion.

The imploding glass spheres achieve maximum levels in the 300–500 Hz range of 250–270 dB *re* 1 μPa, corrected back to 1 m from the source, but the level at other frequencies is much reduced. At the source, the pulse is about 6 ms long.

B. MPL/SIO wireline imploder

The 20 l, cylindrical configuration of the MPL/SIO imploder (Sauter and Dorman, 2003) produces a signal whose level in the 100–200 Hz band is 5–10 dB lower than the single glass sphere signal. We were not able to recognize any of the imploder shots in the IMS hydrophone data.

C. IMS hydrophone recordings

Five of the six successful imploding glass sphere shots were detected at IMS hydrophone stations (Fig. 1). Three were single sphere implosions, recorded at Diego Garcia, and two were five-sphere implosions, recorded at both Diego Garcia and at Cape Leeuwin. All detected shots had source depth ~680 m; a 320 m deep implosion (*sph2*) fired at the same location as a deeper shot (*sph3*) was not detected. Ranges between the single-sphere shots and the receiver were 800–1200 km. The single sphere signal level rises above the noise level for frequencies greater than about 40 Hz (Fig. 4). The duration of the received signal is quite a bit longer than the near-source pulse at almost 400 ms.

The bandwidth of the five-sphere arrivals is similar to that of the single-sphere recordings. Peak amplitudes are not right at the onset of the arrival but are consistently in front of the center of the envelope whose length is about 1 s (Fig. 5). The five-sphere implosions were conducted along the western Australia continental margin with ranges to Diego Garcia of 4168 km (*sph6*) and 4397 km (*sph7*) and ranges to Cape Leeuwin of 1960 and 1465 km (Fig. 1).

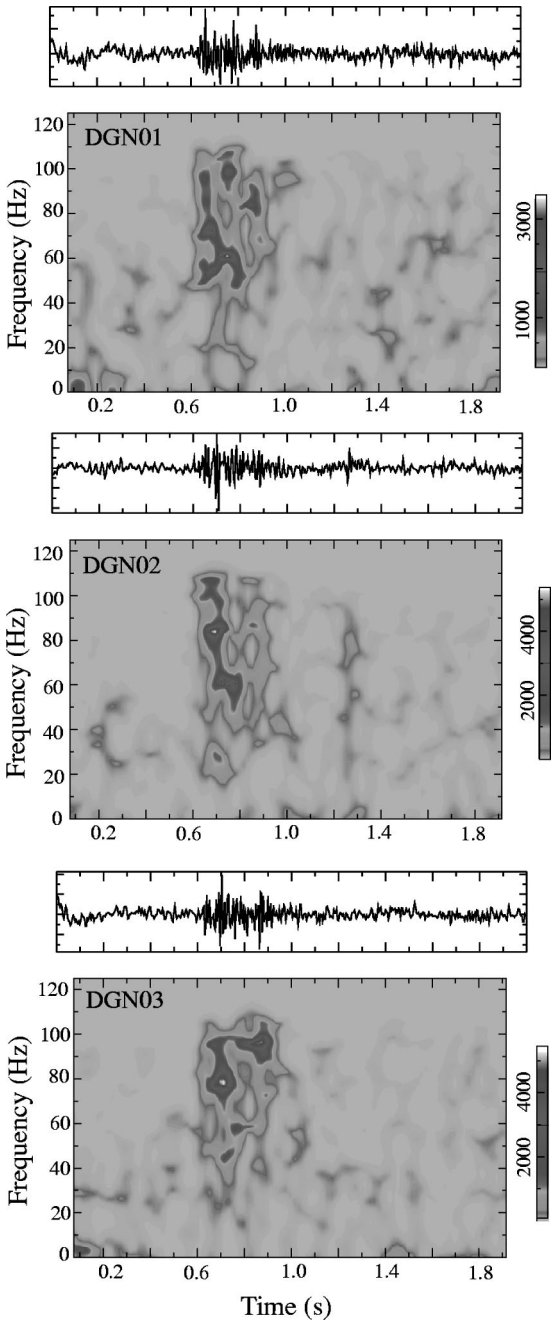


FIG. 4. Spectragrams of the signal from single-glass-sphere implosion *sph1* recorded at each sensor of station DGN. Note the short signal duration and concentration of energy at frequencies above 40 Hz. Unfiltered time sections are above each spectragram (instrument counts, grayshade, show relative level of signal and noise). Time zero corresponds to travel-time at 1.49 km/s for range of 798 km to DGN01.

IV. AIRGUN ARRAY SHOTS

A. Source

The R/V EWING array of 20 airguns with a total volume of 8465 cubic inches was used.[1] Each of our shooting periods lasted about 30 min, and shot interval varied from 57 to 173 s. Ship speed during the shooting was ~4 knots. The peak level of the airgun array signal was measured at 230–240 dB *re* 1 μPa (adjusted to 1 m range) in the 5–60 Hz range (Fig. 6). The duration of the airgun array pulse is about 12 ms. Source depth for the airgun array shots is 9–12 m and the details for each shooting site are listed in Table I.

Shot trigger times are generally accurate to a millisecond. Shot locations generally are accurate to about 10 m; loss of high resolution (P Code) GPS during the last shooting period resulted in greater uncertainty, ~50 m. Corrections for the difference in location between the GPS receiver and the center of the airgun array are incorporated in the airgun shot locations listed in Table I.

Seafloor topography in the vicinity of each shot controls

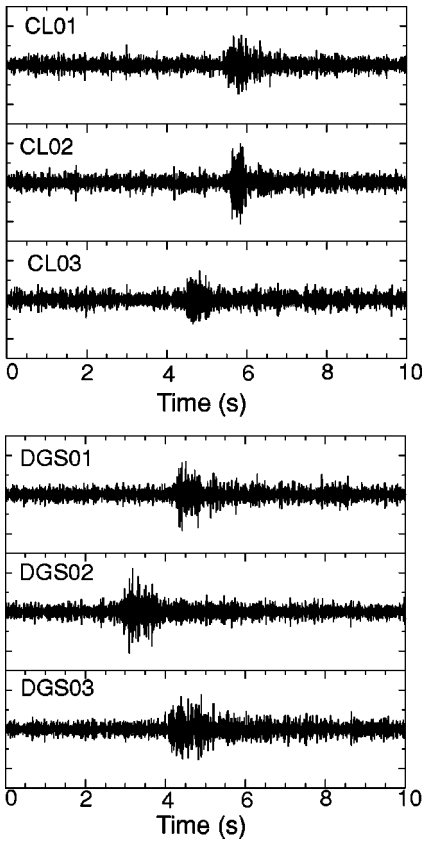


FIG. 5. Recordings of the five-sphere glass ball implosion (*sph7*) at distant IMS hydrophones. Cape Leeuwin time series are shown in the top panels and are high pass filtered with corner at 40 Hz; Diego Garcia station DGS time series are shown in the lower panels with the same filter applied. Time zero corresponds to travel time at 1.49 km/s for range of 4396 and 1465 km to DGS01 and CL01, respectively.

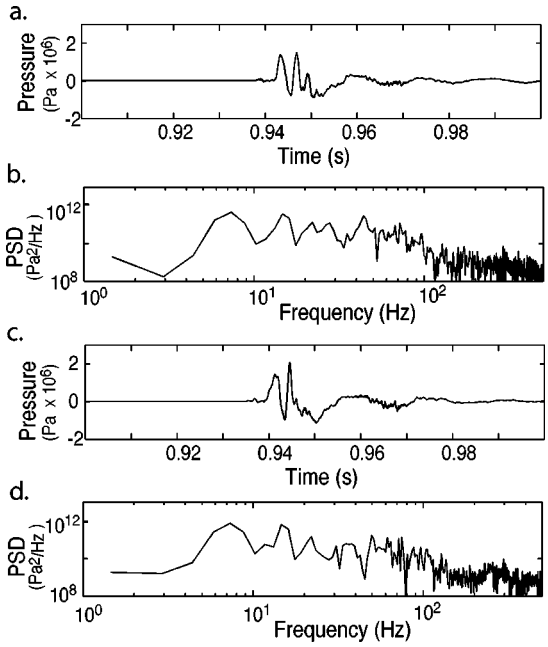


FIG. 6. Time series (a), (c) and power spectral density (b), (d) of two representative airgun array shots as recorded on the near-source hydrophone. Levels are corrected to a distance of 1 m from the source. The source signature is quite consistent in the 5–25 Hz band, with levels differing by less than a few decibels between shots. Variability that occurs in the 25–100 Hz band is illustrated by these two shots.

the efficiency of coupling the airgun energy into a nearly horizontally propagating wave that is guided long distance by the sound channel. Details of the local depth, slope, and seafloor roughness all contribute to this process of transferring energy from the dominantly down-going pulse[1] into horizontally traveling rays.

B. IMS hydrophone recordings

The signal recorded for our series of airgun shooting sites varied as a function of both receiver and source location. Some of this variability was expected as blockage or near-source seafloor conditions changed, but in some cases the increase or decrease in SNR between sites/stations was surprising. Chagos bank blocked propagation to DGS for the first four airgun sites. We expected CL to be blocked from the first site by the Saya de Malha bank and from some of the subsequent five sites, by the Ninety East Ridge (Fig. 1). Propagation from shooting sites *air6–air9* to DGN was blocked by the Chagos bank.

Figures 7–9 illustrate the difference in arrival character between airgun shooting sites and between receiving stations. The records for the source–receiver pairs shown in Figs. 7 and 8(a)–(c) have sufficiently high SNR that the individual shots are visible in filtered sections. Stacking of all shots for a given site, following moveout correction (a time shift that takes into account the small difference in source–receiver range due to ship motion during shooting), increases the SNR of the main arrival and retains features common to the 5–10 Hz bandpassed wave form for the entire shooting period; this is what is shown in Fig. 8. Stacking is ineffective for frequencies greater than 30 Hz; we infer that in this band acoustic energy is more strongly influenced by details of the seafloor topography. The seafloor changes somewhat as the ship progresses along the track within a single site (Fig. 10; ship position cannot be held when the airguns are deployed/firing), so details of the wave form can vary. The recordings of *air2* shots have high SNR in the 5–10 Hz range and the time series data show that the wave train following the initial pulse is short [$\sim$2 s; Fig. 8(a)]. The *air5* shots have high SNR in the 5–40 Hz band [Fig. 7(a)] and the wave duration after the main pulse exceeds 5 s in the 5–10 Hz band [Fig. 8(b)]. Although both had good SNR, transmission loss between the airgun source and the receivers was significantly lower for *air5* shots than for *air2* shots (Fig. 9).

Both DGS and CL recorded the shots for *air9* [Figs. 8(c), 8(d), Fig. 9]. In the 40–60 Hz range, DGS recordings exceed the noise level significantly but the wave forms are not very similar, therefore stacking does not significantly improve SNR. In the 5–10 Hz band, the stacked DGS wave forms are somewhat consistent between the individual hydrophones, showing a narrow peak about 3 s from the onset of the arrival [Fig. 8(c)]. At CL, the record contains significant energy in a broader frequency band for the *air9* shots—30–60 Hz (Fig. 9). The *air9* signals are not discernible in the individual shot records in the 5–10 Hz range, however the arrival is apparent in the stacked section [Fig. 8(d)] where SNR is 2 or less. The stacked *air9* wave forms are not consistent between the individual CL hydrophones.

TABLE I. Airgun array shot lines (start/end).[a]

jd hr:mn	Lat (S)	Ion (E)	Seafloor	Interval	Site	Shot No.
280 19:14	7°46.1′	59°41.6′	2110 m	97 sec	air1	1–9
280 19:27	7 46.2	59 42.5	2145			
280 19:33	7 46.3	59 42.9	2205	173		11–14
280 19:42	7 46.5	59 43.4	2270			
281 02:38	8 8.5	60 30.7	2575	127	air2	22–35
281 03:03	8 9.5	60 32.1	2655			
281 16:48	9 13.7	62 09.8	2645	173	air3	36–46
281 17:17	9 14.7	62 11.5	2720			
282 05:22	10 12.2	63 40.9	3540	57	air4	49–59
282 05:31	10 12.6	63 41.6	3530			
282 05:38	10 12.7	63 41.8	3520	173		61–68
282 05:58	10 13.4	63 42.8	3660			
283 04:33	12 01.8	66 30.6	2965	127	air5	69–83
283 05:03	12 02.6	66 32.2	3330			
284 15:34	14 48.3	71 19.5	4850	173	air6	86–96
284 16:03	14 49.5	71 20.8	4875			
289 02:11	22 55.5	87 02.8	3370	127	air7	99–114
289 02:43	22 54.6	87 00.7	3455			
289 02:45	22 54.6	87 00.7	3445	57		115–128
289 02:58	22 54.3	87 00.0	3550			
289 12:28	23 25.0	88 11.3	2985	173	air8	129–139
289 12:57	23 25.6	88 13.3	3240			
289 12:59	23 25.6	88 13.5	3235	57		140–146
289 13:04	23 25.7	88 13.9	3105			
292:02:10	27 33.2	98 51.6	2710	57	air9	147–170
292:02:32	27 33.7	98 53.0	2800			
292:02:35	27 33.8	98 53.2	2810	127		171–178
292:02:46	27 34.0	98 53.9	2845			

[a]Detailed listing of individual shots available at AIP online archive.

All of the airgun array shots to the west of DGN (*air1–air5*) were visible in the recordings at that site (Fig. 9). The transmission loss was lowest in the 5–25 Hz band for *air5* shots, which were both the closest and had least blockage. A consistent, 5–10 dB lower signal level was observed in DGN recordings for shooting periods *air1–air4* in this VLF band. This difference is several decibels greater than is expected solely based on the difference in range (3 dB). In the 50–100 Hz band, *air1* had the lowest transmission loss at DGN. This site had the shallowest (1900–2200 m) depths and a steady NNE slope. No clear signals of shots *air6–air9* were evident at DGN.

Signal levels exceed noise by several decibels for airgun array shots observed at DGS from sites to the east (*air8–air9*, greater loss at *air7*; Fig. 9). The closest unblocked site (*air6*) did not produce signals at DGS due to deep and dominantly smooth seafloor [Fig. 10(c)] that cannot effectively entrain energy into the sound channel. The signal from *air9* recorded at DGS is clearly visible at a whole-basin-scale source–receiver range of about 4825 km.

The only shots that were clearly visible in the VLF band at Cape Leeuwin were the shots of *air9*, at a range of 1665 km. Blockage of *air8* shots is not known to be significant and CL noise levels during those shots were similar to levels during the *air9* shots so it is not clear why these shots were not observed. The *air7* shots occurred during a somewhat noisier period (Fig. 2) and over the west-dipping slope of the Ninety East Ridge that was not as favorable for entrainment in the sound channel in the direction of CL. However, some energy did propagate the full range. Calculated transmission loss for *air7* is similar to that for *air9* in the 20–90 Hz band where signal was more than 3 dB above background noise (Fig. 9).

C. Determination of near-sea surface source location

None of our shots were recorded at all three stations so we cannot triangulate, in the classic sense, to determine the location of the source. However, beam-forming provides constraints on the back-azimuth when the SNR is sufficiently high in the 5–10 Hz band. Attempts at beam-forming for higher frequencies were not productive. In the time domain, uncertainty in relative hydrophone location ($\sim$10 m, Lawrence, 2003, corresponding to 1/15th of a wavelength at 10 Hz) degrades results from the cross correlation at higher frequency. The spectral estimates are not robust enough above 10 Hz to allow beaming in this domain. Only DGN had sufficiently high SNR in the 5–10 Hz band. Back-azimuths, computed by cross correlating arrivals over 10 s windows at this station have mean errors of 0.45° for *air1–air4*, for all shots with a SNR larger than 1.5. This uncertainty in back-azimuth would yield a location uncertainty of 6–10 km at the shot ranges (known here, which would not be the case for a clandestine test), which vary from 800 to 1260 km.

The cross-correlation estimates of the back-azimuths at DGN for the *air5* sequence were off by as much as 30°, even though the SNR was high in the 5–10 Hz band. This most likely is due to the lengthy coda observed for these shots. As an alternate approach, we determined the azimuth by picking

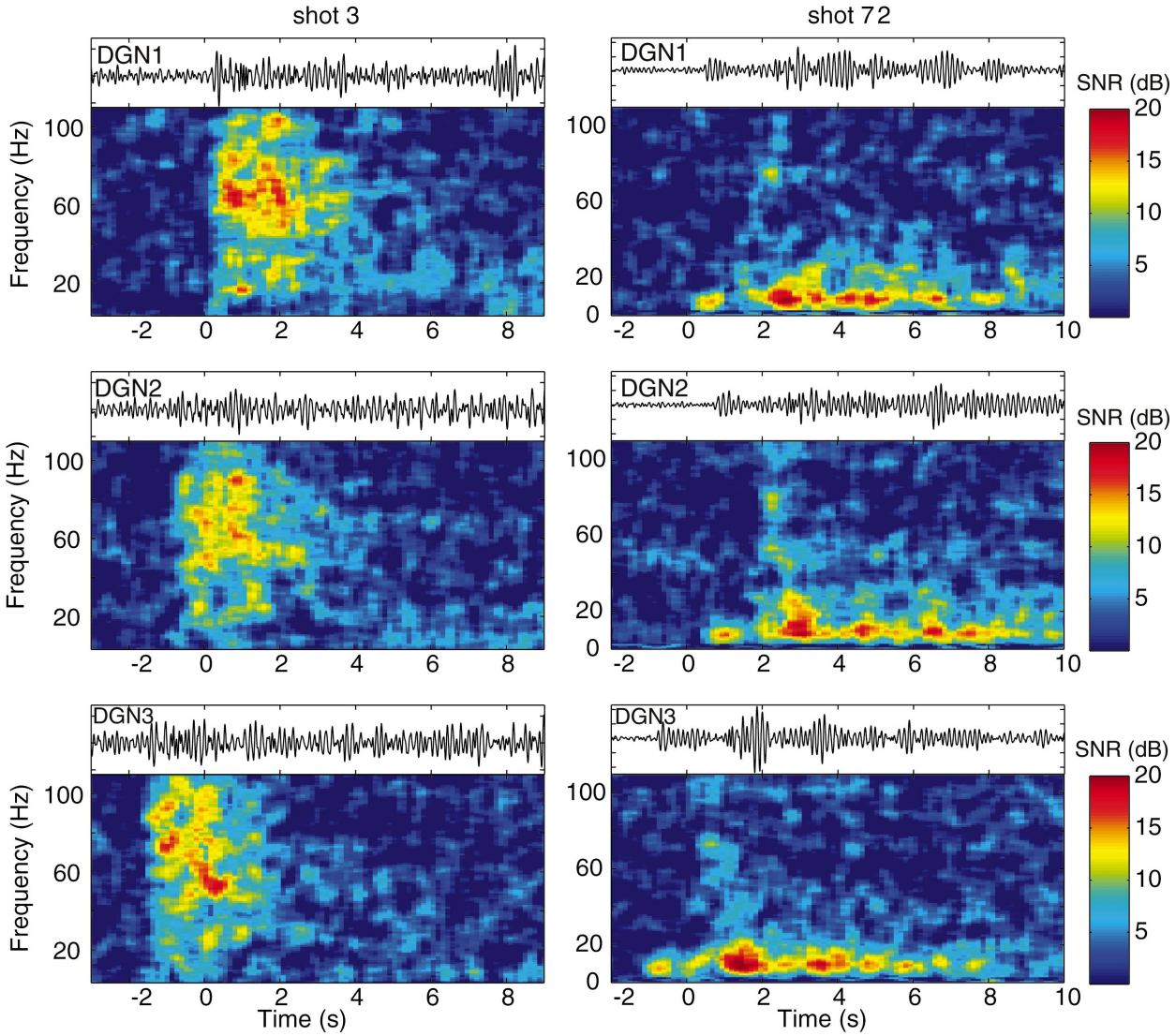


FIG. 7. Examples of individual airgun array shot recordings at the three hydrophones of DGN. Panels on the left show shot 3 from *air1* site. Panels on the right show shot 72 from *air5* site. Unfiltered times series data are plotted above corresponding spectragram in each panel. The spectragrams show signal-to-noise ratio as a function of time and frequency. The spectrum of the signal for each 1 s window is divided by the noise spectrum computed for a 5 s window immediately prior to the shot. Overlap between windows is ~0.9 s.

onset times and solving for the back-azimuths that yield the observed apparent velocity for each hydrophone pair, assuming plane wave propagation. This requires highly accurate time picks because the sensors are so closely spaced. As shown in Fig. 11 for one of the *air5* shots, the initial arrivals at each sensor are highly correlated, so that travel time picks can be made to within 0.05 s. These picks yield a computed back-azimuth with a mean error of 1°, which yields a location uncertainty of 14 km at a distance of 800 km. The later arrivals within the coda are poorly correlated, which suggests that they are being scattered to the receivers along other travel paths. This is consistent with the fact that these shots were located immediately to the east of the Central Indian Ridge (CIR) over a region of complex bathymetry.

The *air9* shot sequence was observed at two stations, which in theory should allow for improved location capability. However, the SNR is too low in the 5–10 Hz band for beam-forming. At higher frequencies, travel time picks cannot be made with sufficient accuracy to make reliable estimates of the apparent velocity, and hence back-azimuths, for either CL or DGS.

Installation of the station at Crozet Island, in 2003, should improve the IMS ability to locate acoustic sources throughout the Indian Ocean. Both the improved triangulation geometry and the added ranging information from the new station will contribute.

V. DISCUSSION

There are three principal methods by which acoustic energy becomes entrained in the oceanic sound channel: direct excitation; downslope propagation; and scattering. Sources near the sea surface, like the airgun shots recorded in this experiment, do not directly excite energy that is channeled between the sea surface and a turning point somewhere above the attenuating seafloor. A very shallow (or very deep) source in the ocean will primarily excite high order acoustic modes that are attenuated due to interaction with the seafloor.

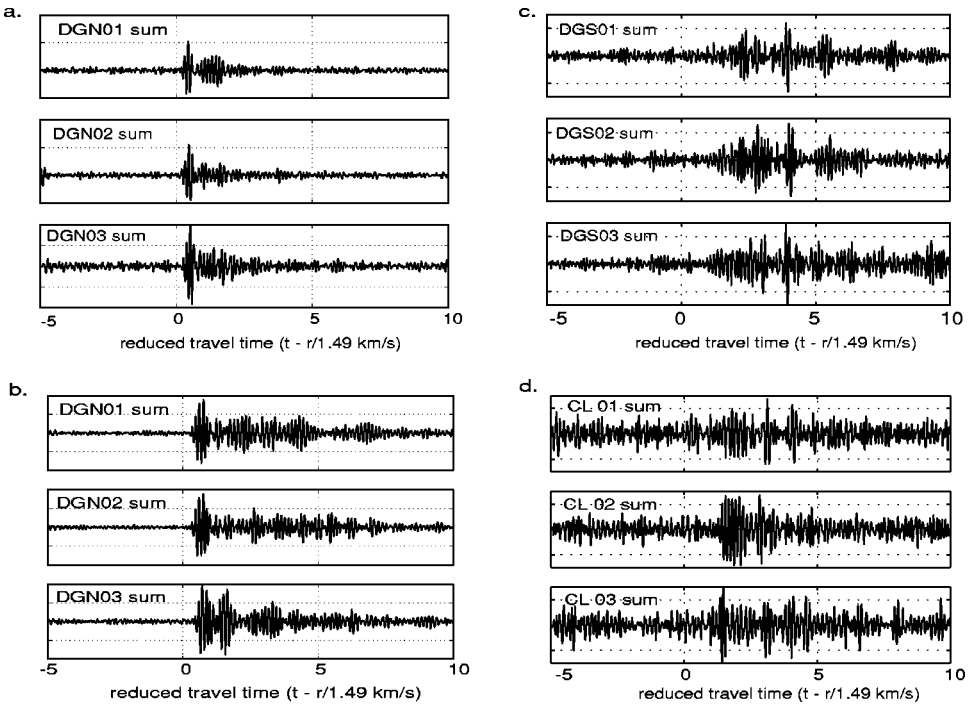


FIG. 8. Examples of airgun array shots recorded at the IMS hydrophones. All shots at each site are stacked after moveout correction. Sections are bandpass filtered 5–10 Hz and time zero corresponds to propagation velocity of 1.49 km/s for each source–receiver range, r. (a) *air2* shots at each of the DGN hydrophones, (b) *air5* shots at the 3 DGN hydrophones, (c) *air9* shots at each DGS hydrophone, (d) *air9* shots at each CL hydrophone.

Some physical process that converts this energy to low order modes that propagate almost horizontally is required. Two mechanisms involving interaction with the seafloor serve to couple high order modes to low order modes, and these are also invoked to describe excitation of acoustic phases, T-waves (Tolstoy and Ewing, 1950), by sub-seafloor sources. These are downslope propagation, which involves repeated interactions of an acoustic phase with a sloping seafloor (Tolstoy and Ewing, 1950), and scattering from a rough seafloor (de Groot-Hedlin and Orcutt, 1999, 2001), which predicts that low order acoustic phases are excited in proportion to the acoustic mode amplitude at the seafloor depth. Thus downslope propagation involves slowly varying bathymetry, while acoustic scattering involves much smaller scale variation in seafloor topography. Park *et al.* (2001) show that T-waves may be excited by a combination of these effects, the key being the occurrence of inhomogeneous structure that enables mode coupling.

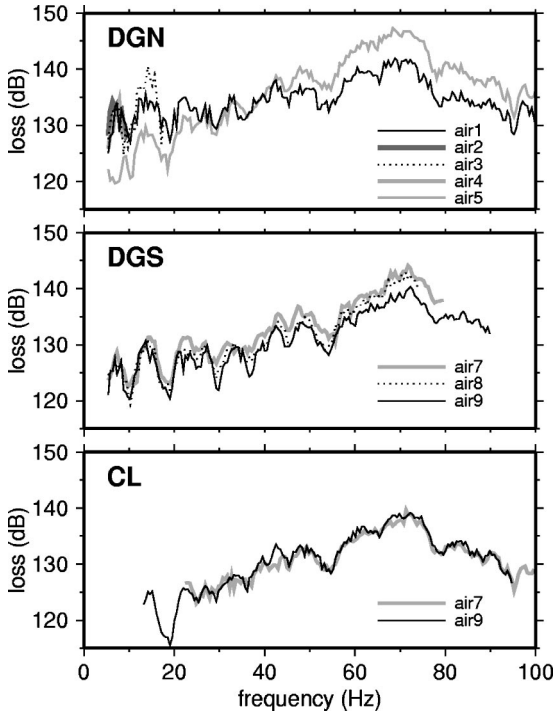


FIG. 9. Transmission loss for each source–receiver pair. The median value of loss for all shots at each site is plotted. Each curve shows only frequencies where signal exceeded current background noise levels by 3 dB. At DGN, *air4* loss is very similar to *air1–air3* losses in the 5–8 Hz band.

Near-source bathymetry was sloped at several of the shooting sites (*air1–air3, air7*). Downslope propagation was a significant factor in coupling high to low order modes. In addition, the T-waves observed in this experiment were probably also excited by scattering of acoustic energy into the sound channel near the source, particularly at *air5*, *air7*, and *air9*. The latter would explain the variability in frequency content of the observed signal between shot sequences, since the amount of scattering depends critically on the roughness scale length in comparison to the acoustic wavelength. This also explains why stacking is ineffective at frequencies above 30 Hz, since even subtle variations in seafloor topography from shot to shot can influence the wave form.

The reason that the *air8* signals were not observed at CL is not certain but topographic variability probably contributes. Kilometer-scale roughness on the eastern flank of the Ninety East Ridge was observed in our swath bathymetry data and these features may scatter energy out of the source–receiver path, thereby reducing the observed signal level for *air8*. The fact that we did observe signals from *air9* probably is due to the existence of a topographic source of coupling just to the south of the shooting site. The slope of Gulden Draak Knoll (Fig. 1) shallows significantly away from our track so this elevated topography could act as a T-wave radiator, both by reflecting direct water waves and through

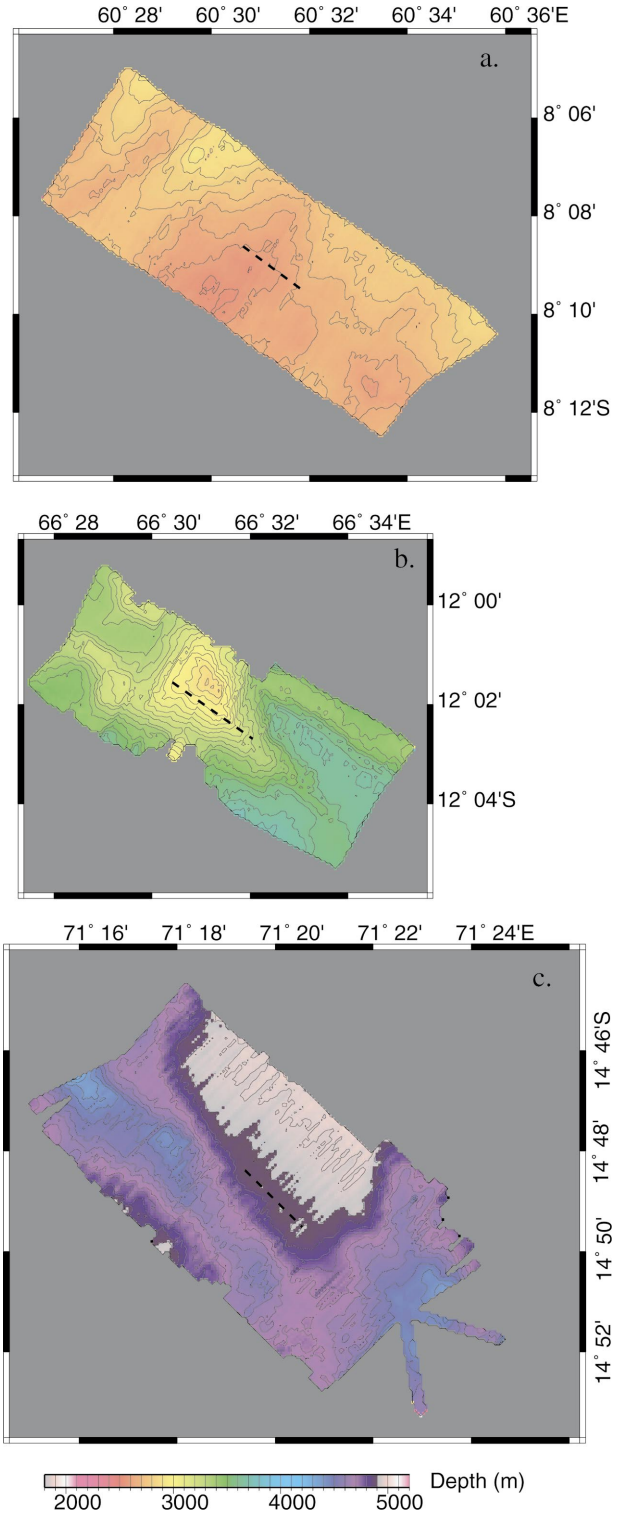


FIG. 10. Hydrosweep bathymetry swaths at representative airgun array shooting sites. Shiptrack runs SE down the center of each swath. Contour interval is 50 m. Coverage shown extends a few kilometers in front of and beyond where the shots were fired (dashed line) to illustrate the broader source region. (a) *air2* site, (b) *air5* site, (c) *air6* site.

seismoacoustic coupling of crustal waves at the slope, within several kilometers of the shots.

The physical process by which low order acoustic phases are excited has implications for modeling their transmission. Since scattering involves much smaller scale lengths than downslope propagation, much greater detail in

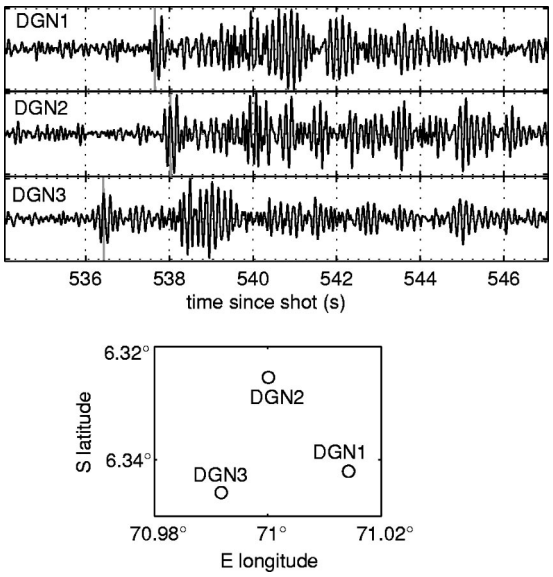


FIG. 11. Arrival time picks for shot 70 (*air5*) used in determining back-azimuth for station DGN. Wave form data for each hydrophone sensor are shown (top) with travel-time pick indicated by gray vertical line. Geometry of the individual sensors at DGN is shown in map view below.

the seafloor bathymetry is required in order to accurately model this phenomenon. The Hydrosweep multibeam swath has a sonar footprint on the seafloor that is on the order of 100 m at depths of ~3000 m, thus we could expect to reasonably model T-waves at frequencies up to 15 Hz for very shallow sources with these data. Outside the vicinity of the source the accuracy of the bathymetry required decreases significantly since the T-wave propagates nearly adiabatically after scattering into low order modes.

First-order estimates of acoustic loss for an 8 Hz signal were obtained using a range-dependent, elastic parabolic equation algorithm (Collins, 1993). Acoustic propagation is two-dimensional along the geodesic path from source to receiver, and constant sub-seafloor elastic structure is assumed (Hamilton, 1980; Laske and Masters, 1997). Oceanic properties (Levitus, 1994) and depths (Smith and Sandwell, 1997) varied along the model path. Quantitative comparison between these estimates and our observations (Fig. 9) is not warranted since acoustic energy from shallow sources interacts strongly with the seabed. A seafloor model more precise than is currently available would be required to justify more complete transmission loss modeling for the airgun source. However, the first-order estimates incorporate the effect of broad slopes and overall basin depth variations so we discuss a few qualitative comparisons. The estimated transmission loss at DGN is significant for sites *air6–9* and at DGS the losses for sites *air1–3* are also quite high. Relative loss levels estimated for DGN from sites *air1* and *air2* follow the observed pattern in the 5–10 Hz range. Compared to the *air1* site, a steeper NNE slope characterizes the *air2* site but depths are greater, so less of the energy is reflected toward the sound channel axis. The estimated loss for *air3–5* is greater for DGN than what is observed, most likely due to rough seafloor along the CIR. On the western side of the Ninety East Ridge, *air7* seafloor was characterized by moderate slope at 3400–3700 m. These

depths are well below the axis of the sound channel and the slope geometry was not optimum so that can explain the lack of signal at DGS. However CL, which first-order estimates indicate should have been even less favorable, did record *air7* shots in the 30–60 Hz band. This can be explained by seismic transmission through the crust for several kilometers to the summit of the Ninety East Ridge and subsequent coupling to acoustic modes at the shallower depths near the summit. Current detailed maps (Fisher, 2003) did not indicate the shallowness of the summit of the ridge where our track crossed it; we measured depths as little as 2500 m to the North of our centerbeam.

VI. CONCLUSIONS

At this stage of our investigation, we can conclude that a large airgun array can indeed be used to generate a signal that contains sufficient energy in the 5–60 Hz band to propagate ocean-basin scales (thousands of kilometers). Whether a given hydrophone station will clearly record the signal is a function of background noise levels as well as topographic conditions near the source and along the path to the receiver. At the time of our experiment, noise levels at DGN were favorable, but at DGS and CL the 5–15 Hz band was characterized by background noise that was too high for clear detection of the airgun array shots. However, several of the shots had good SNR in the 30–60 Hz band. A source with greater output than what the airgun array can consistently couple into the sound channel over variable topography would provide a more consistent means of calibrating the Indian Ocean IMS for the lowest (2–15 Hz) frequency band of interest.

The clear recordings of our glass sphere implosions demonstrate that an ocean-basin network characterization using the five-sphere system as a source is possible provided frequency content below 40 Hz is not required. The duration of the signal, even after over 4000 km of propagation, is short and therefore ideal for travel time measurements. At ranges of several hundred kilometers, the single-sphere sources can reliably produce a signal as well, again dominated by frequencies in the upper half of the 2–100 Hz band of interest for nuclear test monitoring.

In areas of shallow, rough topography, it would be necessary to obtain high-resolution bathymetry data to accurately model the higher frequency aspects of the wave form that may contribute to source characterization. The expense of acquiring such data would tradeoff with the level of apparent potential for the region to become a nuclear test site.

ACKNOWLEDGMENTS

This work was supported by DOE Contract Nos. DE-FC03-01SF22356 and W-7405-ENG-48. The captain and crew of the R/V EWING provided excellent shipboard support during the experiment. Bob Fisher kindly provided his most recent version of the Indian Ocean bathymetric compilation. N. Driscoll, F. Sutherland, C. Hollinshead, and J. Stennett made the deployment of the five-sphere glass imploders a reality. C. Keen provided critical logistical support as cruise schedules and shipping requirements changed at the last minute, due to international events of Fall 2001.

[1]At the time of this writing, details on the airgun array and signal were posted at www.ldeo.columbia.edu/Ewing/airguns/ewing_array.html.

Collins, M. D. (**1993**). ''A higher-order energy-conserving parabolic equation for range-dependent ocean depth, sound speed, and density,'' J. Acoust. Soc. Am. **101**, 1068–1075.

de Groot-Hedlin, C. D., and Orcutt, J. A. (**2001**). ''Excitation of T-phases by seafloor scattering,'' J. Acoust. Soc. Am. **109**, 1944–1954.

de Groot-Hedlin, C. D., and Orcutt, J. A. (**1999**). ''Synthesis of earthquake-generated T-waves,'' Geophys. Res. Lett. **26**, 1227–1230.

Fisher, R. L. (**2003**). *GEBCO Digital Atlas—Centenary Ed.*, British Oceanographic Data Centre, Intergovt. Ocean. Comm. and Intl. Hydrogr. Org., Sheet G.08.

Hamilton, E. L. (**1980**). ''Geoacoustic modeling of the sea floor,'' J. Acoust. Soc. Am. **68**, 1313–1340.

Hanson, J., Le Bras, R., Dysart, P., Brumbaugh, D., Gault, A., and Guern, J. (**2001**). ''Operational processing of hydrocoustics at the Prototype International Data Center,'' Pure Appl. Geophys. **158**, 425–456.

Hanson, J. A., Bowman, R., and Beall, G. (**2002**). ''An advanced concept demonstration for monitoring the Indian Ocean,'' Proceedings of 24th Seismic Research Review, pp. 632–642.

Harben, P. E., Boro, C., Dorman, L., and Pulli, J. (**2000**). ''Use of imploding spheres: An alternative to explosives as acoustic sources at mid-latitude SOFAR channel depths,'' University of California Publ. No. UCRL-ID-139032.

Harben, P. E., and Boro, C. (**2001**). ''Implosion source development and Diego Garcia reflections,'' Proceedings of the 23rd Seismic Research Review, pp. 23–31.

Isaacs, J. D., and Maxwell, A. E. (**1952**). ''The ball-breaker; a deep water signaling device,'' J. Mar. Res. **11**, 63–68.

Laske, G., and Masters, G. (**1997**). ''A global digital map of sediment thickness,'' EOS Trans. Am. Geophys. Union **78**, F483.

Lawrence, M. (**2003**). (private communication).

Levitus, S. (**1994**). *World Ocean Atlas 1994*, NOAA Atlas NESDIS 4, Washington D.C.

Orr, M., and Schoenberg, M. (**1976**). ''Acoustic signatures from deep water implosions of spherical cavities,'' J. Acoust. Soc. Am. **59**, 1155–1159.

Park, M., Odom, R. I., and Soukup, D. J. (**2001**). ''Modal scattering: A key to understanding oceanic T-waves,'' Geophys. Res. Lett. **28**, 3401–3404.

Pulli, J., Farrel., T., and Gibson, R. (**1999**). ''Characterization and utilization of hydroacoustic signals reflected from the continent and bathymetric features,'' Procceedings of the 21st Seismic Research Review, pp. 49–56.

Sauter, A. W., and Dorman, L. M. (**2003**). ''Designing a Deep-Towed/Seafloor Repetitive Sound Source,'' EOS Trans. Am. Geophys. Union **84**, Abstract S52D–0163.

Smith, W. H. F., and Sandwell, D. T. (**1997**). ''Global seafloor topography from satellite altimetry and ship depth soundings,'' Science **277**, 1956–1962.

Tolstoy, I., and Ewing, M. W. (**1950**). ''The T phase of shallow focus earthquakes,'' Bull. Seismol. Soc. Am. **40**, 25–51.

Surface wave focusing and acoustic communications in the surf zone

James C. Preisig
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Grant B. Deane
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093

(Received 22 September 2003; revised 17 May 2004; accepted 18 May 2004)

The forward scattering of acoustic signals off of shoaling surface gravity waves in the surf zone results in a time-varying channel impulse response that is characterized by intense, rapidly fluctuating arrivals. In some cases, the acoustic focusing by the curvature of the wave crest results in the formation of caustics at or near a receiver location. This focusing and the resulting caustics present challenges to the reliable operation of phase coherent underwater acoustic communications systems that must implicitly or explicitly track the fluctuations in the impulse response. The propagation physics leading to focusing are studied with both experimental data and a propagation model using surface wave profiles measured during the collection of the experimental data. The deterministic experimental and modeled data show good agreement and demonstrate the stages of the focusing event and the impact of the high intensity arrivals and rapid fluctuations on the ability of an algorithm to accurately estimate the impulse response. The statistical characterization of experimental data shows that the focusing by surface gravity waves results in focused surface reflected arrivals whose intensity often exceeds that of the direct arrival and the focusing and caustic formation adversely impacts the performance of an impulse response estimation algorithm. [DOI: 10.1121/1.1771591]

PACS numbers: 43.30.Re, 43.60.Cg, 43.60.Dh [EJS] Pages: 2067–2080

I. INTRODUCTION

The surf zone presents a challenging physical environment for underwater acoustic communications systems. This range dependent and highly reverberant environment is complicated by acoustic scattering from the shoaling gravity wave field and attenuation of acoustic signals by dense plumes of bubbles. The surface wave field can cause significant fluctuations in the delay of surface scattered signals, the breaking surf injects dense plumes of bubbles in the water column resulting in highly variable scattering and attenuation of propagating signals, and wave front focusing of surface reflected signals results in the formation of transient caustics. Prior work has examined characteristics of sonar performance in the surf zone environment[1] and the influence of bubble clouds on acoustic propagation.[2] While all of the above listed acoustic effects impact the performance of acoustic communications systems, this paper will analyze the characteristics of surface wave focusing and transient caustics and their potential impact on phase coherent underwater acoustic communications systems.

Wave front focusing caused by acoustic reflections from shoaling surface waves is a significant propagation phenomena in the near-shore environment (see, for example, Ref. 3 for a treatment of caustics and deterministic surface scattering). As will be shown in this paper, the caustics resulting from this focusing often have amplitudes much greater than those of the direct arrival, occur at delays much greater than those of the direct arrival, are highly transient in nature often appearing and disappearing within a few hundred milliseconds, and may exhibit abrupt phase shifts.

In the surf zone environment where the time-varying channel impulse response has significant multipath-induced delay spread, phase coherent signal demodulation relies implicitly or explicitly on accurate estimation of this impulse response. It will be shown later in this paper that the characteristics of the surface wave focusing and the resulting transient caustics described in the previous paragraph pose challenges to the reliable estimation of the channel impulse response and therefore the reliable operation of such signal demodulation algorithms.

To illustrate the importance of acoustic wave front focusing by surface gravity waves, consider Fig. 1. This figure shows the mean and maximum magnitude squared (intensity) of the channel impulse response measured 40 m from a source in the surf zone as a function of relative delay. These sample statistics were calculated using observations over a 9 s time period as a surface wave passed overhead. The details of the processing to generate these statistics are contained in Sec. IV and the Appendix. The direct and bottom arrivals are relatively stable and show little difference between their mean and maximum values. However, the surface reflected arrivals are highly dynamic. While their mean values are significantly below those of the direct and bottom arrivals, their peak values exceed those of the direct and bottom arrivals by as much as 7 dB. At first glance, these increased magnitudes are surprising since surface scattering and geometrical spreading would lead one to expect the kind of magnitude decay observed in the mean values. However, the

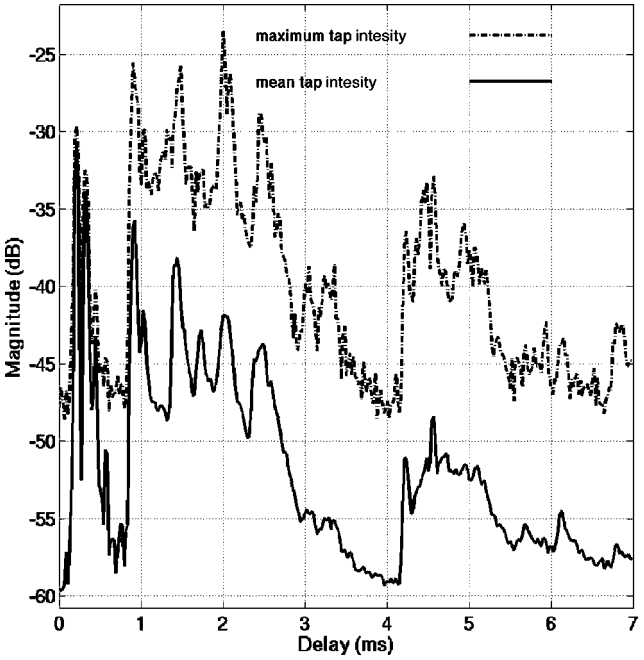


FIG. 1. Maximum and mean intensity of estimated time-varying channel impulse response. The first and second signficant arrivals, both occurring before 0.5 ms in delay, are the direct and bottom bounce arrivals, respectively. The next four significant arrivals occurring in the interval from approximately 0.9 to 2.5 ms in delay represent the surface bounce, surface-bottom bounce, bottom-surface bounce, and bottom-surface-bottom bounce paths, respectively. The arrivals occurring after 2.5 ms in delay all represent paths with more than one surface bounce.

analysis in this paper will show that acoustic focusing by surface gravity waves can lead to surface scattered arrivals that are significantly greater in amplitude than direct and bottom scattered paths. As noted earlier, these high amplitude arrivals are highly transient in nature and pose challenges to acoustic communications systems in the surf zone.

This paper combines results from theoretical analysis, analysis of data generated using an acoustic propagation model, and the analysis of experimental data. A description of the propagation experiment labeled Wavefronts II from which data was obtained can be found in Sec. II. Section III begins with a description of the channel impulse response both as measured during the Wavefronts II experiment and modeled using the Wavefronts acoustic propagation model.[4] It then goes on to describe the anatomy of an acoustic focusing event and concludes with an analysis of the impact of the acoustic focusing on the estimation of the channel impulse response and the performance of phase coherent acoustic communications systems. Section IV presents statistical characterizations of the channel impulse response and performance of the least-squares algorithm used to estimate it. Section V summarizes the contributions of the paper. The Appendix details the processing methods used to estimate the channel impulse response and scattering functions observed during the experiment.

The impulse response intensities, scattering function intensities, received signal levels, and signal estimation residual error (SER) levels presented in the paper are all represented in dB. Unless otherwise noted, the reference for these is the estimated intensity of the direct arrival of the channel impulse response one meter from the source transducer during the Wavefronts II experiment.

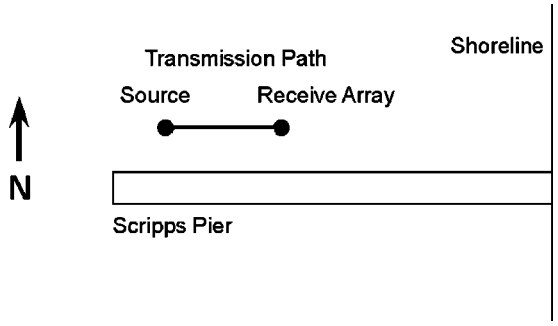


FIG. 2. Wavefronts II experiment top view.

II. THE WAVEFRONTS II EXPERIMENT

The Wavefronts II transmission experiment took place in the surf zone in approximately 6 m deep water, 30 m north of Scripps Pier in December 2001. Top and side views of the experiment geometry are shown in Figs. 2 and 3. Broadband signals with center frequencies ranging from 12 to 26 kHz were generated with an ITC 1007 source jetted 2.0 m above the seafloor. The signals were transmitted 38 m inshore to an array of 3, ITC 6050 C hydrophones vertically spaced at 0.5 m intervals, with the bottom hydrophone 1.51 m above the seafloor. A reference hydrophone was deployed at the same depth as and 0.71 m shoreward of the source and was used to monitor the source signal and level. The experiment geometry was designed to allow surface reflected arrivals to be time-separated from other paths with the available source bandwidth and this was achieved.

An array of 8 pressure sensors was deployed just above the seafloor along the acoustic propagation path to allow shoaling surface gravity waves to be monitored simultaneously with the acoustic transmissions. The sensors of the pressure array were sampled at a rate of 5 Hz. The water depth and pressure sensor locations were surveyed by divers during the experiment, and are identified in Fig. 3. The solid line shows a least-squared fit to the surveyed depths. The linear regression analysis shows that the seafloor had an almost constant slope of 2.0° along the propagation path. The sea surface corresponds to an actual surface gravity wave profile measured during the time of the transmissions, and illustrates the fact that generally only one wave crest was found between the source and the receive array. This is an important point that we will return to later.

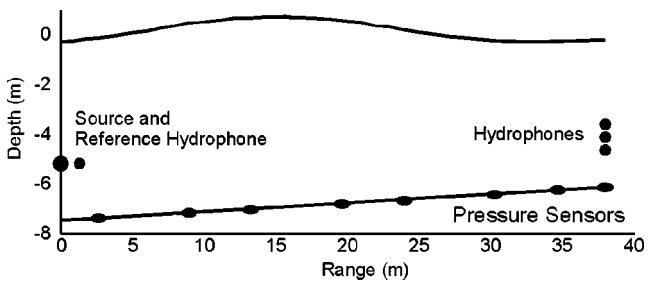


FIG. 3. Wavefronts II experiment side view.

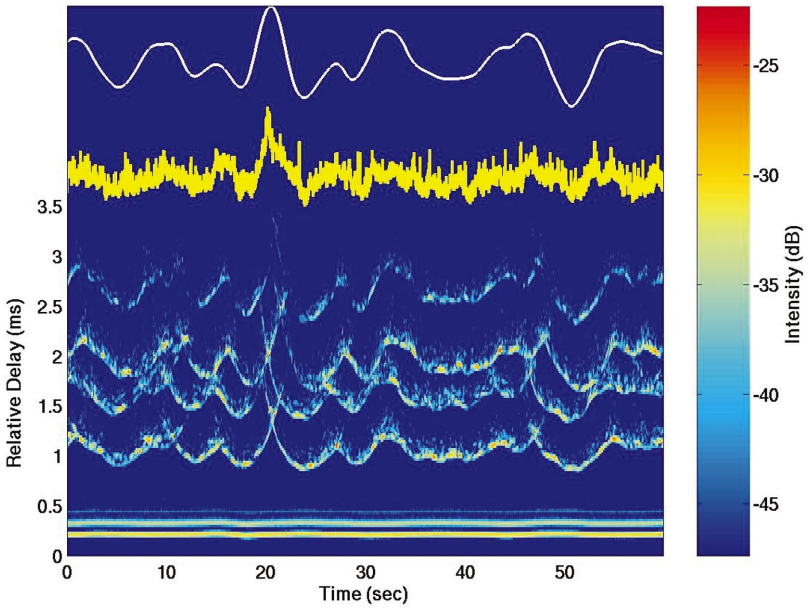


FIG. 4. Surface wave height, SER, and intensity of estimated time-varying channel impulse response. The horizontal lines at the bottom represent the overlapping direct arrival and first bottom bounce. The time-varying arrivals, in order from bottom to top, are the first surface bounce, the surface-bottom bounce, the bottom-surface bounce, and the bottom-surface-bottom bounce. The top white line shows the measured surface wave height near the specular reflection point of the first surface scattered path. The trough to peak excursion on this plot is 1.21 m. The yellow line below the surface wave height is a plot of the magnitude of signal estimation residual error (SER) realized by the algorithm used to estimate the channel impulse response. This plot is in dB and the minimum to maximum error excursion is 10.74 dB.

The sound speed during the experiment was measured to be 1503 m/s. The sound speed, absorption, and density of the seafloor can be estimated using Buckingham's geoacoustic theory for sediments provided the mean grain size of the sand is known.[5] The mean diameter of grains in a sample of sand collected from La Jolla Shores beach was analyzed photographically and estimated to be 573 microns. The geoacoustic parameters corresponding to this diameter and taken to be representative of the seafloor are a density of 2048 kg/m^3, a sound speed of 1757 m/s (corresponding to a critical angle in the seafloor of 31.2°) and an absorption of 0.85 dB per wavelength.

The transmit and receive signals were sampled at a rate of 96 kHz. Transmit signals were generated with center frequencies of 12, 18, and 26 kHz, and were prefiltered to provide an approximately flat system frequency response over a bandwidth of the inverse of the pulse or symbol duration around the center frequency of the signal. Signals were transmitted continuously for approximately 60 s with a several second pause between each transmission period. Thirty such 60 s transmissions were made with each signal. Three signal formats were transmitted. The first was a binary phase shift keyed (BPSK) signal modulated by continuous repetitions of a 4095 point maximum length shift register sequence (m-seq).[6] The symbol rates for the 12, 18, and 26 kHz signals were 16 000, 24 000, and 48 000 symbols per second, respectively. The second transmitted signal consisted of a single cycle of a pulse at the 12, 18, or 26 kHz center frequency with a repetition rate of 16 pulses per second. The final transmitted signal consisted of interleaved single cycle pulses and a BPSK signal modulated with 13 symbol Barker code. The data analyzed and presented in this paper is the 18 kHz m-seq data.

III. A SURFACE WAVE FOCUSING EVENT AND ITS IMPACT ON ACOUSTIC COMMUNICATIONS

The intensity of the channel impulse response estimated from data collected during the Wavefronts II experiment, the surface wave height at the time of data collection, and the signal estimation residual error (SER) realized by the algorithm (see Appendix) used to estimate the channel impulse response are shown in Fig. 4. While the delay spread of the impulse response is on the order of 7 ms, this figure shows the arrivals for only the first 3.5 ms of the arrival structure. This includes the direct and first bottom bounce arrivals as well as all of the arrivals with a single surface bounce. A low SER indicates that the algorithm is providing a good estimate of the channel impulse response.

The variations in arrival time and intensity of the surface scattered paths are clearly associated with the passage of surface gravity waves over the experiment transmission path. Transient caustics are formed by the wave front focusing property of shoaling surf. As open ocean swell approaches the coastline and begins to interact with the sea floor, it refracts so that wave fronts become parallel to the isobaths and it grows in amplitude. As it shoals, each wave crest forms an acoustic mirror with its own characteristic shape and focusing properties. The surf therefore forms a series of time-varying mirrors that create moving focal regions of scattered sound.

Several important features can be observed in the data. The first is that the passage of a peak and trough results in an asymmetric pattern of arrival time fluctuations and, most significantly, signal estimation errors. The passage of a trough results in a single arrival that first decreases and then increases in delay so the range rate of the arrival is approxi-

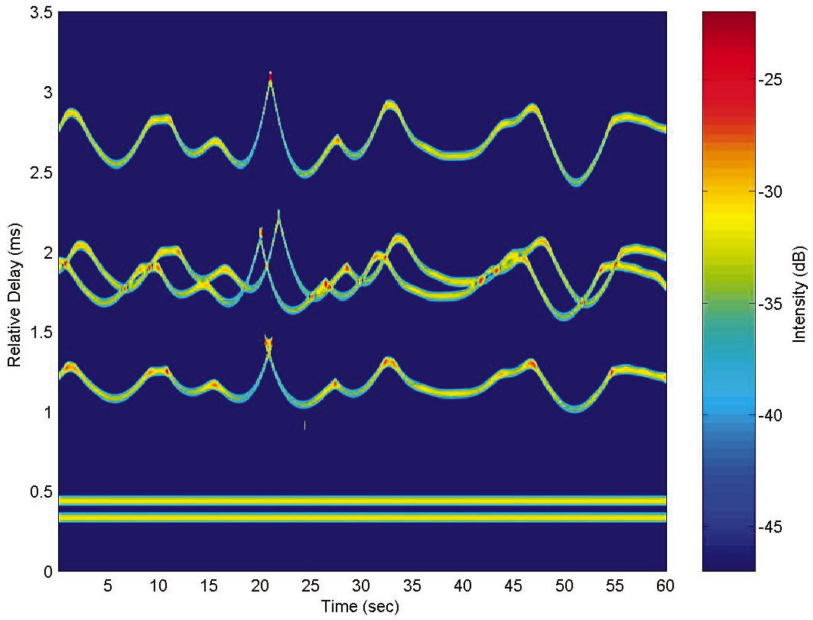


FIG. 5. Intensity of modeled time-varying channel impulse response.

mately zero. The term ''range rate'' is used herein to mean the apparent rate at which the location of an arrival is increasing or decreasing in delay. In contrast, the passage of a peak often results in two arrivals, one increasing and the other decreasing in delay, crossing with high range rates. The high range rates result in a large signal estimation error. This is most noticeable with the large wave crest that passed over the specular point at a time of 20 s in the plot. This pattern will be discussed in detail in Secs. III B and III C.

The second feature is that the intensity of the received signal at the caustics is larger than the direct arrivals, despite the fact that they have travelled along a longer path and been scattered by small-scale surface roughness. The caustics can appear at a delay that is significant (from a communications context) with respect to the direct arrival. In addition, the caustics can appear and disappear in the span of a few hundred milliseconds.

In Sec. III A which follows, the agreement of the estimated impulse response and the impulse response predicted with the Wavefronts acoustic propagation model is illustrated and limitations are discussed. In Sec. III B, the model results are used to illustrate the formation of a transient caustic and its relevant features during the passage of a single wave peak observed during the Wavefronts II experiment. This section is concluded with Sec. III C which uses acoustic data from the same wave peak event to further illustrate the relevant features of the transient caustic and their impact on the performance of a least squares channel estimation algorithm.

A. The Wavefronts acoustic propagation model and comparison with experimental data

A new method of modeling underwater sound propagation, called Wavefront Modeling, has been developed by Tindle.[4] The method is based on a Hankel transform-generalized Wentzel–Kramers–Brillouin (WKB) solution of the wave equation. The resulting integral leads to a new form of ray theory which is valid at relatively low frequencies and allows evaluation of the acoustic field on both the illuminated and shadow sides of caustics and at cusps where two caustics meet to form a focus. The method is much faster than other standard methods, is able to handle rapid range dependence, and the phase, amplitude, and travel time of broadband acoustic pulses are obtainable directly from a simple graph of ray travel time as a function of depth at a given range. The model can handle the following features of the environment or acoustic field: real but smooth surface waves from experimental data, range dependent smooth bathymetric variations, and cusps and caustics. The model can propagate real waveforms enabling direct comparison of modeled received waveforms and waveforms received during field experiments.

The Wavefronts model calculated the first 3.5 ms of the impulse response arrival structure shown in Fig. 5. The surface wave field used as an input to the model for this 60 s period was the same as that measured during the time when the acoustic signals used in generating Fig. 4 were gathered. The model shows good qualitative agreement with the impulse response estimates generated from the experimental data. The model results show the high intensity of the surface scattered arrivals as each wave peak passes and a lack of high intensity arrivals during the passage of wave troughs.

Of equal importance are the differences between the experimental data and model results. There is fine scale structure apparent in the data that is not seen in the model results. It is believed that this structure is due to the presence of small scale surface roughness that could not be detected by the bottom-mounted pressure sensors used during the experiment. This surface roughness is therefore not reflected in the model results. There are also limitations inherent in the least squares algorithm used to estimate the channel impulse response from the experimental data (see Appendix). The primary limitation is the 25 ms averaging time used by the

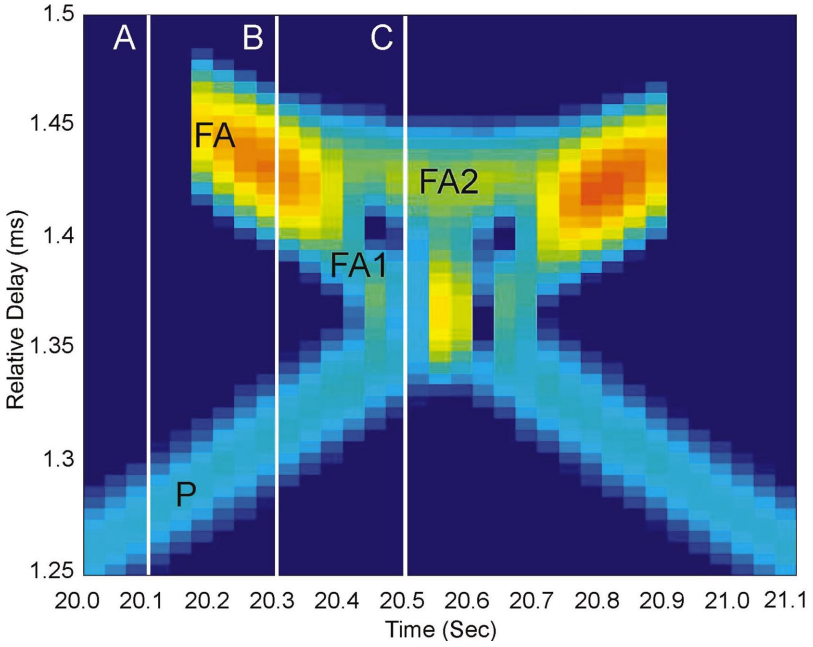


FIG. 6. Intensity of modeled time-varying channel impulse response during focusing event. The times labeled A, B, and C correspond to the three times for which the angle/depth diagrams are shown in Fig. 9. The arrival labels (P, FA, FA1, and FA2) correspond to the similarly labeled arrivals in Figs. 7, 9, and 10. Time A corresponds to the shadow zone of the focused arrival FA. Time B is just after the formation of the caustic. At time C, the saddle points for the focused arrival have diverged and the arrival has split into two arrivals labeled FA1 and FA2.

algorithm which can both smear estimates of arrivals in time and highly attenuate arrivals appearing at Doppler shifts of greater than about 20 Hz.

The same model and experiment results focusing on the first surface bounce arrival in the vicinity of the time of 20 s are shown in Figs. 6 and 7, respectively. Noting the slight offset of the plots in the delay variable, the model results are seen to capture the essential features of the caustic seen in the experimental data. Notwithstanding the fine scale structure present in the data and absent from the model results, the model does provides a sufficiently detailed reproduction of the data to warrant its use as an analysis tool.

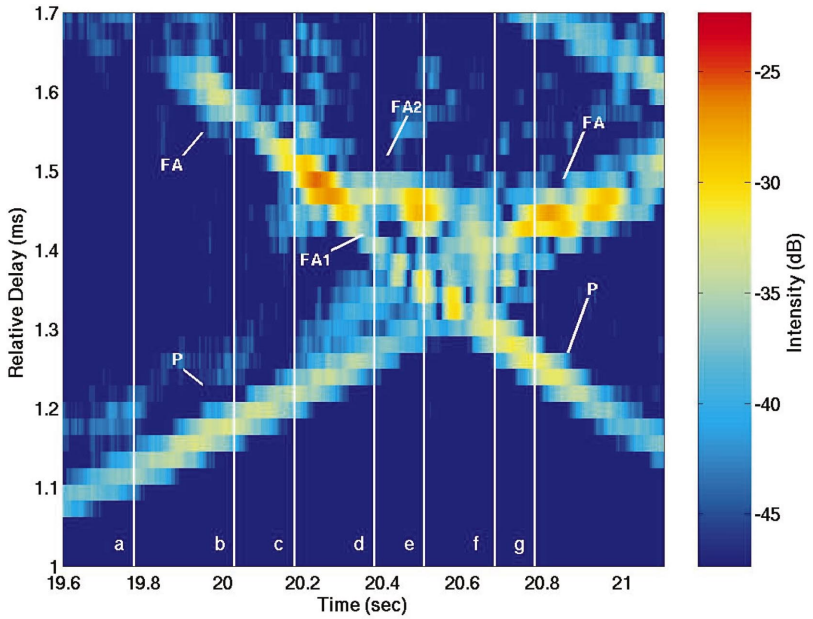


FIG. 7. Intensity of estimated time-varying channel impulse response during focusing event. The vertical white lines, labeled a–g mark significant points of time in the evolution of the surface scattered arrival. The arrival labels (P, FA, FA1, and FA2) correspond to the similarly labeled arrivals in Figs. 6, 9, and 10. The single ray arrival is the retreating primary arrival is labeled P. At time b, the advancing folded wave front arrival, labeled FA has appeared. The arrival has not yet formed a caustic. At time c the caustic is forming as the folded arrival gains in intensity. At time d the caustic has passed, the saddle points are diverging, and the folded wave front arrival has split into two arrivals labeled FA1 and FA2. The later of the two arrivals (FA2) is scattering off of the peak of the wave crest and is therefore stationary in delay and exhibits no Doppler shift. The evolution of the wave front focusing continues through times e–g. FA1 becomes the primary arrival and is advancing. FA2 and P eventually merge to form a caustic at time g near the end of the focusing event.

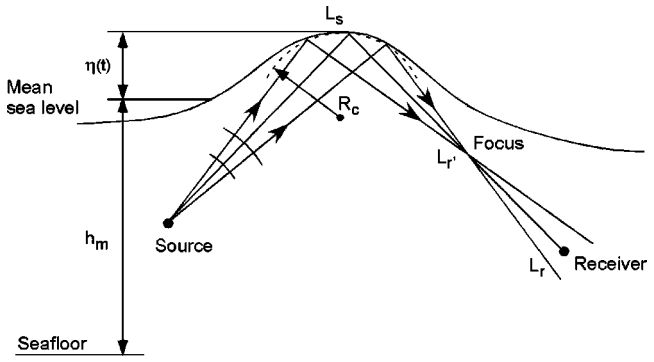


FIG. 8. Qualitative description of acoustic focusing by surface waves. The source to surface wave distance (L_s) and the surface wave to focal point distance ($L_{r'}$) distance are related by the wave focal length (L_f).

B. The anatomy of a caustic

This section begins with a qualitative description of sound focusing by shoaling gravity waves and then illustrates the underlying physical structure with model calculations for the wave that caused the intensification of the first surface bounce arrival about 20 s into Fig. 4. A schematic depiction of wave field focusing is shown in Fig. 8. The shoaling surface gravity wave has an instantaneous height $\eta(t,r)$ as a function of time, t and range r, and the local radius of curvature of the wave crest at the specular reflection point is R_c. The mean water depth is h_m and the total water depth is $h_w(t,r)=h_m+\eta(t,r)$. Sound emitted by the source, L_s meters from the wave crest is shown focused $L_{r'}$ meters from the wave crest, due to scattering by the (assumed locally spherical) surface. The relationship between L_s and $L_{r'}$ is determined by the wave focal length, L_f.

The focusing properties of the shoaling wave crests depends on the detailed shape of the air–water interface at the time of pulse reflection, which includes asymmetry in onshore direction (nonlinear effects result in the leading edge of the crest being steeper than the trailing edge) and scattering from features smaller than an acoustic wavelength. Based on measurements of shoaling wave crests taken from the pressure array deployed during the Wavefronts II experiment, it is assumed that to first order the shape of wave crest can be approximated as a cylindrical shell. As explained in Ref. 7, the normal-incidence focal length of a spherical section depends on the section radius of curvature and the relationship $L_f=R_c/2$ is assumed for the focal length of the waves.

The Wavefronts model calculation of the impulse response intensity during the passage of the wave crest over the acoustic transmission path is shown in Fig. 6. The intensity versus delay and time clearly shows the ''butterfly'' pattern characteristic of the passage of a folded wave front past the receiver. The two intense spots on the wings are caustics, which occur on the boundary between insonified and acoustic shadow regions. Note that the passage of the wave crest does not result in a single arrival increasing and then decreasing in delay. Instead, the surface wave curvature and wave front folding results in two distinct arrivals, one increasing in delay and the other decreasing in delay, which merge and the separate as the wave crest passes through the specular region. As will be shown in Sec. III C, this structure in the wave field results in a significant Doppler spread of the channel scattering function during the focusing event. In the figures showing detailed arrival structure, different arrivals bear the labels, P, FA, FA1, and FA2. The label P indicates a primary surface scattered arrival not subject to strong surface wave focusing while the label FA (with or without a number) indicates a surface scattered arrival arising from a folded wave front which results in surface wave focusing.

The detailed arrival structure of the focused wave field has been analyzed at the three times, A, B, and C annotated in Fig. 6. The results are shown in Fig. 9. The three plots on the left show the depth versus launch angle of small segments of acoustic wave fronts at the receiver range, propagated through the water column and interacting with both the surface and bottom. The segment of wave front between launch angle 1° to −7° corresponds to the direct-path arrival, and the segment between −7° and −28° corresponds to the surface-interacting wave fronts. The plots on the right show ray paths which pass through the receiver after reflection off the surface. The launch angle of these rays corresponds to the points in the left hand plots at which the depth versus launch angle curve intersects a line at the receiver depth. Those points on the surface-interacting segment of the curves have been annotated with a small circle. Points of intersection on other segments of the curve (e.g., the direct-path segment, etc.) also correspond to arrivals at the receiver, but do not correspond to the portion of the channel impulse response considered here.

At time A in Fig. 9, there is only a single intersection point, corresponding to a single ray path connecting the source and the receiver. Accordingly, there is only a single arrival at the receiver at time A in Fig. 6. The region at time A between 1.4 and 1.5 ms delay is a shadow zone for the folded wave front that appears between times A and B. At time B, there are three points of intersection, two of which occur spaced closely together around −20°. The two near-by saddle points form the caustic that first appears between A and B, while the third point remains distinct. At time C, all three points are distinct. If the geometry is such that the receiver lies right on the focal point, then all three saddle points merge. This pattern of arrivals is the underlying structure resulting in the advancing and retreating arrivals observed in the data in Figs. 4 and 7.

The following subsection will illustrate the impact of both the rapid intensification and the crossing of advancing and retreating arrivals during wave focusing on the estimation of the time-varying channel impulse response.

C. The impact of a caustic

The transient caustic examined in the prior subsection has several characteristics of importance to the performance of acoustic communications algorithms. These include the sudden appearance of the folded wave front, the advancing and retreating wave fronts in the surface scattered arrival after the appearance of the folded wave front, and rapid changes in the amplitude of the arrival. There may also be a rapid phase shift associated with the caustics but the current data does not provide sufficient information to verify this

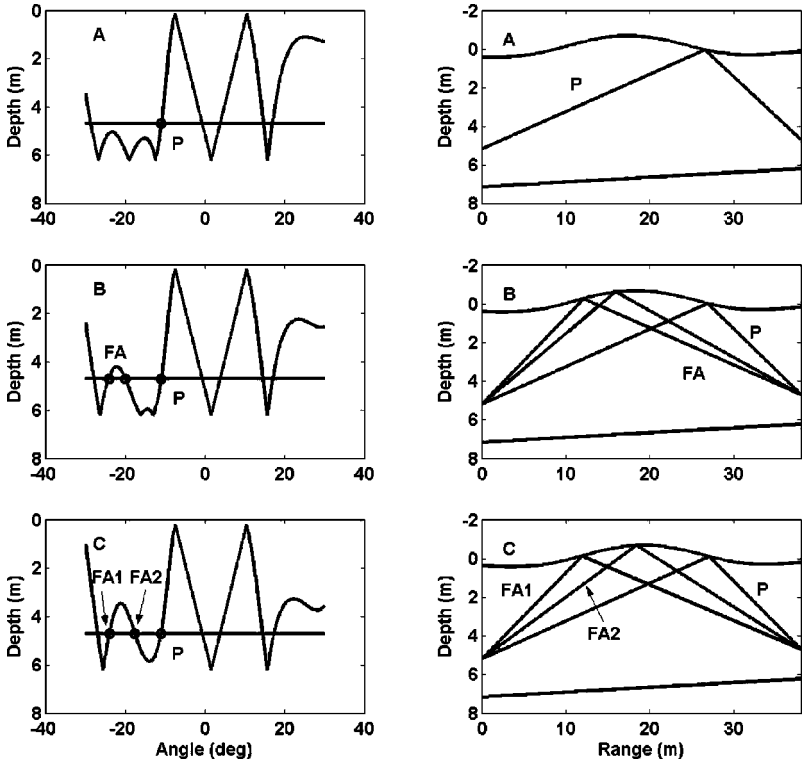


FIG. 9. Modeled angle/depth and ray path diagrams during focusing event. The subfigures labeled A, B, and C correspond to the three times marked in Fig. 6. The arrival labels (P, FA, FA1, and FA2) correspond to the similarly labeled arrivals in Figs. 6, 7, and 10. The divergence of the saddle points between times B and C on the angle/depth diagram results in the formation of two distinct ray paths and arrivals as shown.

hypothesis. The model results shown in Fig. 6 were generated using an actual surface wave profile from the Wavefronts II experiment. The experimental data corresponding to this wave profile are shown in Fig. 7.

There is a close agreement between the model and experimental results. Comparing the two figures, it is clear that the wave front has been folded back on itself resulting in the formation of caustics and multiple arrivals in the focal region. As was noted previously, there is also some fine structure in the experimental data that is not present in the model results. This is likely due to the limitations of the measurements of the surface wave field used as input data to the Wavefronts model. In addition, as mentioned earlier the finite window averaging used by the least squares channel identification algorithm may result in there being rapidly fluctuating features in the channel impulse response not represented in Fig. 7.

The vertical white lines, labeled a–g in Fig. 7 mark significant points of time in the evolution of the surface scattered arrival. The estimated channel scattering function for the surface scattered arrival at each of these points in time is shown in the subfigures of Fig. 10. The scattering function shows the distribution of energy in frequency of the fluctuations of each tap of the channel impulse response (i.e., the horizontal axis) as a function of delay (i.e., the vertical axis). The relative delay axis in the scattering function figures matches the relative delay axis in the impulse response estimate. Energy present at higher frequencies represents arrival energy at a particular delay that is fluctuating more rapidly than energy present at lower frequencies. A description of the scattering function representation of time-varying channels is given in Ref. 8. The distribution of energy in Doppler frequency as a function of relative delay at any of the 8 labeled times in Fig. 7 can be observed by looking at the same relative delay of the corresponding subfigure in Fig. 10.

The movement of energy in the Delay/Doppler plane as the wave crest passes can be explained in terms of the arrivals in Fig. 7 and the model results in Fig. 9. Time a in Fig. 7 corresponds to the shadow zone for the folded wave front. The single ray arrival is the retreating primary arrival (labeled P in Figs. 6–10) and therefore has a negative Doppler shift. This corresponds to time A in Figs. 6 and 9. At time b in Fig. 7, the advancing folded wave front arrival, labeled FA has appeared with a corresponding positive Doppler shift. The arrival has not yet formed a caustic. This feature is not represented in the model results. In the model results, the caustic occurs at the first appearance of the folded wave front arrival. The difference is most likely due to the small scale roughness of the actual sea surface. At time c in Fig. 7, the caustic is forming as the folded arrival gains in intensity. Accounting for both the primary and folded wave front arrivals, the Doppler spread of this surface scattered arrival is approximately 30 Hz. This time corresponds to time B in Figs. 6 and 9.

Time d in Figs. 7 and 10 corresponds to time C in Figs. 6 and 9. At this time, the caustic has passed, the saddle points shown in Fig. 9 are diverging, and the folded wave front arrival has split into two arrivals labeled FA1 and FA2. The later of the two arrivals (FA2) is scattering off of the peak of the wave crest and is therefore stationary in delay and exhibits no Doppler shift. The earlier of the two arrivals (FA1) is scattering off of the trailing edge of the wave crest and continues to be advancing in delay and exhibit a positive Doppler shift. The evolution of the wave front focusing continues through times e–g. The arrival FA1 becomes the primary arrival and is advancing. The arrivals FA2 and P eventually merge to form a caustic at time g near the end of the focusing event.

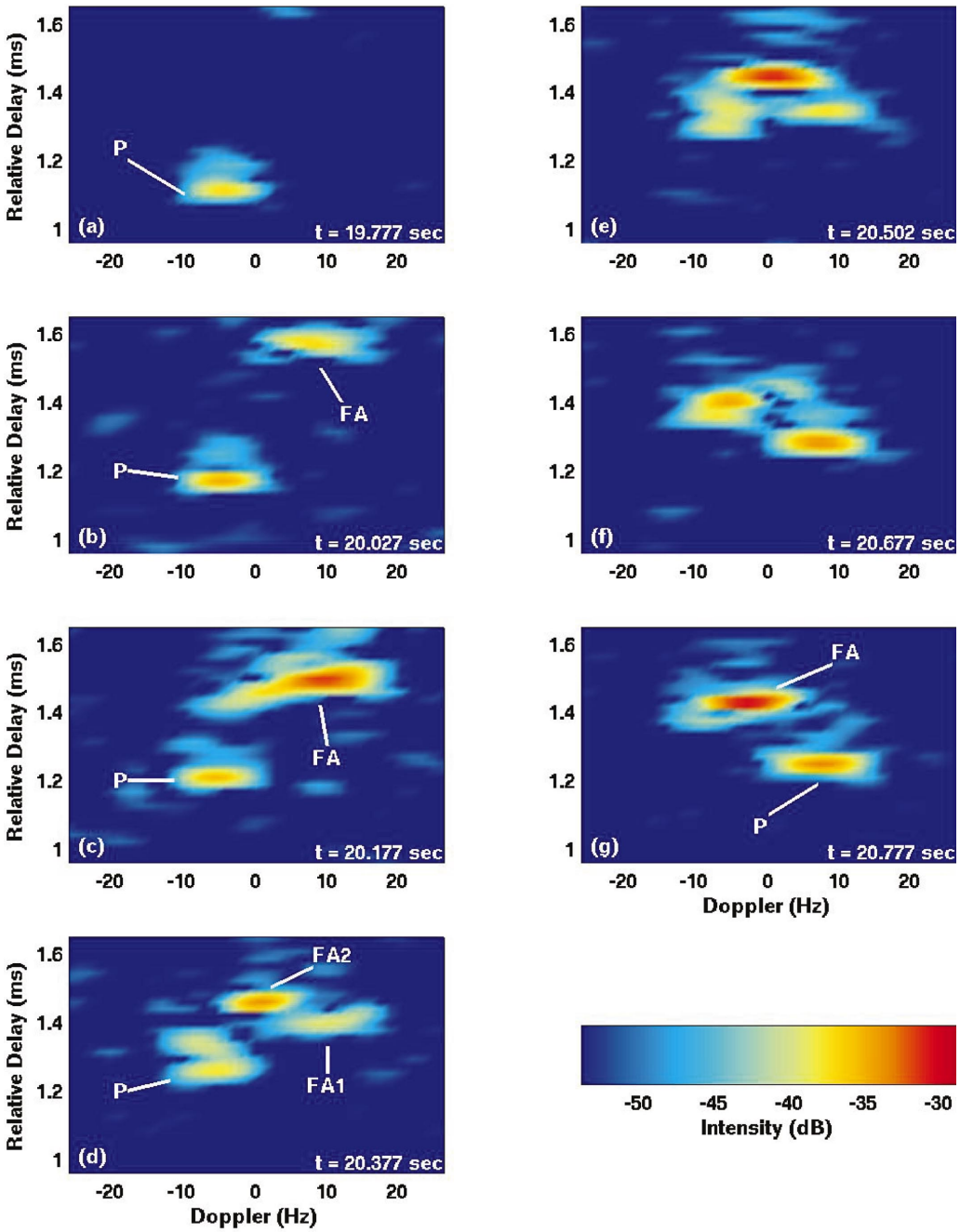


FIG. 10. Estimated channel scattering functions during focusing event. The subfigure labels (a–g) correspond to the similarly labeled points in time in Fig. 7. The arrival labels (P, FA, FA1, and FA2) also correspond to the similarly labeled arrivals in Figs. 6, 7, and 9. The initial primary arrival P is scattering off of the retreating edge of the wave crest and has a negative Doppler shift. The initial folded arrival is scattering off of the advancing edge of the wave crest and has a positive Doppler shift. When the folded arrival splits into two arrivals, the arrival FA2 is scattering off of the peak of the wave crest and has no Doppler shift. FA2 and the initial P arrivals merge to form the final focused arrival scattering off of the retreating edge of the wave crest (FA in subfigure g) with a negative Doppler shift. FA1 becomes the final primary arrival (P in subfigure g) which is scattering off of the advancing edge of the wave crest and has a positive Doppler shift. The scattering function estimates were generated with a matched filtering algorithm described in the Appendix and have a frequency resolution of approximately 6.4 Hz.

The simple model of a flat sea surface moving up and down in elevation as used in Ref. 9, and references therein is not sufficient to explain the rate of path length fluctuation for the first surface scattered arrival observed in Fig. 7 and the resulting Doppler spreads observed in Fig. 10. Using the maximum value for $d\eta/dt$ observed during the passage of the wave, the maximum rate of path length change for this arrival is predicted by the simple model to be 0.23 m/s. The maximum rate of path length fluctuation calculated from the data shown in Fig. 7 is approximately 0.75 m/s (arrival FA at time c in Fig. 7). This difference by over a factor of 3 is due to the effect of the scattering of the focused arrivals off of the advancing and retreating surface wave fronts that propagate at the wave speed (the surface wave speeds observed during the experiment were approximately the 7.5 m/s) and the movement of the scattering point for each arrival along the surface of the wave crest. In addition to the increase in Doppler shift magnitude, focusing from wave crests results in

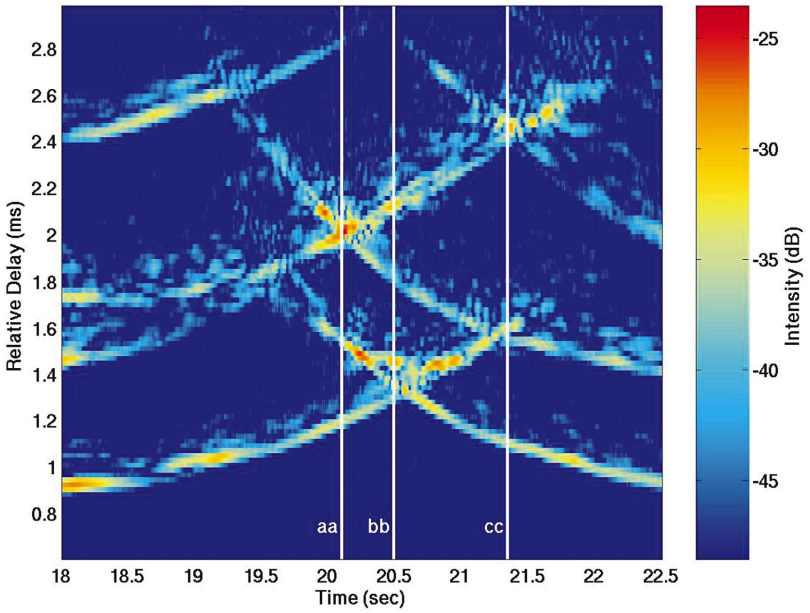


FIG. 11. Three surface scattered arrivals of estimated time-varying channel impulse response during the focusing event. The center of the focusing events for the surface-bottom, surface, and bottom-surface arrivals are labeled aa, bb, and cc, respectively. The sharp peaks in the SER shown in Fig. 12 are labeled in the same manner.

micropath arrivals which exhibit Doppler shifts of opposing signs corresponding to reflections from surface regions on either side of the wave crest maximum. This aspect of the phenomenon would be impossible to reproduce by replacing the wave crest with a moving horizontal surface and presents a problem for commonly implemented phase coherent equalizers.

The impact of the focusing event on the performance of a channel estimation algorithm is illustrated by Figs. 11 and 12. The estimated impulse response in an expanded region in both time and delay around the focusing event is shown in Fig. 11. The focusing event at the time labeled bb is the surface arrival focus shown in Fig. 7. The arrival whose focusing event occurs at aa is the surface-bottom arrival while the arrival whose focusing event occurs at cc is the bottom-surface arrival. The received baseband signal energy and the SER of the channel estimation algorithm are shown in Fig. 12. Note that before the passage of the wave crest, the mean error falls between 8 and 10 dB below the signal energy. As the wave crest passes, the error climbs to within approximately 3 dB of the received signal energy. This increase in error is a direct result of the increase in the time variation of the channel impulse response as described by the Doppler spreads observed in Fig. 10. Note also the sharp spikes on the signal estimation error with the passage of the center of each focusing event (labeled aa, bb, and cc in Figs. 11 and 12). At these times, the received signal estimation errors are within 1–2 dB of the received signal levels. This indicates that the channel is highly dynamic at these times and there may be features of the channel impulse response that are not accurately captured in the channel estimates shown in Figs. 7 and 11. The resolution of these features will be addressed in future work.

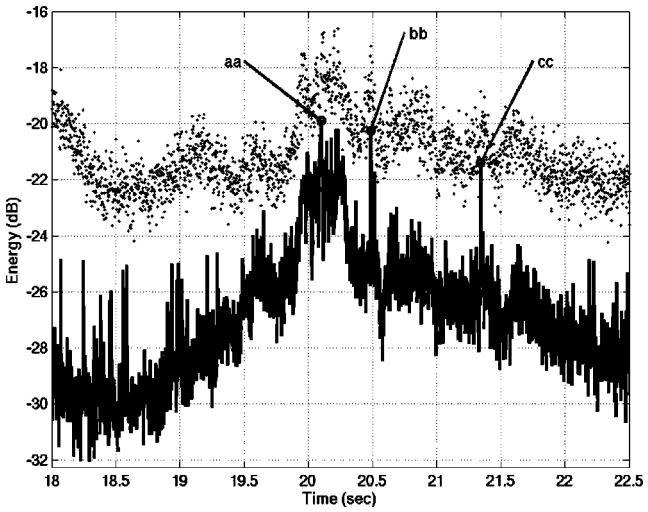


FIG. 12. Signal estimation residual error (SER) and received signal energy during the focusing event. The dots represent received signal energy and the line represents the SER of the least squares channel estimation algorithm. Both quantities are averaged over 1 ms intervals. The sharp peaks in SER labeled aa, bb, and cc correspond in time with the times at the center of the focusing events for the surface-bottom, surface, and bottom-surface arrivals, respectively, as shown in Fig. 11.

A number of the observed features of acoustic focusing events will have detrimental impacts on many current phase coherent demodulation algorithms for underwater acoustic communications. Both channel estimate based algorithms [e.g., channel estimate based decision feedback equalizer (CE-DFE), maximum likelihood sequence estimator] and direct adaptation equalizers (e.g., direct adaptation decision feedback equalizer, DFE) must be able to either explicitly or implicitly track the dominant energy in the time-varying channel impulse response. While in other environments the rapidly varying arrivals may have suffered enough scattering losses to be insignificant, the results here show that the contrary is true in the presence of focusing by surface waves. The focused arrivals have high intensities, often exceeding

the intensity of the direct arrival, and at the same time are rapidly varying. Therefore, communications algorithms in the surf zone must be able to track these arrivals.

A number of innovations have been made to enable algorithms to track channel fluctuations or reduce algorithm complexity. However, these innovations are based upon a number of simplifying assumptions and they work well as long as the assumptions are satisfied. The features of acoustic focusing events observed here violate most of the important assumptions. For example, the Phase Locked Loop/Signal Resampling approach to addressing path length fluctuations[10,11] that has been highly successful in other environments is based upon the assumption that a single time-varying Doppler shift/range rate adequately describes the time variation of the channel impulse response. This assumption is clearly violated by both the surface scattered arrival which is significantly Doppler spread and the entire impulse response that consists of fairly stable direct and bottom arrivals in addition to all of the time-varying surface interacting arrivals. In fact, a DFE based upon this approach has not been successful in demodulating the data collected during the focusing event analyzed here.

Another approach to tracking the time-varying channel impulse response is to estimate the channel scattering function periodically during the reception of communications signals and then to track the impulse response for only those points in the Delay/Doppler plane that have significant energy.[12,13] This technique relies on the assumption that the distribution of energy in the Delay/Doppler plane is stable for periods long enough to allow for initialization and operation of a channel tracking algorithm and the transmission of a packet of data. The evolution of the scattering functions shown in Fig. 10 clearly violates this assumption and would result in a failure of this algorithm.

Finally, channel sparsing algorithms are used to reduce the number of taps of the channel impulse response that are tracked. This sparsing increases the rate of channel fluctuations that can be tracked as well as reduces the computationally complexity of the algorithms.[14,15] Most sparsing techniques assume that the distribution of energy in the delay variable of the impulse response is relatively stable, i.e., the location of the ''significant'' taps in the channel impulse reponse is slowly varying. The data shown in Figs. 4, 7, and 11 shows clearly that this assumption is violated for the environmental conditions under which this data was collected. Again, this would result in the failure of these algorithms. Some techniques[14] do not make this assumption and use other techniques to facilitate tracking of rapid channel fluctuations. However, even these techniques have not been successful in demodulating the data analyzed here.

The results here show that the phase coherent underwater acoustic communications techniques that rely on accurate channel estimation would not be capable of reliable operation in the surf zone environment in the conditions analyzed. This illustrates the need for future work on either improving the ability of channel estimation algorithms to track the fluctuations associated with focusing by surface waves or developing demodulation algorithms that are robust with respect to errors in channel estimates.

IV. STATISTICAL CHARACTERIZATION OF SURFACE WAVE FOCUSING

The m-seq transmissions made during the Wavefronts II experiment occurred in three 10 min intervals spread over 2 h. This section presents some sample statistics gathered over these three 10 min periods. These statistics are calculated from the estimated channel impulse response at 5 ms intervals, the signal estimation residual error (SER) of the least squares channel estimation algorithm (see Appendix), and the surface wavefield as measured by the array of pressure sensors (see Sec. II).

The measures of SER and surface scattered arrival intensity were examined. These two measures are important for communications applications. Arrival intensity is important because high intensity arrivals present both an opportunity and challenge. Arrivals with high intensity result in a higher signal to noise ratio in the received signal which improves the potential for improved estimation of the transmitted data by the receiver. However, in order to realize this improved data estimation performance, phase-coherent demodulation algorithms must be able to accurately track the phase and amplitude of the arrival. SER is a measure of the ability to do this. When the arrival can be accurately tracked (low SER), the data demodulation algorithm will be able to exploit the higher received signal energy to improve the estimation of the received data. Otherwise (high SER), the received energy represents unmodeled signal and will appear to the receiver to be contaminating noise.

Some of the statistics below were calculated for the entire data set and some were calculated for only that portion of the data that was collected within 100 ms of the passage of a surface wave. This is referred to as a ''wave focusing event.'' The determination of the time of each wave focusing event was made by time-aligning the surface wave height as measured for each passing wave near the specular point for the surface scattered path with the time series of estimates of the channel impulse response. The fluctuations in the channel impulse response corresponding to the passage of the wave were identified and the center of the focusing event determined. Statistics regarding the behavior of the channel impulse response and the channel estimator were gathered from data within 100 ms of this center.

Each focusing event was also classified regarding whether or not it resulted in the formation of a caustic. This was done by examining the estimated channel impulse response for the first three surface reflected paths (surface, surface-bottom, bottom-surface) over the duration of the focusing event. If any of these arrivals showed a butterfly like pattern characteristic of a caustic, then the focusing event was classified as having resulted in the formation of a caustic. Otherwise, it was classified as not having resulted in the formation of a caustic.

The sample pdfs of the log intensity $[10*\log_{10}(\text{intensity})]$ of the estimated direct and first surface bounce path arrivals are shown in Fig. 13. In contrast to the results shown in Fig. 1 which show the maximum and mean intensity for each delay tap in the estimated time-varying impulse response, Fig. 13 shows the intensity statistics for estimated arrivals which, for the case of the surface bounce path ar-

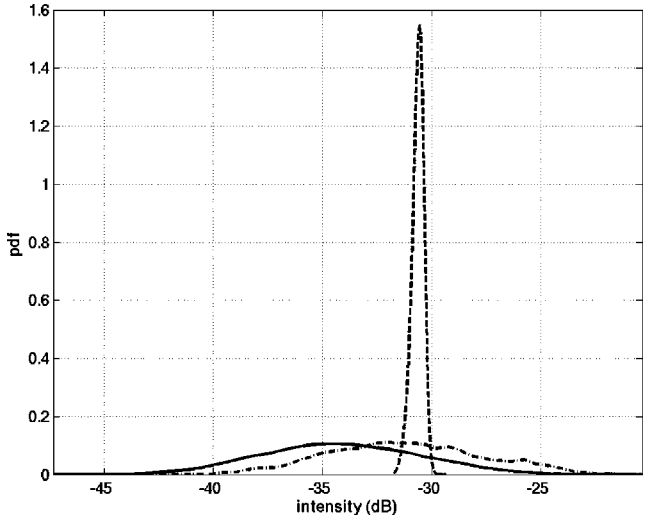


FIG. 13. Sample pdf of $10*\log_{10}$ of the intensities of estimated direct and surface bounce path arrivals. The dashed line is the pdf for the direct arrival, the solid line is the pdf for the first surface bounce path arrival. The dashed–dotted line is also the pdf for the first surface bounce path arrival but using data only from within the interval extending from 100 ms before to 100 ms after the time of each wave focusing event.

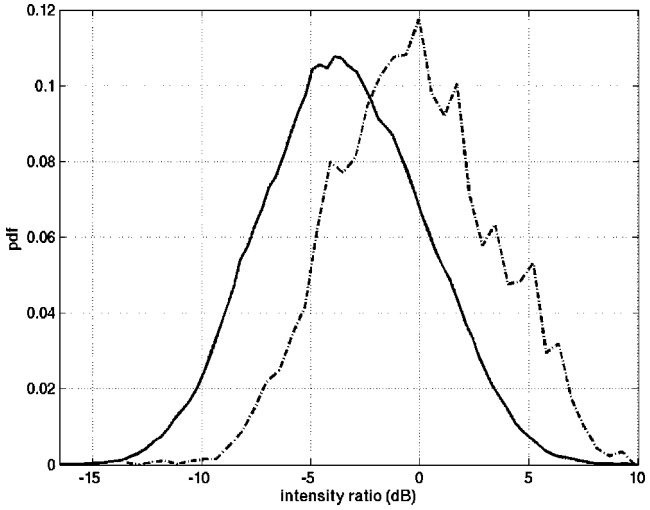


FIG. 14. Sample pdf of $10*\log_{10}$ of the ratio of the intensities of the direct path to surface bounce path arrivals. The *x*-axis scale indicates the amount by which the intensity of the first surface bounce path arrival exceeded that of the direct path arrival. The solid line shows the pdf calculated using the entire 30 min data set. The dashed–dotted line shows the pdf calculated using data only from within the interval extending from 100 ms before to 100 ms after the time of each wave focusing event.

rival, move in tap position (delay) as the surface elevation fluctuates. The pdf for the surface bounce path arrival is calculated both over the entire data collection period and over the intervals within 100 ms of each wave focusing event. The mean values for the peak intensity of the estimated direct arrival, the surface path arrival, and the surface path arrival within 100 ms of each wave focusing event, are -30.63 dB, -32.62 dB, and -29.50 dB, respectively.

The pdf for $10*\log_{10}$ of the ratio of the peak intensity of the estimated surface bounce path to the estimated peak intensity of the direct path for each estimated channel impulse response is shown in Fig. 14. The pdfs were calculated both over the entire data set and over only the data taken within 100 ms of each wave focusing event. For the full data set, the intensity of the estimated surface bounce path exceeds that of the direct path 17.7% of the time. For the data falling within 100 ms of each wave focusing event, the intensity of the estimated surface bounce path exceeds that of the direct path 45.7% of the time.

The impact of surface wave focusing on the intensity of the surface scattered arrival are clearly illustrated by Figs. 13 and 14. Separate intensity statistics were also calculated for focusing events that showed the formation of a caustic and those that did not. The intensity pdfs for these two classes of focusing events were virtually identical. This indicates that, to the extent that the channel fluctuations are accurately tracked by the channel estimation algorithm, the intensity statistics during the passage of a wave do not depend on whether or not a caustic is formed.

The mean of the SER conditioned upon the delay of the first surface bounce path arrival is shown in Fig. 15. The data clearly shows that the mean SER increases as the delay increases. Since the delay of this arrival is maximum at or near the focusing event, this shows that the SER is generally maximized at or near the focusing events. This conditional mean was calculated from using the entire data set.

Scatter plots of the mean SER taken over intervals extending from 100 ms before to 100 ms after each focusing event are shown in Fig. 16. There were some focusing events in the data for which the data are not included in Fig. 16. These data were from events involving waves for which multiple wave peaks overlapped or other features made it difficult to calculate either a wave height or focal length for the portion of the wave that resulted in the focusing. The data for the events that formed caustics shows a clear trend of increasing SER with both a decreasing focal length and increasing wave height. The data for events that did not form caustics shows only a weak dependence of SER on focal length but a clear trend of increasing SER with increasing wave height. This is in contrast to the lack of a dependence

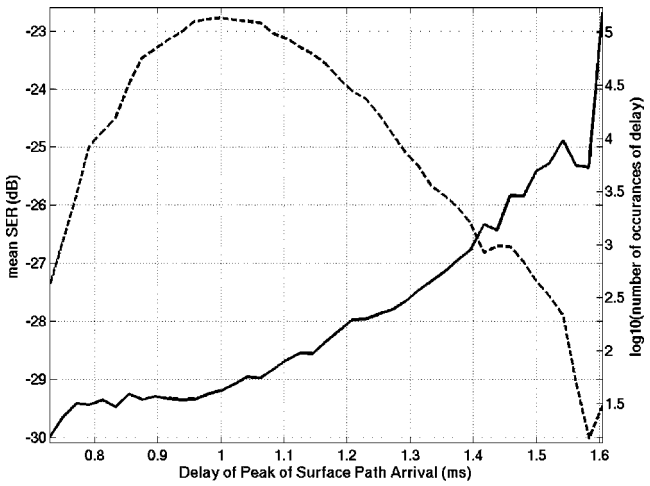


FIG. 15. Sample mean SER conditioned on the delay of the peak of the first surface bounce path arrival. The solid line (scale on left vertical axis) is the mean SER and the dashed line (scale on right vertical axis) represents the number of occurrences in logarithmic units of the peak of the first surface bounce path arrival being located at the indicated delay.

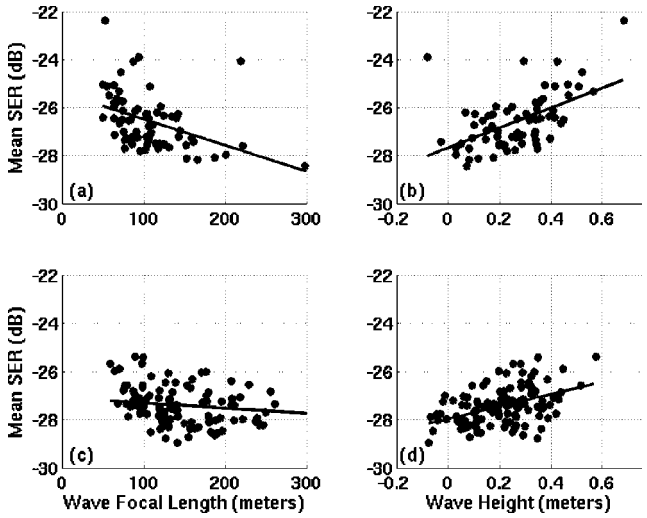


FIG. 16. Sample mean SER during the passage of each surface wave as a function of wave focal length and wave height. The surface wave events are divided into those which produced a clear caustic focusing event [(a) and (b)] and those which did not [(c) and (d)]. The line in each plot is a linear fit to the data in each plot. The mean for each wave was taken over an interval extending from 100 ms before to 100 ms after the time of each wave focusing event. For comparison, the mean SER over the entire 30 min data set is −28.61 dB.

of arrival intensity statistics on whether or not a caustic is formed as discussed earlier.

The total number of events recorded was 185. Of these, 123 events had surface wave focal lengths of greater than 100 m. Thirty five (28%) of these events showed the formation of a caustic while 88 (72%) did not. Sixty two events had surface wave focal lengths of less than 100 m. Thirty four (55%) of these events showed the formation of a caustic while 28 (45%) did not. These statistics show that the likelihood of the formation of a caustic increases for surface waves with focal lengths approaching the propagation path length of approximately 40 m. Of the 34 events with focal lengths less than 100 m and that showed the formation of a caustic, 13 (38%) resulted in a SER of greater than −25.76 dB. Of the 28 events with focal lengths less than 100 m and that did not show the formation of a caustic, only 4 (14%) resulted in a SER of greater than −25.76 dB. Thus, the increased likelihood of the formation of a caustic by short focal length waves also results in a generally increased SER indicating poor tracking of the channel impulse response and an inability to exploit the increased surface arrival energy.

The data shows that the SER statistics depend on whether or not a caustic is formed but the arrival intensity statistics do not. This difference may be the result of several factors. First, the result may accurately represent the propagation physics. Second, the small scale surface roughness that is not measured by the pressure sensor array (see Sec. II) may be an important element in controlling the intensity fluctuations of the surface scattered arrival. That is, there may be surface waves that have scales large enough to focus sound with an 8.3 cm wavelength but which are too small to be accurately measured by bottom mounted pressure sensors in approximately 6 m of water. These waves may also have surface normal velocities that are too small to result in significant Doppler shifts in the surface scattered arrivals. If this were the case, these small scale waves may significantly impact the surface scattered arrival intensity statistics but not the SER statistics and would not be reflected in the measured surface wave data. Finally, the higher SER during focusing events in which a caustic is formed may indicate that there are higher intensity arrivals during these events that the channel estimation algorithm is unable to track. Hence, these higher intensity arrivals would not be reflected in the intensity statistics presented here. The currently available data is not sufficient to resolve this issue.

V. CONCLUSIONS

The acoustic focusing caused by the scattering of signals off of surface gravity waves in the surf zone gives rise to arrivals that have both high intensity and are rapidly fluctuating. In the experimental data presented, the intensity of these arrivals at times exceed that of the direct arrival by almost 10 dB. During the peak of the passage of a surface wave, the simultaneous presence of advancing and retreating wave-fronts in the surface scattered arrival results in a Doppler spread of the arrival by up to 30 Hz. The high intensity and rapid fluctuation of the surface scattered arrival during a focusing event significantly degrades the performance of a least squares algorithm for estimating the channel impulse response. This will significantly degrade the performance of phase coherent acoustic communications systems that must either implicitly or explicitly estimate the channel impulse response.

The evaluation of surface wave and acoustic data from a single focusing event shows close agreement between model and experimental results. The model data clearly shows the important features in the focusing event including the formation of caustics. The experimental data clearly shows a peak in signal estimation residual error (SER) at a particular phase of the focusing event. However, the currently available data is not sufficient to resolve the precise features of the channel impulse response that results in an error spike at these times. The evaluation of data from 185 passing surface waves over 30 min of data collection showed that for those 69 focusing events that showed the formation of a caustic, the SER showed a strong increase with the decreasing focal length of the wave and with the increasing height of the wave. For those events that did not show the formation of a caustic, the SER showed little dependence on the focal length of the wave but did show a increase with increasing wave height.

The overall conclusion that can be drawn is that surface wave focusing causes an increase in the amplitude of the surface scattered arrivals at a receiver and a degradation in the ability of an estimation algorithm to track the fluctuations in the channel impulse response. For events involving the formation of a caustic at the receiver, the degradation can be related to the wave parameters of focal length and wave height. Additional work is needed to further resolve the characteristics of the fluctuations in the channel impulse response in the vicinity of the caustic and the primary cause of the degradation in algorithm performance. Two methods of resolving these issues would be to conduct tests with shoaling surface waves propagating in a well controlled tank environment and to conduct tests using the transmission of short

acoustic pulses with a level sufficient to yield reliable estimates of the channel impulse response without the need for temporal averaging.

ACKNOWLEDGMENTS

The authors are pleased to acknowledge Dr. Chris Tindle for many insightful discussions and creating a version of his Wavefronts model suitable for use in the surf zone. We would also like to acknowledge Dr. Dale Stokes for his assistance with the field work. This work has been supported by ONR Grant Nos. N00014-96-1-0120, N00014-00-1-0303, N00014-99-1-0274, and N00014-00-1-0048. This paper is WHOI Contribution No. 11018.

APPENDIX: CHANNEL IMPULSE RESPONSE AND SCATTERING FUNCTION ESTIMATION

The acoustic signals received from each of the experiments were processed to yield estimates of the time-varying impulse response of the acoustic channel. The received signals for the maximum length shift register sequence (m-seq) transmissions (see Sec. II) were modulated to baseband, lowpass filtered, and then downsampled to a rate of two samples per symbol. The resulting baseband signals were then used in combination with the original data sequence as the inputs to a least squares channel identification algorithm to generate estimates of the baseband channel impulse response. The duration of the modulation pulse used in the transmitted signals was approximately 0.04 ms. The impulse response estimates derived from the data are actually estimates of the channel impulse response convolved with this modulation pulse and have resolution in delay of approximately 0.04 ms.

The particular form of the baseband channel impulse response estimated by this method is the input delay-spread function.[8] The channel input/output relationship for the sampled input delay-spread function is $y[n]=\Sigma_{m=0}^{N_c} g^*[n,m]d[n-m]$, where $g[n,m]$ denotes the baseband input delay-spread function as a function of output sample index, n, and sample delay, m, N_c is the number of causal samples in the delay dimension of the input delay-spread function ($N_c=167$ was used), the superscript $*$ denotes complex conjugate, $d[n]$ is the baseband transmitted data sequence, and $y[n]$ is the sampled baseband output of the channel. Letting

$$\mathbf{g}[n]=\begin{bmatrix} g[n,0] \\ \vdots \\ g[n,N_c] \end{bmatrix}, \quad \mathbf{d}[n]=\begin{bmatrix} d[n] \\ \vdots \\ d[n-N_c] \end{bmatrix},$$

and $v[n]$ denote the baseband observation noise, the baseband received signal $y[n]$ is given by $y[n]=\mathbf{g}[n]^h\mathbf{d}[n]+v[n]$. Here, the superscript h denotes Hermitian and boldface letters denote vectors.

The least squares algorithm used a sliding, rectangular data window to accommodate time variation in the channel impulse response. The delay spread of the estimated channel impulse response was 7 ms (168 symbol periods) and the rectangular averaging window was 25 ms (600 symbol periods) in length. The algorithm was run with averaging windows ranging from 10 to 45 ms and the best results in terms of minimizing the signal estimation residual error were achieved using the 25 ms averaging window. At each symbol period, the estimated impulse response vector is given by $\hat{\mathbf{g}}[n]=\arg\min_g \Sigma_{k=0}^{599}|y[n-k]-\mathbf{g}^h\mathbf{d}[n-k]|^2$. This estimated impulse response was saved every 5 ms (120 symbol periods). The fractional spacing of the received baseband signal at 2 samples per symbol was accommodated by running 2 independent channel identification algorithms (one for odd numbered samples and the other for even number samples) and interleaving the resulting channel impulse response estimates.

The signal estimation residual error (SER) is given by $e[n]=y[n]-\hat{\mathbf{g}}[n-1]^h\mathbf{d}[n]$ and is used as a measure of how well the estimated channel impulse response approximates the actual channel impulse response. That the energy in $e[n]$ is a reasonable measure of the agreement between the estimated and actual channel impulse response is justified in the following paragraph.

The estimate of the impulse response at time n using received signal data collected up to time $(n-1)$ is denoted by $\hat{\mathbf{g}}[n-1]$. Then the residual prediction error can be expressed as

$$\begin{aligned} e[n] &= y[n]-\hat{\mathbf{g}}[n-1]^h\mathbf{d}[n] \\ &= (\mathbf{g}[n]-\hat{\mathbf{g}}[n-1])^h\mathbf{d}[n]+v[n]. \end{aligned}$$

Assuming that the baseband observation noise and data sequence are zero-mean and uncorrelated, the baseband data sequence has a variance of one and is a white sequence, the variance of the baseband observation noise is σ_v^2, and that the impulse response is uncorrelated with both the baseband observation noise and data sequence, the variance of the residual prediction error is give by

$$E|[e[n]]=\sigma_e^2[n]=\|\mathbf{g}[n]-\hat{\mathbf{g}}[n-1]\|^2+\sigma_v^2.$$

Here the expectation is with respect to the observation noise and baseband transmitted data sequence. Thus, with the assumption that the variance of the observation noise is constant, σ_v^2, the signal estimation residual error is a reasonable measure of the 2-norm of the difference between the actual and estimated channel impulse response.

The channel scattering function[8] was estimated by matched filtering frequency shifted versions of the received baseband signal with a windowed 4095 point m-seq. The window was used to reduce sidelobe levels at the expense of mainlobe width. The window used was a Kaiser window with a shape parameter of 3. This yielded a maximum sidelobe level of -23.8 dB relative to the peak mainlobe level and a two-sided mainlobe 3 dB width of 6.4 Hz.

[1] J. A. Smith, "Performance of a horizontally scanning Doppler sonar near shore," J. Atmos. Ocean. Technol. **10**, 752–763 (1999).

[2] D. M. Farmer, G. B. Deane, and S. Vagle, "The influence of bubble clouds on acoustic propagation in the surf zone," IEEE J. Ocean. Eng. **26**, 113–124 (2001).

[3] K. L. Williams, J. S. Stroud, and P. L. Marston, "High-frequency forward scattering from Gaussian spectrum, pressure release, corrugated surfaces. I. Catastrophe theory modeling," J. Acoust. Soc. Am. **96**, 1687–1702 (1994).

[4] C. T. Tindle, "Wavefronts and waveforms in deep-water sound propagation," J. Acoust. Soc. Am. **112**, 464–475 (2002).

[5] M. J. Buckingham, "Precision correlations between the geoacoustic pa-

rameters of an unconsolidated, sandy marine sediment,'' J. Comput. Acoust. **9**, 101–123 (2001).
[6] J. G. Proakis, *Digital Communications*, 3rd ed. (McGraw–Hill, New York, 1995), Chap. 13, pp. 724–729.
[7] G. B. Deane, ''The beam forming properties of a concave spherical reflector with an on-axis receiver,'' J. Acoust. Soc. Am. **106**, 1255–1261 (1999).
[8] P. A. Bello, ''Characterization of randomly time-variant linear channels,'' IEEE Trans. Commun. Electron. **CS-11**, 360–393 (1963).
[9] M. Stojanovic, ''Recent advances in high-speed underwater acoustic communications,'' IEEE J. Ocean. Eng. **21**, 125–136 (1996).
[10] M. Stojanovic, J. Catipovic, and J. Proakis, ''Adaptive multichannel combining and equalization for underwater acoustic communications,'' J. Acoust. Soc. Am. **94**, 1621–1631 (1993).
[11] M. Johnson, L. Freitag, and M. Stojanovic, ''Improved Doppler tracking and correction for underwater acoustic communications,'' Proc. ICAASP'97, Munich, Germany, 1997, pp. 575–578.
[12] T. Eggen, A. Baggeroer, and J. Preisig, ''Communication over Doppler spread channels—Part I: Channel and receiver presentation,'' IEEE J. Ocean. Eng. **25**, 62–71 (2000).
[13] T. Eggen, J. Preisig, and A. Baggeroer, ''Communication over Doppler spread channels—Part II: Receiver characterization and practical results,'' IEEE J. Ocean. Eng. **26**, 612–621 (2001).
[14] M. Stojanovic, L. Freitag, and M. Johnson, ''Channel-estimation-based adaptive equalization of underwater acoustic signals,'' OCEANS'99 Conference, Seattle, WA, 1999, pp. 985–990.
[15] M. Stojanovic, ''Efficient acoustic signal processing based on channel estimation for high rate underwater information,'' J. Acoust. Soc. Am. (submitted).

Acoustic absorption measurements for characterization of gas mixing

Aurelien Cottet, Yedidia Neumeier, David Scarborough, Oleksandr Bibik, and Tim Lieuwen[a)]
School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0150

(Received 30 March 2004; revised 28 June 2004; accepted 2 July 2004)

Controlling and/or monitoring the degree of mixing between constituents of a multicomponent media is a key problem in a variety of applications. Monitoring such mixing processes necessarily requires capabilities for quantification of the level of ''mixedness.'' However, quantification of *molecular* mixedness levels, as opposed to macroscale mixture uniformity, is difficult. This paper demonstrates the use of acoustic absorption measurements to characterize an average level of molecular mixedness between gases across the wave propagation path. This approach takes advantage of the fact that over a large frequency range, acoustic damping is dominated by vibrational relaxation processes. The vibrational relaxation frequency for a particular gas is often a strong function of the other species it is in molecular contact with. Thus, the relaxation frequency of each species in a multicomponent gas mixture varies with the level of molecular mixedness of the constituent species. This paper presents the results of example calculations and experiments demonstrating the feasibility of this approach and the significant sensitivity of acoustic absorption levels upon gas mixedness; e.g., measurements reported here show acoustic amplitude differences of up to a factor of 10 between identical gas mixtures whose only difference is the level of mixedness of their constituents.
[DOI: 10.1121/1.1785631]

PACS numbers: 43.35.Ae, 43.20.Hq [RR] Pages: 2081–2088

I. INTRODUCTION

Controlling, enhancing, and/or monitoring the degree of mixing between constituents of a multicomponent media is a key problem in a variety of applications; e.g., in industrial chemical processes or in developing fuel/air mixers for low emissions combustors.[1–3] In particular, for applications where the rate of chemical reactions is controlled by the molecular collision rate between mixture constituents, control and/or enhancement of mixedness at the molecular level, as opposed to simply macroscale entrainment, is needed. Optimization of such mixing enhancement strategies necessarily requires capabilities for quantification of the level of ''mixedness.'' However, measuring *molecular* levels of mixedness, as opposed to macroscale mixture uniformity, is difficult. For example, consider visual or photographic assessment of the degree of mixing between two nonreacting species with different colors. Such a system can readily determine the degree of macroscale mixture uniformity. However, this visual assessment necessarily ''averages out'' the features of the mixture over the entire range of scales between that of the measurement system's resolution and the molecular free path.

Various definitions are used in the literature to quantify mixedness. The simplest is simply the ratio of the spatial and/or temporal standard deviation (σ) to the mean (μ) of the species concentration, σ/μ. An entropy measure, defined spatially as $\int Y(x)\ln(Y(x))dx$, where $Y(x)$ denotes the spatial profile of the species mass fraction, is also frequently used,[4] as well as related definitions that use other normalization or scaling factors.

The objective of this paper is to demonstrate an acoustic technique to characterize the level of gas mixedness at the molecular level. The use of acoustic absorption measurements as gas phase mixture diagnostics has received limited attention to date. The majority of existing acoustic diagnostic techniques have relied on sound speed or Doppler shift measurements to infer quantities such as average temperature, composition of binary mixtures, or flow velocity.[5–8] Absorption measurements for diagnostic purposes have been reported in some recent studies for gas composition characterization[8] or particle sizing.[9] The technique described here is unique in that it based upon the fact that the relaxation frequency for a particular gas is often a strong function of the other species it is in molecular contact with. Thus, the relaxation frequency of a dual component gas mixture varies with the level of molecular mixedness of the constituent species.

Before proceeding to the detailed discussion, we provide a brief review of acoustic absorption via molecular relaxation processes. Acoustic absorption processes arise from viscosity, thermal conductivity, gas diffusion in multicomponent mixtures, and rotational and vibrational relaxation of internal energy modes.[10–12] Of most interest here are vibrational relaxation processes, which damp acoustic waves through the following mechanism: gas compression associated with the acoustic wave cause perturbations in the local translational energy. Through molecular collisions, the translational energy is redistributed to the rotational and vibrational degrees of freedom. During low frequency acoustic perturbations, en-

[a)]Electronic mail: tim.lieuwen@aerospace.gatech.edu

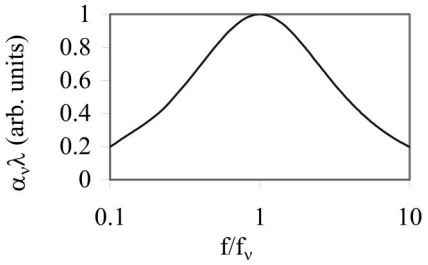


FIG. 1. Dependence of the acoustic absorption coefficient upon the normalized frequency.

ergy is fed from the translational to the vibrational energy modes of the gas during the compressive phase of the acoustic cycle and returned during the rarefaction phase. While gas translational and rotational modes equilibrate very quickly (generally within a few collisions), the vibrational mode requires extensively longer (several thousand collisions). Thus, even at relatively low frequencies the internal energy exchange processes cannot respond sufficiently quickly to acoustic fluctuations, with the effect that energy is not removed from or returned to the dilatational disturbance in phase with its oscillations. These nonequilibrium effects cause the acoustic wave to be damped. Maximum acoustic damping per wavelength, λ, of acoustic propagation distance occurs at the ''relaxation frequency,'' f_v. Following Pierce,[13] Fig. 1 plots the dependence of the acoustic absorption upon normalized frequency, f/f_v. In the figure, the quantity α_v refers to the acoustic damping coefficient due to vibrational relaxation; i.e., a wave propagating through a distance x is reduced in amplitude by the factor $\exp(-\alpha_v x)$. These vibrational nonequilibrium effects provide the dominant damping mechanism in most gases over a wide range of frequencies; e.g., for frequencies less than about 1 MHz in air.[13] An additional, relatively minor, effect of these nonequilibrium phenomenon is to slightly increase the speed of sound; i.e., a typical change is on the order of 0.1% of the nominal sound speed.

The fundamental principle motivating this study is the fact that the *relaxation frequency of a particular gas is often a function of the other species with which it is in molecular contact*. Numerous studies have demonstrated this point in various gases over a wide range of temperatures and pressures, such as nitrogen,[14,15] oxygen,[16,17] carbon monoxide,[18] and carbon dioxide.[19–24] For example, Fig. 2 reproduces Knudsen and Fricke's measurements of the vibrational relaxation frequency, f_v, of carbon dioxide (CO_2) mixtures upon the concentration of water vapor (H_2O), methanol (CH_3OH), and several other gases.[21] These data show that f_v in CO_2 increases substantially with the addition of these gases. For example, the addition of only 0.5% of any of these gases increases the relaxation frequency of CO_2 by a factor of at least 5. We should point out that more recent measurements of these relaxation frequency characteristics have, in some cases slightly modified the value of the change in relaxation frequency (e.g., see Ref. 25 for the H_2O case), but not the overall trends illustrated in the plot.

This sensitivity of a species' relaxation frequency to the presence of another species implies that the relaxation frequency of a binary mixture composed of two or more gases with such properties will vary with their level of molecular mixedness. In other words, if the two gases are completely unmixed, the vibrational relaxation rates of each species is simply equal to its values in isolation. In contrast, if the gases are mixed, the relaxation frequency has a different value that depends upon the relative constituent concentrations. The objective of this paper is to demonstrate this point.

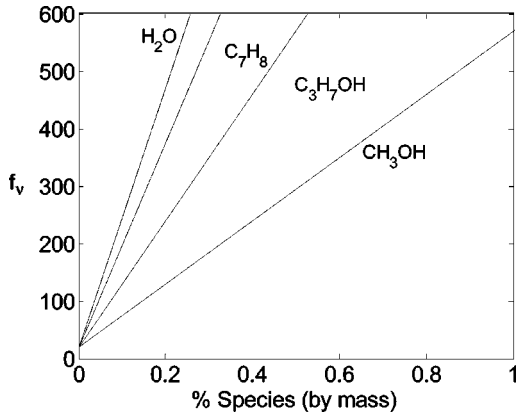


FIG. 2. Dependence of f_v in carbon dioxide mixtures with the indicated species, as reported by Knudsen and Fricke (see Ref. 21).

The rest of the paper is organized in the following manner: The following section provides theoretical calculations using simulated mixedness profiles and tabulated relaxation frequency data. The objective of this section is to theoretically demonstrate the feasibility of the concept. Then, we present experimental results for CO_2–propane (C_3H_8) and CO_2–H_2O mixtures that verify the strong effect of mixedness levels upon acoustic absorption levels. Data included here show that in some cases, the acoustic amplitude changes by a factor of 10 between two identical gas mixtures, whose only difference is the level of mixedness of their constituent gases.

II. EXAMPLE CALCULATIONS

This section presents theoretical calculations illustrating the effects of gas mixedness on acoustic absorption using simulated mixedness profiles and standard acoustic absorption formulas. These theoretical calculations use the following formula's presented by Pierce[13] for acoustic absorption and neglect the effect of absorption by gas diffusion. The total absorption coefficient, α, can be written as $\alpha=\alpha_{\mathrm{cl}}+\sum_v \alpha_v$, where α_{cl} is the classical absorption coefficient. It represents the sum of the viscous and thermal absorption coefficients and is given by

$$\alpha_{\mathrm{cl}}=\frac{\omega^2\eta}{2c^2}\left(\frac{4}{3}+\frac{\eta_B}{\eta}+\frac{(\gamma-1)}{\mathrm{Pr}}\right), \tag{1}$$

where η, η_B, γ, Pr, c, and ω denote the dynamic viscosity, bulk viscosity, ratio of specific heats, Prandtl number, sound speed, and angular frequency, respectively. Note that the earlier equation assumes that the frequency is much less than the rotational and translational relaxation frequencies, a reasonable assumption for the $f<100$ kHz frequency range ex-

perimentally explored in the next sections. The vibrational absorption coefficient, α_v, equals

$$\alpha_v = \left[\frac{\pi(\gamma-1)}{\gamma}\frac{1}{\lambda}\right]\left[\frac{c_{v,V}}{c_V}\right]\left[\frac{\omega\tau_v}{1+(\omega\tau_v)^2}\right], \tag{2}$$

where τ_v and $c_{v,V}/c_V$ denote the vibrational relaxation time and specific heat due to the vibrational modes, normalized by the total specific heat. Note that the fundamental principle behind the technique described here is the dependence of this relaxation time, τ_v, upon mixedness level. The specific heat due to the vibrational modes, $c_{v,V}$, equals

$$\frac{c_{v,V}}{R} = \sum_j \frac{(\theta_{v,j}/T)^2 e^{\theta_{v,j}/T}}{(e^{\theta_{v,j}/T}-1)^2}, \tag{3}$$

where $\theta_{v,j}$, T, and R denote the vibrational temperature of the molecule's jth degree of freedom, gas temperature, and gas constant, respectively. The summation is over the j number of vibrational degrees of freedom for the specific molecule.

Example calculations using these equations are next presented. For these calculations, CO_2 was used as the primary gas because of its relatively low vibrational temperature, resulting in high levels of absorption at room temperatures in the 10–90 kHz range. Many other candidate species with higher vibrational temperatures (e.g., N_2 or O_2) have low levels of vibrational mode excitation at room temperature, making the total amount of absorption relatively small at these frequencies and over distances on the order of the experiment size. Water vapor was used as the second gas. The calculations use the values of $c=266$ m/s, $\eta=0.000\,085$ m^2/s, $T=291$ K, Pr$=0.78$, $\eta_B/\eta=0.6$, $\theta_{v,1}=\theta_{v,2}=959$ K, and $\theta_{v,3}=1920$ K. Relaxation time values, τ_v, of each particular CO_2-H_2O combination were determined from Fig. 2.

Figure 3 plots the dependence of the total absorption, $\exp(-\alpha y)$, upon frequency and water vapor concentration, using Eqs. (1)–(3) and the relaxation frequency data from Fig. 2. This calculation was performed assuming a distance $y=18$ cm, corresponding to the size of the experiment described in the next section. The figure shows that, at a fixed H_2O level, the absorption coefficient monotonically increases with frequency. Note, however, the nonmonotonic dependence of the absorption coefficient upon H_2O level at a fixed frequency. This nonmonotonicity is due to the shift in relaxation frequency with H_2O level. The nonlinear dependence of the acoustic absorption level upon (in this case) H_2O content shown in the figure is key to the technique being pursued here. Because of this nonlinearity, the total absorption of a sound wave traversing the medium depends upon the local concentration values, and not just the integrated sum across the line of sight. If these relations were purely linear, the absorption level would not change with mixedness levels. For example, consider the mixing of equal parts of a pure CO_2 and $CO_2+0.1\%$ H_2O mixture, such that the final mixture consists of $CO_2+0.05\%$ H_2O. Figure 3 shows that the difference in absorption levels between the mixed and unmixed states is relatively small at frequencies below about 30 kHz, but rapidly increases at higher frequencies.

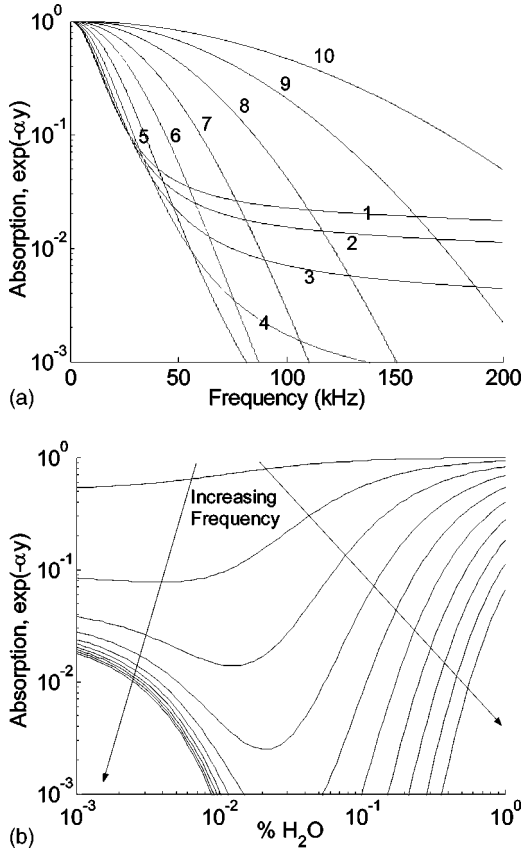


FIG. 3. Dependence of total acoustic absorption, $\exp(-\alpha y)$, across a distance $y=0.18$ m upon frequency for several CO_2-H_2O mixtures; curves numbered to indicate (left) H_2O levels of 0.001%, 0.002%, 0.005%, 0.01%, 0.02%, 0.05%, 0.1%, 0.2%, 0.5%, and 1% by mass and (right) frequencies of 10, 30, 50, 70, 90, 110, 130, 150, 170, and 190 kHz.

To quantify the change in absorption levels with mixing, consider a spatially nonuniform mixture of CO_2 and H_2O, where the spatial profile of the H_2O is plotted in Fig. 4. These profiles closely simulate those produced in the experiments described in the next section. They represent a two-dimensional analogue of the mixing between an inner and an outer, coflowing annular gas stream, a geometry commonly encountered in a variety of applications.

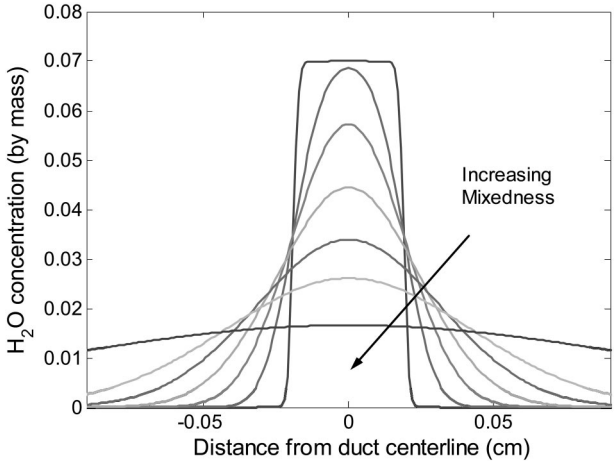


FIG. 4. Assumed spatial dependence of H_2O mass fraction profile for model problem.

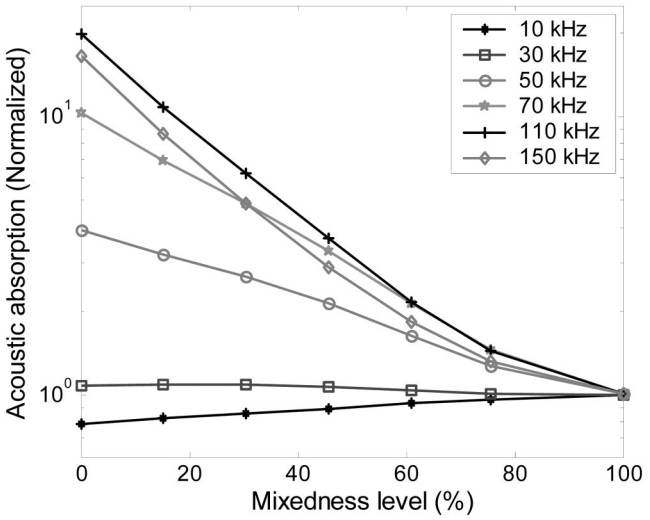


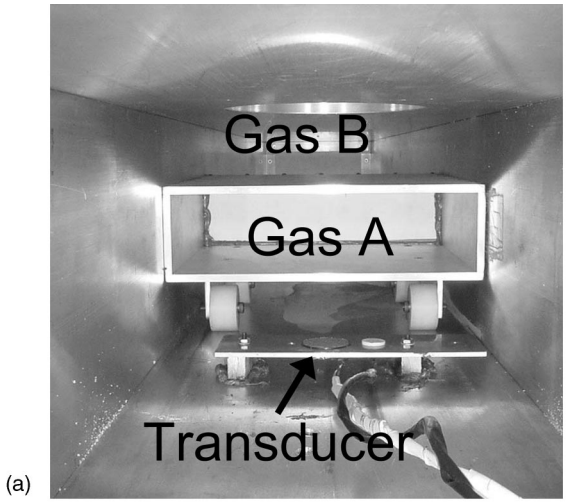
FIG. 5. Dependence of acoustic absorption upon mixedness levels at frequencies of 10, 30, 50, 70, 90, 110, 130, and 150 kHz using assumed water vapor profile illustrated in Fig. 4.

Figure 5 plots the dependence of the acoustic absorption upon mixedness levels at several frequencies. The mixedness value indicated in the figure was quantified by linearly scaling the ratio of the profile's standard deviation to its mean, σ/μ, between zero and unity

$$\text{mixedness level} = 1-(\sigma/\mu)/(\sigma_{\max}/\mu) = 1-\sigma/\sigma_{\max}, \tag{4}$$

where $\sigma_{\max}$ denotes the maximum value of the profile standard deviation (i.e., corresponding to its most unmixed state). As such, mixedness values of zero and unity correspond to the completely unmixed and mixed cases, respectively.

At each frequency, the absorption level is normalized by its value in the perfectly mixed case. The expression $\exp(-\int \alpha dy)$ was used to determine the total absorption (i.e., it is assumed that acoustic reflections are negligible), where α was calculated at each location using the local composition. The figure shows that rather large differences exist between the wave amplitude in the perfectly mixed and unmixed cases. In this example, the acoustic amplitude increases and decreases with mixedness levels for frequencies below and above 30 kHz, respectively. At 30 kHz, the acoustic amplitude exhibits minimal sensitivity to mixedness level, but the amplitude increases by about 25% at 10 kHz and decreases by 2000% at 110 kHz. The sensitivity of the absorption levels upon mixedness peaks at about 110 kHz in this example. Note that this frequency of peak sensitivity is a function of the mixture composition, a fact which can be appreciated from the results in Fig. 3. The differences between acoustic amplitude in the perfectly mixed and unmixed cases becomes monotonically smaller as the mixedness levels increase from the normalized levels of zero to unity. As such, it seems unlikely that the differences would be significant enough to distinguish accurately between two nearly homogeneous mixtures, say with mixedness levels of 0.95 and 0.99. However, reasonable differences exist even between mixedness levels of 0.8 and 1.0 at the higher frequencies. For example, the amplitude decreases by 50% in the 75–150 kHz frequency range.



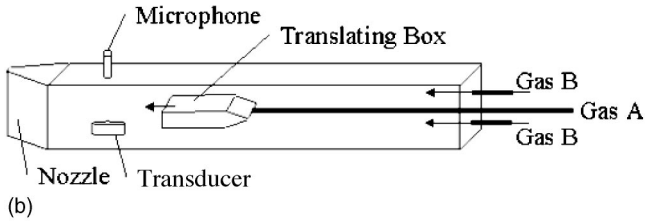


FIG. 6. Photo (flow coming out of page) and drawing of facility developed for acoustic absorption measurements.

III. EXPERIMENTAL FACILITY

Having theoretically demonstrated the feasibility of gas mixedness characterization using acoustic absorption measurements, this section describes companion experiments to demonstrate the technique. A photograph and drawing of the facility are shown in Fig. 6. It consists of a 244 cm long, 29×21 cm section aluminum duct that is capped at one side and open at the other. A translating box 31×29×6.4 cm with one open end slides inside the duct. All joints in the duct are sealed to prevent contamination of the mixture inside. Although removed for the photograph, a nozzle is fitted to the end of the box to minimize back diffusion of gas from the ambient air into the box. This nozzle is fitted to flexible hosing connected to the building exhaust. Two gases of arbitrary composition, denoted as gas "A" and gas "B," flow through the main chamber and the translating box and mix at the translating box exit. Great effort was expended to maintain a "clean" flow so that the two gases mix in a steady, laminar fashion. In order to minimize flow unsteadiness and flow separation, the translating box was fabricated to have an aerodynamic profile. First, its width tapers down at the back to a thin edge so that the outer flow can cleanly divide around it. In addition, the edge of the box where the two gases initially mix is machined to a 0.05 cm width to minimize the wake thickness.

For this experiment, gases A and B consisted of two CO_2 sources with different levels of some secondary gas.

Both H_2O and propane (C_3H_8) were used as the secondary gases. However, as will be discussed further later, the substantially different molecular weight between H_2O and CO_2 introduced a host of measurement problems. For this reason, the majority of data were taken with C_3H_8, which has essentially the same molecular weight as CO_2. Dry, industrial grade CO_2 (99.98% purity, <10 ppm water vapor) flows through the main part of the chamber and around the translating box (gas A). The water vapor concentration of the CO_2 actually entering the box was higher, however, due to such factors as leakage at fittings and hygroscopic absorption into the plastic tubing. In cases where H_2O was used as the secondary gas, CO_2 was humidified by passing it through a bubbler. Its humidity was systematically varied by adjusting the relative concentrations of this stream and a second dry CO_2 stream, which is then introduced through the translating box. H_2O/CO_2 mass ratios of 0%–0.3% were achieved with this arrangement. The water vapor concentration was measured with a Tri-Sense relative humidity sensor, model EW-37000-00. In cases where C_3H_8 was used as the secondary gas, a flow control panel was used to mix the desired concentrations from 0% to 20% by mass of C_3H_8 in CO_2. The C_3H_8 concentration was measured with an infrared absorption technique. Gas was removed from the measurement volume via a 1.3 cm diameter sampling probe. Light from a broadband (2.5–14 μm) thermal radiation source was passed through the gas in a cell and into an infrared spectrometer set at 3.3 μm, the approximately wavelength of maximum C_3H_8 infrared absorption (CO_2 absorption at this wavelength is negligible). The dependence of the spectrometer output voltage to C_3H_8 concentration was determined from off-line calibration experiments that spanned the C_3H_8 ranges encountered in these experiments.

All gas flows are metered with calibrated rotameters with velocities from 1 to 5 cm/s. This velocity range was specified based upon diffusive mixing calculations, with the requirement that the two gases could be completely mixed when the translating box is positioned the farthest distance from the transducer (120 cm). As such, the degree of gas mixing can be varied via the box location and/or the gas velocity.

Acoustic disturbances are generated with a 38 mm diameter, type 616341 electrostatic Polaroid transducer. Acoustic measurements are obtained with a 1/8 in. type 4191 Bruel and Kjaer microphone. The speaker and microphone are situated near the downstream end of the duct and can be seen in Fig. 6. The gas was acoustically interrogated by either driving the transducer with a persistent sinusoidal signal or with bursts of 10–50 acoustic cycles at the desired frequency, followed by a pause in order to allow the reverberant field to damp out.

IV. UNMIXEDNESS RESULTS

In a typical experiment, two initially unmixed gas streams enter the facility in the positions noted in Fig. 6. At the point where gas A exits the translating box, it mixes with gas B. As these gases convect down the duct and mix, the acoustic amplitude is measured at different axial locations, accomplished by pulling the translating box away from the transducers.

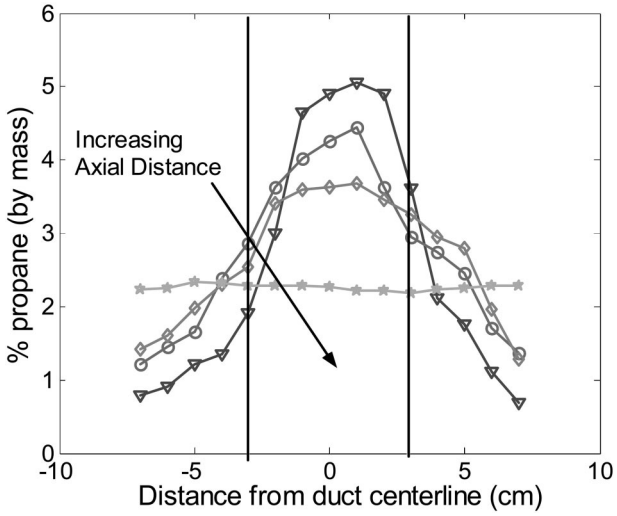


FIG. 7. Measured transverse propane concentration profile at mixing distances of 0, 2.5, 5, and 120 cm (axial flow velocity=3 cm/s). Vertical lines denote translating box edges.

Typical transverse C_3H_8 concentration profiles at different axial locations are shown in Fig. 7. The two vertical lines indicate the edges of the translating box. It can be seen that the profile at the translating box exit is smoothed out somewhat from the discontinuous change in concentration, at least partially due to the averaging of the 1.3 cm diameter measurement probe in the transverse direction. As expected, the profiles smooth out with increasing axial mixing distance and, at 120 cm, the propane concentration is completely uniform. Ideally, the total integrated C_3H_8 concentration across the flow at each axial location should stay nearly constant. Typical differences were on the order of 1%–5%, but never more than 10%, providing an indication of the accuracy of these concentration measurements.

The typical dependence of the acoustic amplitude upon axial mixing distance, normalized by its values at the perfectly mixed $x=120$ cm axial location, are shown in Fig. 8. The figure shows the substantial, generally monotonic, increase in acoustic amplitudes as the gases mix. Data are not shown for 80 or 90 kHz because the absorption was so high that the amplitude could not be accurately measured in all cases. The dependence of the ratio of the acoustic amplitude in the most unmixed and mixed, $A(x=0)/A(x=1.2\,\mathrm{m})$, cases upon frequency is plotted in Fig. 9 (the result from another test with a slightly higher propane concentration is also included). It shows that the acoustic amplitude increases by factors of roughly 2 and 10 in the 30 and 70 kHz cases, making the point that the effect of mixedness upon acoustic amplitude is by no means minor! The figure also shows a comparison of the predicted dependence of this ratio upon frequency, calculated using the measured absorption at each C_3H_8 level in the perfectly mixed case (calibration measurement of the absorption levels in the perfectly mixed case are described in the Appendix) and assuming that the total absorption could be approximated as $\exp(-\int \alpha dy)$. The two re-

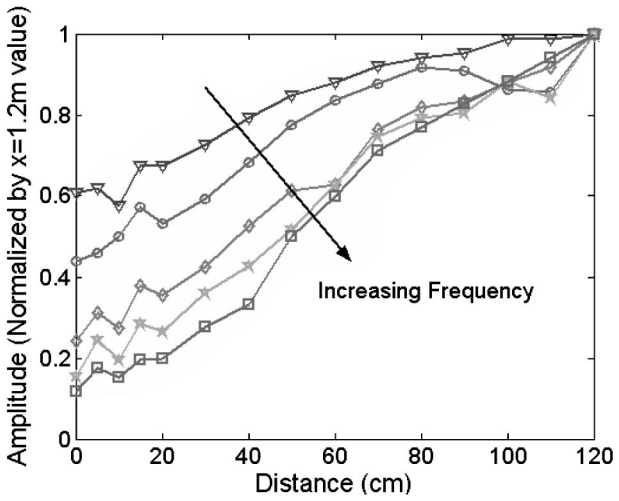


FIG. 8. Dependence of 30, 40, 50, 60, and 70 kHz amplitude signal upon axial distance between gas stream origination point and acoustic transducer (initial propane concentration of 8.5%, axial flow velocity=3 cm/s).

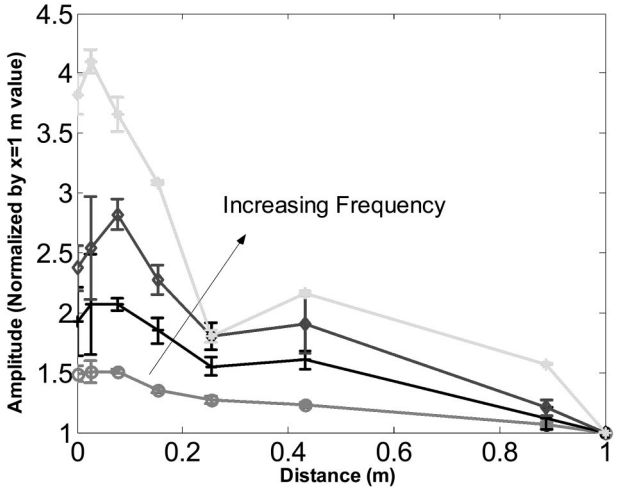


FIG. 10. Dependence of 30, 40, 50, and 60 kHz amplitude signal upon axial distance between gas stream origination point and acoustic transducer (approximate H_2O concentration at translating box exit is 0.045%). Error bar denotes level of fluctuation of signal amplitude.

sults have similar qualitative trends, although the theoretically predicted frequency sensitivity of the amplitude ratio is significantly less than what was experimentally observed. These differences are due in part to the sensitivity of the theoretical result to the exact details of the C_3H_8 profile. For example, the true profile likely changes much more sharply at the translating box edges than is shown in Fig. 7, but is smeared out in the measurements because of the transverse averaging of the measurement probe. Sensitivity studies were performed to evaluate the manner in which the predicted amplitude ratio would change with deviations in C_3H_8 profile. By using a ''flat-top'' exit profile whose concentration changed discontinuously from 2% in the outer flow to 8.25% in the translating box, the frequency sensitivity changed to approximately 0.85 and 0.35 at 30 and 70 kHz, respectively. These values are much closer to the measurements.

Analogous data were obtained for the case where water vapor was used at the secondary gas. In these studies, accurate measurements of the transverse water profile concentrations could not be made due to the poor sensitivity of the humidity sensor at the low concentrations used. In addition, humidity levels and pressure amplitude measurements were often very unsteady. Obtaining good, repeatable data from this facility was challenging because of the strong role of buoyancy effects at these low flow velocities. Consequently, good data were obtained under limited conditions where the H_2O concentration remained relatively steady. Note that the H_2O concentration tends toward a vertically stratified profile (i.e., highest water concentrations at the top) at large mixing distances.

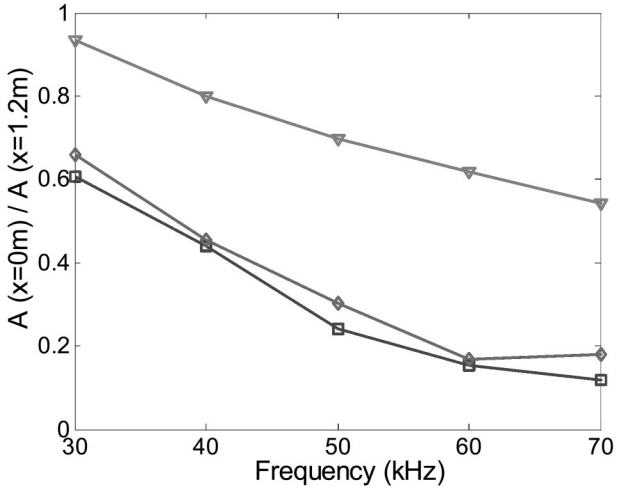


FIG. 9. Comparison of theoretical, ▽, (based upon measured C_3H_8 concentration at box exit in 8.5% propane case) and measured (initial C_3H_8 concentrations at translating box exit of 8.5%, ◇, and 9.4%, □) ratio of acoustic amplitude at $x=0$ and 120 cm. Axial flow velocity=3 cm/s.

Figure 10 illustrates the dependence of the 30–60 kHz acoustic amplitude, normalized by its value at $x=1.0$ m. It shows a monotonic decrease in acoustic amplitude with increased mixedness. The amplitude of the error bar corresponds to the fluctuation level of the signal. Note that the decrease in amplitude observed here, compared to the increase observed in the earlier data, is simply due to the concentration levels of the secondary gas; opposite trends can be observed with either gas at other concentration levels. The dependence of the ratio of the acoustic amplitude at the two axial extremes $(A(x=0)/A(x=1.0\,\mathrm{m})$ upon frequency is plotted in Fig. 11. As earlier, the error bar indicates the level of fluctuation of the signal amplitude. It shows that the acoustic amplitude decreases by factors of roughly 1.5 and 4 in the 30 and 60 kHz cases.

While the data shown in Figs. 8–11 deliberately focused on steady state, laminar mixing cases so that the corresponding gas concentration profiles could be readily measured, ample data were obtained in cases where unsteady mixing is clearly prominent. For example, by intentionally causing the flow velocities of the outer flow and the translating box to differ, the unsteady rollup of the shear layer between the two streams can be clearly observed. This behavior can be seen in Fig. 12, where the average velocity of the outer flow has intentionally been set 60% higher than in the translating box.

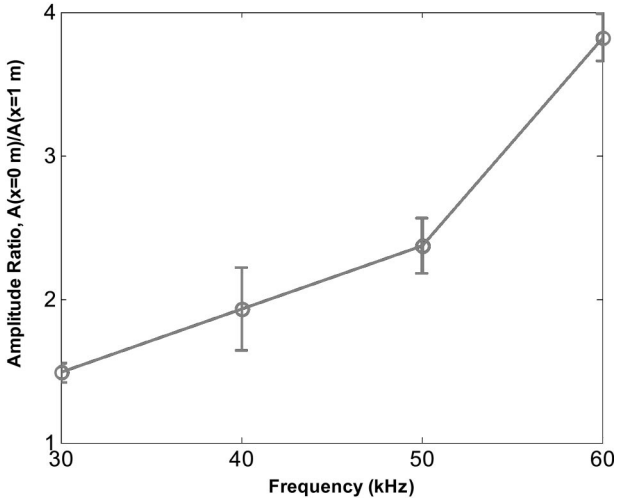


FIG. 11. Dependence of ratio of acoustic amplitude at $x=0$ and 100 cm upon frequency. Error bars denote level of fluctuation of signal.

Pure CO_2 is used in both flows in curve (1) to illustrate that the amplitude is essentially constant in the absence of unsteady species mixing, demonstrating that the oscillation observed in the signal amplitude in other cases is not due to velocity fluctuations. Approximately 30% propane is added to the translating box flow in curves (2–7), which are taken at successive axial locations downstream. At the translating box exit (curve 2), the signal oscillates only slightly. The level of amplitude oscillations (as well as the average amplitude) grows monotonically with downstream distances of 5, 15, and 30 cm (curves 3, 4, and 5). The unsteadiness in amplitude is particularly evident at 30 cm. The amplitude of oscillations then decreases with further distance, indicating that the two streams are nearly mixed. Although not shown, at 120 cm the signal is essentially flat with no oscillation.

V. CONCLUDING REMARKS

This paper has theoretically and experimentally demonstrated the feasibility of using acoustic absorption measure-

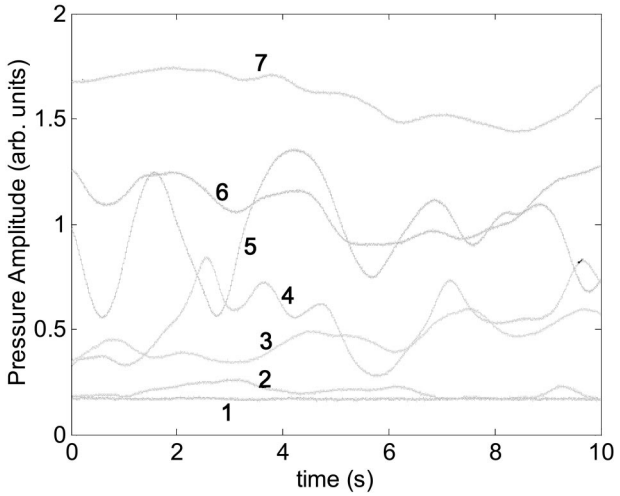


FIG. 12. Temporal dependence of amplitude of 50 kHz signal with (1) pure CO_2, and (2)–(7) $C_3H_8-CO_2$ mixtures at axial distances of $x=0$, 5, 15, 30, 40, and 60 cm (average velocities of 1.3 and 2.1 cm/s in the translating box and outer flow, respectively; 30% propane in translating box flow in cases 2–7).

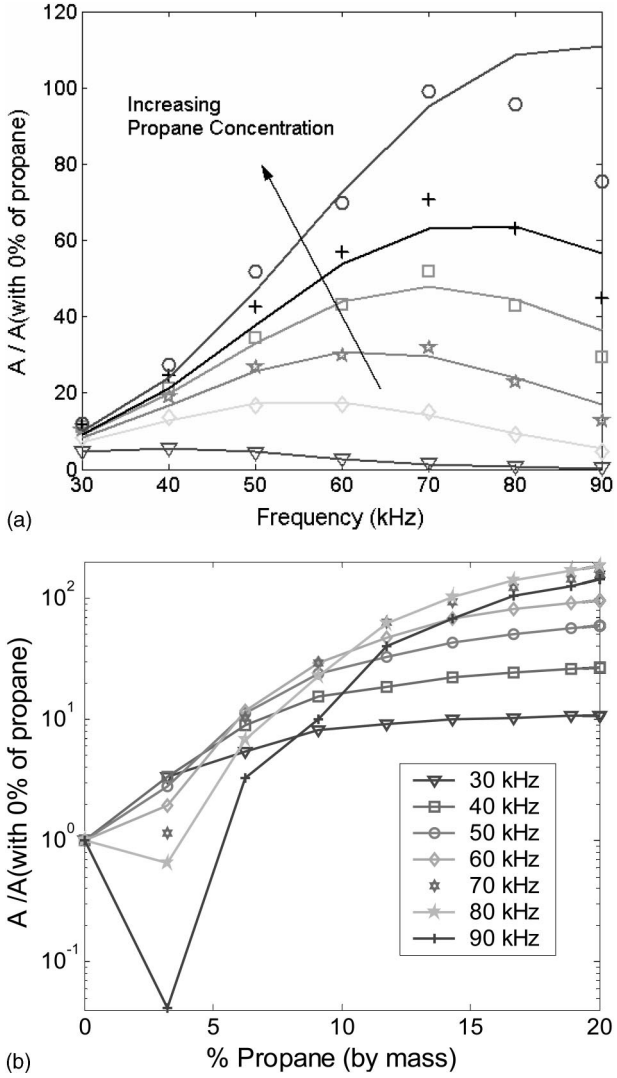


FIG. 13. Dependence of measured (symbols) and, in left plot, fitted (line) acoustic amplitude at the indicated propane concentration, normalized by its pure CO_2 value at C_3H_8 concentrations of 0%, 3.2%, 6.3%, 9.1%, 11.7%, 14.3%, 16.6%, 18.9%, and 20%. Only measured data are shown in bottom plot.

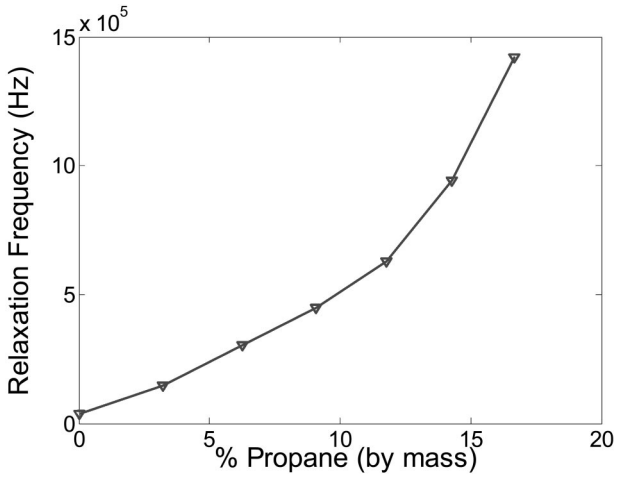


FIG. 14. Measured dependence of f_v in $C_3H_8-CO_2$ mixtures using data shown in Fig. 13.

ments to characterize levels of gas mixedness along a line of sight. In closing it should be pointed out the implementation of the technique is gas specific, depending upon the nature of the vibrational-translational and vibrational–vibrational interactions between the gases being mixed. The technique may not work in all situations, such as with two monotonic gases or two gases with little effect upon each other's vibrational relaxation rates. However, a possible way to extend the technique to such cases would be to mix one or more of the gases with a third gas that does impact the other's relaxation rates.

APPENDIX—PROPANE CALIBRATION MEASUREMENTS

A literature review did not find data on C_3H_8 effects on CO_2 relaxation rates, so extensive calibration data of these mixtures at room pressure and temperature were obtained so that the unmixedness measurements could be compared to a theoretical value. We obtained these calibration data by measuring the acoustic absorption between 30 and 90 kHz across uniform flowing mixtures in the facility shown in Fig. 6, using propane levels ranging between 0%–20% by mass. These data are shown in Fig. 13. In order to facilitate theoretical predictions, we fitted these data by assuming the absorption is only due to CO_2. While this fit is most physically meaningful at lower C_3H_8 concentrations, it nonetheless provides a convenient method for collapsing the absorption data into a simple set of coefficients for insertion into a theoretical model. Rather than attempting to backout an absorption coefficient by determining the amplitude at two different wave propagation distances with the appropriate diffraction corrections, we normalized the measured amplitude at each C_3H_8 concentration level by the amplitude using pure CO_2. The best fit value of the relaxation frequencies at each propane level were determined using these data and Eqs. (1)–(2). The resulting curves are plotted with the data in Fig. 13, as well as the calculated relaxation frequencies that were used to generate them in Fig. 14. Figure 13 shows that the curve fit satisfactorily fits the data, except in the 70–90 kHz regions at the two highest C_3H_8 levels, probably for the reasons given earlier.

ACKNOWLEDGMENT

This research was supported by the Army Research Office, Dr. David Mann, technical monitor.

[1] J. Oldshue, *Fluid Mixing Technology* (Chemical Engineering, New York, 1983).
[2] A. H. Lefebvre, *Gas Turbine Combustion* (Edwards Brothers, Ann Arbor, 1999), Chap. 9, pp. 311–375.
[3] J. M. Ottino, *The Kinematics of Mixing: Stretching, Chaos, and Transport* (Cambridge University Press, New York, 1989).
[4] R. Everson, D. Manin, L. Sirovich, and M. Winter, "Quantification of mixing and mixing rate from experimental observations," AIAA J. **36**, 121–127 (1998).
[5] W. Wright, D. Schindel, D. Hutchins, P. Carpenter, and D. Jansen, "Ultrasonic tomographic imaging of temperature and flow fields in gases using air-coupled capacitance transducers," J. Acoust. Soc. Am. **104**, 3446–3455 (1998).
[6] M. Mori, Y. Takeda, T. Taishi, N. Furuichi, M. Aritomi, and H. Kikura, "Development of a novel flow metering system using ultrasonic velocity profile measurement," Exp. Fluids **32**(2), 153–160 (2002).
[7] H. Winter, E. Mair, and H. Schoenework, "Device for monitoring the concentration of an air-vapor mixture," U.S., Patent No. 4,424,703 (7 May, 1984).
[8] S. Phillips, Y. Dain, and R. Lueptow, "Theory for a gas composition sensor based on acoustic properties," Meas. Sci. Technol. **14**, 70–75 (2003).
[9] P. Mougin, D. Wilkinson, K. Roberts, and R. Tweedie, "Characterization of particle size and its distribution during the crystallization of organic fine chemical products as measured *in situ* using ultrasonic attenuation spectroscopy," J. Acoust. Soc. Am. **109**, 274–230 (2004).
[10] A. Bhatia, *Ultrasonic Absorption* (Dover, New York, 1967).
[11] H. O. Kneser, "Relaxation processes in gases," in *Physical Acoustics*, edited by W. P. Mason (Academic, New York, 1965), Vol. IIA, pp. 133–202.
[12] H. Bass, "Introduction to elastic constants of gases," in *Handbook of Elastic Properties of Solids, Liquids, and Gases*, edited by M. Levy, H. Bass, and R. Stern (Academic, New York, 2001), Vol. IV, pp. 235–265.
[13] A. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York, 1991).
[14] R. Millikan and D. White, "Vibrational energy exchange between N_2 and CO. The vibrational relaxation of nitrogen," J. Chem. Phys. **39**, 1803–1808 (1963).
[15] H. Bauer and H. Roesler, "Relaxation of the vibrational degrees of freedom in binary mixtures of diatomic gases," in *Molecular Relaxation Processes* (Academic, New York, 1966).
[16] D. White and R. Millikan, "Vibrational relaxation of oxygen," J. Chem. Phys. **39**, 98–101 (1961).
[17] D. White and R. Millikan, "Vibrational relaxation in air," AIAA J. **2**, 1844–1846 (1965).
[18] C. Von Rosenberg, R. Taylor, and J. Teare, "Vibrational relaxation of CO in nonequilibrium nozzle flow, and the effect of hydrogen atoms on CO relaxation," J. Chem. Phys. **54**, 1974–1987 (1971).
[19] T. Cottrell and M. Day, "The effect of noble gases on vibrational relaxation in carbon dioxide," in *Molecular Relaxation Processes* (Academic, New York, 1966).
[20] R. Higgs and R. Torborg, "Ultrasonic absorption in carbon dioxide-water vapor mixtures," J. Acoust. Soc. Am. **42**, 1038–1040 (1967).
[21] V. O. Knudsen and E. Fricke, "The absorption of sound in CO_2, N_2O, COS, and in CS_2 containing added impurities," J. Acoust. Soc. Am. **12**, 255–259 (1940).
[22] J. W. Lewis and F. D. Shields, "Vibrational relaxation in carbon dioxide/helium mixtures," J. Acoust. Soc. Am. **41**, 100–102 (1967).
[23] C. B. Moore, R. E. Wood, B. L. Hu, and J. T. Yardley, "Vibrational energy transfer in CO_2 lasers," J. Chem. Phys. **46**, 4222–4231 (1967).
[24] J. C. Stephenson, R. E. Wood, and C. B. Moore, "Vibrational relaxation of laser excited CO_2-polyatomic mixtures," J. Chem. Phys. **56**, 4813–4816 (1972).
[25] J. W. Lewis and K. P. Lee, "Vibrational relaxation in carbon dioxide/water vapor mixtures," J. Acoust. Soc. Am. **38**, 813–816 (1965).

Lamb mode conversion in a plate with a delamination

G. Shkerdin[a] and C. Glorieux[b]
Laboratorium voor Akoestiek en Thermische Fysica, Departement Natuurkunde, K. U. Leuven, Celestijnenlaan 200 D, B-3001 Leuven, Belgium

(Received 18 May 2004; revised 25 June 2004; accepted 28 June 2004)

A theoretical study of Lamb mode propagation through a part of a plate containing a finite-length, delamination parallel with the surface is presented. In the delamination boundary region, which is taken parallel with the free plate surface, noncontact boundary conditions are assumed. The calculation is based on a modal decomposition method. As a result of diffraction on a delamination the incident Lamb mode is efficiently converted into Lamb modes with wave numbers close to the wave number of incident mode. The transmission coefficient of a Rayleigh wave incident on a delamination located near the surface has an oscillating dependence on the delamination parameters and has a pronounced minimum where there is a strong conversion into transmitted Lamb modes. Inversely, using the method of phase conjugation, a proper incident Lamb mode combination can be efficiently converted into a single transmitted Lamb or Rayleigh wave. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1784435]

PACS numbers: 43.35.Ns, 43.35.Zc [YHB] Pages: 2089–2100

I. INTRODUCTION

The interaction of acoustic waves with defects inside materials has been widely studied during the past years. This problem is very important in the field of nondestructive characterization of materials. In particular delaminationlike defects are likely to grow in mechanically constrained structures, and finally lead to global failure or malperformance. The propagation of Lamb modes in plates turns out to be quite convenient to study the peculiarities of acoustic wave interaction with delaminations situated both in the depth and near the surface of a plate. Interaction of Lamb modes with different kinds of defects including delaminations was extensively studied.[1–8] In particular, B. A. Auld and M. Tan[1,2] predicted the reflection of Lamb modes from vertical delaminations normal to the plate surfaces and this effect was later studied by many authors (see, for example, Refs. 6–8). Special attention was also given by several authors to calculating the interaction of Lamb waves with defects in materials for technological applications.[9–13] Modal decomposition and finite element methods were used and compared.[7]

S. Rokhlin studied Lamb modes diffraction on symmetrical, semi-infinite, and finite delaminations parallel to the plate surfaces and situated exactly in the middle of the plate, when there is no mode conversion between symmetric and antisymmetric Lamb modes.[3–5] Solutions found in these papers include the existence of nonpropagating Lamb modes to satisfy the boundary conditions. Also reflection and transmission of an incident Gaussian acoustic beam through the plate with a horizontal delamination were analyzed and it was shown that the existence of a delamination leads to a considerable phase shift of the nonspecular lobe in reflected beam.[14]

The interaction of Lamb modes with a delamination was mostly analyzed for the case of a thin plate when only the lowest symmetric and antisymmetric Lamb modes exist. This is justified by the problems of Lamb modes testing. In practice, it is difficult to excite a pure, single Lamb mode in a multimode plate. Taking into account that Lamb modes are dispersive, this leads to problems in making a correct interpretation of the signal. From this point of view the multimode regime is not attractive to experimentally study the effect of mode conversion on a delamination in a multimode plate. On the other hand, the lowest symmetric and antisymmetric Lamb modes in a multimode plate can combine to a nondispersive Rayleigh wave. Therefore the study of Lamb mode conversion on a delamination in a multimode plate gives the possibility to analyze the effect of conversion between the Rayleigh wave and other Lamb modes.

The present paper is concerned with two-dimensional propagation of Lamb modes through a (horizontal) plate containing a horizontal delamination for the case of a multimode plate. The calculation is based on a modal decomposition method. In the first part of the paper the fundamentals of the method are given and an infinite system of coupled equations for mode amplitudes is derived. Lamb mode conversion on a delamination is studied in the second part of the paper. Special attention is paid to the case of conversion between a Rayleigh wave and Lamb modes.

II. THE METHOD OF ACOUSTIC FIELD CALCULATION

The two-dimensional geometry under consideration is shown in Fig. 1. The incident acoustic wave is represented by Lamb modes propagating along the Z direction in an isotropic plate of thickness d containing a finite rectangular horizontal delamination of length L. It is supposed that plate and delamination surfaces are stress-free, and that the delamination thickness is larger than normal elastic displace-

[a]Permanent address: Institute of Radio Engineering and Electronics of Russian Academy of Sciences, Vvedensky sq., 1, Fryazino, Moscow region, Russia.
[b]Author to whom correspondence should be addressed. Also with Fonds voor Wetenschappelijk Onderzoek—Vlaanderen. Electronic mail: christ.glorieux@fys.kuleuven.ac.be

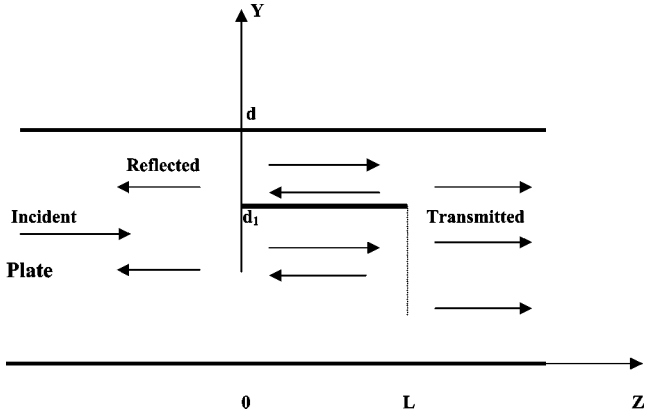


FIG. 1. Geometry of Lamb mode propagation in the structure under consideration.

ments at its surfaces. On the other hand, the delamination thickness Δ is taken small enough; $\Delta \ll \min(d,\lambda_s)$ (λ_s is the shear wave wavelength in a plate), such that the delamination can be treated as infinitesimally thin. The distance between the delamination and the lower plate surface is d_1.

The interaction of acoustic waves with the delamination leads to the generation of reflected acoustic waves in the region $z<0$, reflected and transmitted acoustic waves in the region $0<z<L$, and transmitted acoustic waves in the region $z>L$. It is possible to expand all waves in separate Lamb eigenmodes for the regions $z<0$; $z>L$; $0<z<L$, $0<y<d_1$ and $0<z<L$, $d_1<y<d$, since the eigenmodes represent a complete and orthogonal set of modes in each region. One can obtain the profiles of these eigenmodes by solving the elastic equations, using stress–strain relations, and applying stress-free boundary conditions at the plate borders.[15] We write the components of the stress tensor S_{ik}^n and the mechanical displacement vector U_i^n ($i=y,z$ and $k=y,z$) of the nth Lamb mode as follows:

$$(S_{ik}^n, U_i^n) = (\sigma_{ik}^n, u_i^n)\mathrm{Exp}(i(q_n z - \omega t)) + (\sigma_{ik}^n, u_i^n)^* \,\mathrm{Exp}(-i((q_n)^* z - \omega t)). \quad (1)$$

Here q_n and ω are Lamb mode wave number and angular frequency, and an asterisk means complex conjugation. Displacements U_i^n and stresses S_{ik}^n are coupled by stress–strain relations.

There are several types of orthogonality conditions. The most general orthogonality condition, also applicable to absorbing plates, is written down for the geometry of Fig. 1 (see the Appendix):

$$\int_0^d dy(\sigma_{yz}^n u_y^m + \sigma_{yz}^m u_y^n - \sigma_{zz}^n u_z^m - \sigma_{zz}^m u_z^n) = A(n)\,\delta_{n,m} \quad (2a)$$

with $\delta_{n,m}$ the Kronecker delta symbol. For $n=m$, $A(n)$ is a normalization factor. Since the product of stress and displacement changes sign when the propagation direction is reversed, in Eq. (2a) the normalization constants of forward and backward propagating modes have opposite signs. For nonabsorptive plates, also another type of orthogonality condition exists. It can be obtained from Eq. (2a) by the substitutions: $(u_y^m, \sigma_{zz}^m) \rightarrow (u_y^m, \sigma_{zz}^m)^*$, $(u_z^m, \sigma_{yz}^m) \rightarrow -(u_z^m, \sigma_{yz}^m)^*$, rendering for propagating modes[16]

$$\int_0^d dy(\sigma_{yz}^n (u_y^m)^* + \sigma_{zz}^n (u_z^m)^* - u_y^n (\sigma_{yz}^m)^* - u_z^n (\sigma_{zz}^m)^*) = B(n)\,\delta_{n,m}. \quad (2b)$$

Here, for $n=m$, $B(n)$ is proportional to the energy flow. Orthogonality conditions (2b) exist only in nonabsorbing plates. Although these conditions are directly connected with acoustic wave flow of energy, they are only convenient for propagating modes, since all nonpropagating modes normalization factors $B(n)$ are equal to zero, i.e., all modes are orthogonal to themselves. Therefore, to find the expansion coefficients, orthogonality conditions (2a) are used in the paper.

Equation (2a) is valid for the regions $z<0$ and $z>L$. Analogous relations exist for the regions with $0<z<L$. One can verify that all pairs of Lamb eigenmodes with different wave numbers are orthogonal using condition Eq. (2a). Forward and backward propagating Lamb modes are also orthogonal, since for this case $q_z^n = -q_z^m$ and all terms in Eq. (2a) are cancelled. In general, both for propagating and nonpropagating modes, the normalization factor $A(n)$ is different from zero.

In the region $z>L$, the expansion of the transmitted (index tr) mechanical displacement vector $\vec{U}^{tr}$ in right-going eigenmodes $\vec{u}_{tr}^n$ is given by

$$\vec{U}^{tr} = \sum_n C_n^{tr} \vec{u}_{tr}^n Exp^{i(q_n z - \omega t)}, \quad (3)$$

with C_n^{tr} expansion coefficients, which are to be determined. In the region $z<0$, the expansion of the incident (index in) mechanical displacement vector $\vec{U}^{in}$ is similar, with coefficients C_n^{in}, and with the same right-going eigenmodes $\vec{u}_{tr}^n$ as in the $z>L$ region. An analogous expression can be given for the reflected (index ref) acoustic wave $\vec{U}^{ref}$ with the substitutions: $C_n^{tr} \rightarrow C_n^{ref}$, $\vec{u}_{tr}^n \rightarrow \vec{u}_{ref}^n$, and $q_n \rightarrow -q_n$.

For the region $0<z<L$, there are different expansions for the subregions $0<y<d_1$ and $d_1<y<d$. The expansion of the right-going (index tr) mechanical displacement vectors $\vec{U}_{\pm}^{tr}$ in the eigenmodes $\vec{u}_{\pm}^{tr}$ is found by applying the following substitutions in Eq. (3): $C_n^{tr} \rightarrow C_{n,\pm}^{tr}$, $\vec{u}_{tr}^n \rightarrow \vec{u}_{tr}^{n,\pm}$ and $q_n \rightarrow q_{n,\pm}$. The $\pm$ signs mark respectively transmission in the upper ($d_1<y<d$) and lower ($0<y<d_1$) parts of the delamination region. The expansion of the left-going (index ref) acoustic waves $U_{\pm}^{ref}$ in the region $0<z<L$ is given by

$$\vec{U}_{\pm}^{ref} = \sum_n C_{n,\pm}^{ref} \vec{u}_{ref}^{n,\pm} Exp^{i(-q_{n,\pm}(z-L) - \omega t)}. \quad (4)$$

The expressions for the expansions of stress tensor components are completely analogous to Eqs. (3) and (4).

To find the expansion coefficients of the transmitted and reflected waves the boundary conditions at the vertical planes $z=(0,L)$ can be used. They are found by applying continuity conditions for σ_{yz}, σ_{zz} and for the mechanical displacement vector components for $0 \leqslant y < d_1$ and $d_1 < y \leqslant d$. The boundary conditions at the horizontal interfaces $y=(0,d_1,d)$ in the plate and delamination regions are fulfilled automatically, because all proper Lamb modes satisfy the free boundary conditions at these interfaces.

The boundary conditions have to be fulfilled for every depth y. However, here we use a procedure to reduce this system of equations. Consider the integrated quantities $R^m_{1,\pm}$:

$$R^m_{1,\pm}=\int_{a_\pm}^{b_\pm} dy(S^{\rm tot}_{yz,\pm}u^{m,\pm}_{y,tr}+U^{\rm tot}_{y,\pm}\sigma^{m,\pm}_{yz,tr}-S^{\rm tot}_{zz,\pm}u^{m,\pm}_{z,tr} - U^{\rm tot}_{z,\pm}\sigma^{m,\pm}_{zz,tr}), \tag{5a}$$

where $S^{\rm tot}_{ik,\pm}$, $U^{\rm tot}_{i,\pm}$ $(i=y,z,\ k=y,z)$ are components of the stress tensor, respectively the mechanical displacement vector of the total acoustic field on the right side of the plane $z=0$, $a_-=0$, $a_+=b_-=d_1$, $b_+=d$. Time-dependent exponents are omitted.

Using orthogonality condition Eq. (2) one obtains

$$R^m_{1,\pm}=A^\pm(m)C^{tr}_{m,\pm}. \tag{5b}$$

Here $A^\pm(m)$ is the normalization constant for the upper and lower part of the plate with the delamination. On the other hand, using the boundary conditions at $z=0$, Eq. (5) can be rewritten as follows:

$$\begin{aligned}R^m_{1,\pm}&=A^\pm(m)C^{tr}_{m,\pm}\\&=\int_{a_\pm}^{b_\pm} dy(S^{\rm tot}_{yz}u^{m,\pm}_{y,tr}+U^{\rm tot}_{y}\sigma^{m,\pm}_{yz,tr}-S^{\rm tot}_{zz}u^{m,\pm}_{z,tr} - U^{\rm tot}_{z}\sigma^{m,\pm}_{zz,tr})\\&=\sum_n (C^{in}_n(K^\pm_{n,m}+I^\pm_{n,m})+C^{ref}_n(K^\pm_{n,m}-I^\pm_{n,m})),\end{aligned} \tag{6}$$

where $S^{\rm tot}_{ik}$, $U^{\rm tot}_i$ are components of the stress tensor, respectively the mechanical displacement vector of the total (incident+reflected) acoustic field on the left side of the plane $z=0$, $I^\pm_{n,m}=\int_{a_\pm}^{b_\pm}dy(\sigma^n_{yz,tr}u^{m,\pm}_{y,tr}-u^n_{z,tr}\sigma^{m,\pm}_{zz,tr})$, $K^\pm_{n,m}=\int_{a_\pm}^{b_\pm}dy(\sigma^{m,\pm}_{yz,tr}u^n_{y,tr}-u^{m,\pm}_{z,tr}\sigma^n_{zz,tr})$.

To derive Eq. (6) the stress tensor and the mechanical displacement vector of the incident acoustic field on the left side of the plane $z=0$ were expanded in eigenmodes of the plate without delamination. These modes are not orthogonal to eigenmodes of the plate with delamination so that nonzero overlap integrals $I^\pm_{n,m}$, $K^\pm_{n,m}$ appear.

The same procedure is repeated for the following integral:

$$R^n_2=\int_0^d dy(S^{\rm tot}_{yz}u^n_{y,ref}+U^{\rm tot}_{y}\sigma^n_{yz,ref}-S^{\rm tot}_{zz}u^n_{z,ref} - U^{\rm tot}_{z}\sigma^n_{zz,ref}). \tag{7}$$

Equation (2a) and boundary conditions leads to the equation as follows:

$$A(n)C^{ref}_n=\sum_{m,\pm} (C^{ref}_{m,\pm}e^{iq_{m,\pm}L}(K^\pm_{n,m}+I^\pm_{n,m}) + C^{tr}_{m,\pm}(I^\pm_{n,m}-K^\pm_{n,m})). \tag{8}$$

The same kind of equation is written down at the plane $z=L$:

$$A^\pm(m)C^{ref}_{m,\pm}=\sum_n C^{tr}_n(K^\pm_{n,m}-I^\pm_{n,m}), \tag{9}$$

$$A(n)C^{tr}_n=\sum_{m,\pm} (C^{tr}_{m,\pm}e^{iq_{m,\pm}L}(K^\pm_{n,m}+I^\pm_{n,m}) + C^{ref}_{m,\pm}(I^\pm_{n,m}-K^\pm_{n,m})). \tag{10}$$

In Eqs. (8) and (9) normalization constants $A(n)$ and $A^\pm(m)$ are written for the forward propagating modes with positive real values of wave numbers.

As a result we obtain an infinite system of linear equations (6), and (8)–(10) in the unknown mode conversion coefficients $C^{tr,ref}_n$ and $C^{tr,ref}_{m,\pm}$. Because of the way the equations were constructed, there are exactly as many equations as unknowns. Typically there are a finite number of propagating, and an infinite number of nonpropagating mode coefficients. In order to reduce the calculation time, one has to restrict the number of modes taken into account, and thus solve for an approximate solution. Taking into account only transmitted propagating modes in all parts of the structure under consideration and setting the other coefficients (of propagating reflected and nonpropagating modes) equal to zero, one obtains the following solution of this system:

$$C^{tr}_p=\sum_{n,m,\pm} C^{in}_n e^{iq_{m,\pm}L}\frac{(K^\pm_{p,m}+I^\pm_{p,m})(K^\pm_{n,m}+I^\pm_{n,m})}{A(p)A^\pm(m)}. \tag{11}$$

In the summation, the index n counts over the incident modes, while the index m counts over all transmitted propagating modes. Though this simplification introduces inaccuracies in the calculation, we will show in the numerical treatment for thin plates in the next section that for many cases, provided the reflection coefficients are not too large, it is possible to get some assessment of the introduced error, using some energy considerations. In principle, taking into account the small number of propagating reflected modes would not greatly complicate the calculation. However, this would not useful, since the error introduced by neglecting an infinite number of nonpropagating modes would make the values of the reflection coefficients very unreliable.

The general expression of acoustic flow of energy P_z can be written down as follows:[17]

$$P_z=-\sum_{k=y,z}\int_0^d dyS_{kz}\frac{\partial U_k}{\partial t}, \tag{12}$$

where the stress tensor S_{kz} and the mechanical displacement vector U_k components are given by

$$(S_{kz}(y,z,t);U_k(y,z,t))=(S_{kz}(y,z);U_k)\mathrm{Exp}(-i\omega t) + (S_{kz}(y,z);U_k)^*\,\mathrm{Exp}(i\omega t). \tag{13}$$

Substituting the proper expansions [see Eq. (3)] in Eq. (13) and taking into account the mode orthogonality, the time averaged expression $\langle P_z\rangle$ for a nonabsorbing plate and $z<0$ can be written down as follows:

$$\langle P_z\rangle=\sum_{n_p} (|C^{in}_{n_p}|^2-|C^{ref}_{n_p}|^2)F_{n_p}, \tag{14}$$

where $C_{n_p}^{in}$ are expansion coefficients for the incident acoustic wave, $F_{n_p}=\mathrm{Im}(\int_0^d dy(\sigma_{zz,tr}^{n_p}(u_{z,tr}^{n_p})^*-(\sigma_{yz,tr}^{n_p})^* u_{y,tr}^{n_p}))$, and only propagating Lamb modes n_p are taken into account. The contributions of all nonpropagating modes to Eq. (14) is equal to zero.

An analogous expression $\langle P_z\rangle$ for $z>L$ is given by

$$\langle P_z\rangle=\sum_{n_p}|C_{n_p}^{tr}|^2F_{n_p}. \tag{15}$$

Taking into account the continuity of σ_{yz}, σ_{zz} and of the mechanical displacement vector components at $z=0$, L, one obtains that P_z is continuous at $z=0$, L. Therefore, for a nonabsorptive plate $\langle P_z\rangle$ does not depend on z and the following energy conservation equation is valid:

$$\sum_{n_p}|C_{n_p}^{in}|^2F_{n_p}=\sum_{n_p}(|C_{n_p}^{tr}|^2+|C_{n_p}^{ref}|^2)F_{n_p}. \tag{16}$$

Since Eq. (16) must always be satisfied, deviations from it, expressed by

$$\delta=(\Sigma_{n_p}(|C_{n_p}^{in}|^2-|C_{n_p}^{tr}|^2-|C_{n_p}^{ref}|^2)F_{n_p})/\Sigma_{n_p}(|C_{n_p}^{in}|^2)F_{n_p},$$

can be used to quantify a possible inaccuracy in calculated coefficients. The energy conservation mismatch parameter δ is obviously also closely related to the mismatch of the boundary conditions. In the next section, we will make use of this feature.

III. RESULTS OF NUMERICAL CALCULATIONS AND DISCUSSION

The propagation of a Lamb mode over a delaminated area leads in general to reflection. Significant reflection can occur when the phase velocities of incident Lamb modes and Lamb modes in the region $0<z<L$ are considerably different (see also Ref. 3). Such a situation takes place in the case of a thin plate. The phase velocities of fundamental modes S0 and A0 are shown in Fig. 2(a) for an aluminum plate with shear and longitudinal velocities equal to resp. 3440 and 5570 m/s, as a function of the dimensionless parameter q_td (q_t is the wave number of shear wave). The phase velocity dispersion is especially strong for $q_td<(5-6)$, when, due to the strong thickness sensitivity of the modes, the existence of the delamination can lead to a strong reflection.[3–5] To calculate transmission and reflection coefficients with satisfactory accuracy, it is necessary to take into account for this case a large number of nonpropagating modes.

As an example, Figs. 3(a) and 3(b) show calculation results for the case where there are only two propagating modes: A0 and S0. The dependences of the total transmission and reflection coefficients, defined as $(T,R)=(|C_{A0}^{tr,ref}|^2F_{A0}+|C_{S0}^{tr,ref}|^2F_{S0})/(|C_{A0}^{in}|^2F_{A0})$, and the relative energy conservation error versus the number of modes taken into account, are shown for an aluminum plate, with an incident A0 mode with frequency $f=1.5$ MHz, plate thickness $d=1$ mm ($q_td\approx2.74$), $d_1=0.8$ mm and for different values of the delamination length L. The acoustic displacement and characteristic equations for Lamb modes are given in the Appendix. Calculations were carried out solving Eqs. (6), and (8)–(10), truncating the infinite series as follows: as a starting point only two propagating modes S0 and A0 were taken into account. In the next step the only (for this q_td value) damped antisymmetric Lamb mode (with purely imaginary wave number) was added to the set of modes (three modes in total). Then pairs of the lowest symmetric and antisymmetric nonpropagating inhomogeneous Lamb modes with complex wave numbers were taken into account (five modes in total). Next, another pair of symmetric and antisymmetric inhomogeneous Lamb modes with larger values of real and imaginary parts of their wave numbers were taken into account (seven modes in total), and so on.

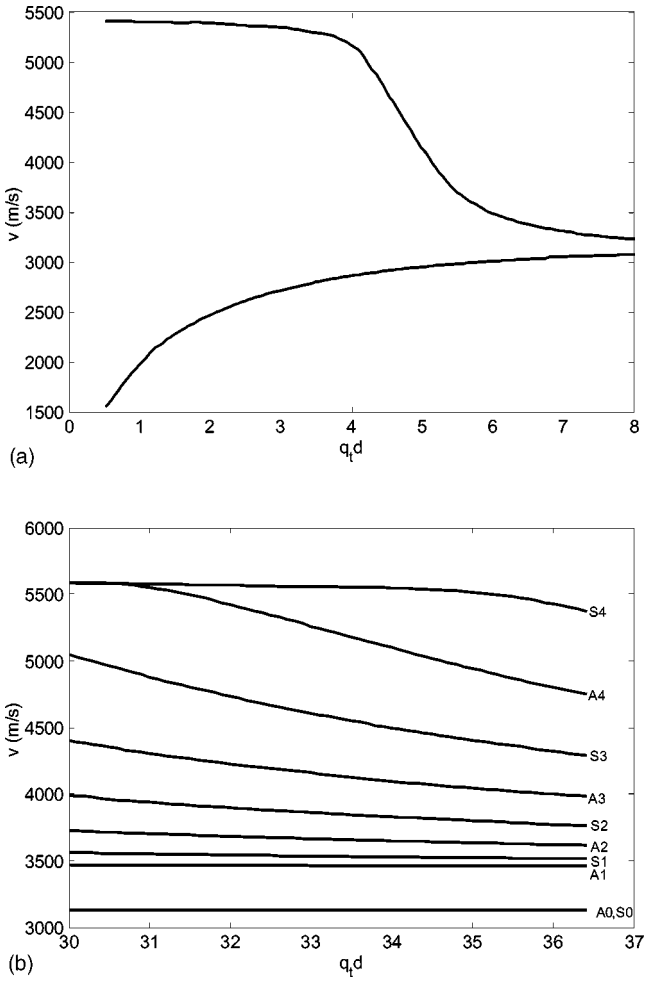


FIG. 2. (a) Dependence of fundamental S0 (upper curve) and A0 (lower curve) mode phase velocities on dimensionless plate thickness q_td. (b) Dependence of the velocity of the first five symmetric and five antisymmetric Lamb modes on the dimensionless plate thickness q_td.

In Figs. 3(a) and (b) results for $L=0$ are also given. This case corresponds with the limit of very small delamination length when $L\to0$ and the delamination reduces to a stress-free point. In this case the transmission coefficient $T\to1$, but, in taking the limit, formally, the local acoustic field nearby this short delamination represented by nonpropagating modes still exists. Therefore, it is necessary to take into account the contribution of these nonpropagating modes to obtain a rigorous solution. Figure 3(a) illustrates that $T\to1$ and taking into account only a small number of modes results in a small error for this case.

Calculation results for $L=2$ mm and $L=5$ mm are

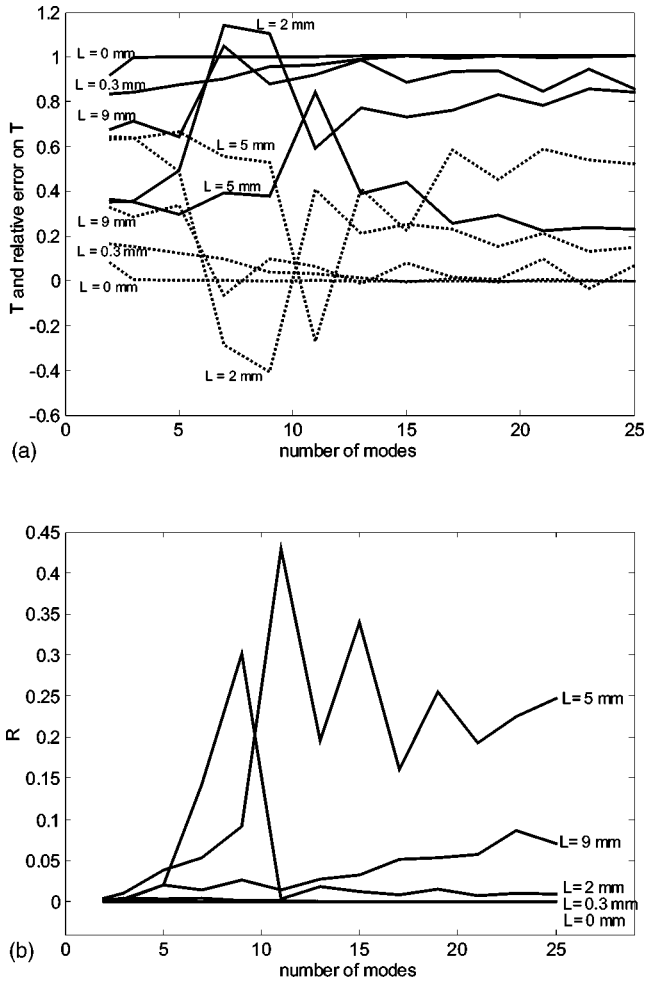


FIG. 3. (a) Dependence of the total transmission coefficient (full lines) and relative error δ (dotted lines) as a function of the number of modes taken into account for the case of an incident antisymmetric mode A0 in a plate. The calculations were done for an acoustic wave frequency $f=1.5$ MHz, plate thickness $d=1$ mm, and delamination depth $d_1=0.8$ mm. (b) Dependence of the total reflection coefficient versus the number of modes taken into account in the calculation, for the case of an incident antisymmetric mode A0. The calculation parameters are the same as in (a).

strongly oscillating as a function of the number of modes number and here the reflection can be considerable. Calculations show that the transmitted wave mainly contains the A0 mode (cf. the incident mode).

Figure 3(a) reveals that the initial error δ, obtained when only propagating modes are taken into account, is strongly indicative for more extensive mode expansions. If the initial error δ is considerably smaller than 1 ($\delta \leqslant 0.2$), then addition of nonpropagating modes reduces the error. In the case of large initial error $\delta \geqslant (0.4–0.5)$, taking into account of nonpropagating modes results in strong oscillations of δ around zero and reliable results can be obtained only by taking into account a very large number of nonpropagating modes. Reflection in this case can be also considerable. The strongly indicative value of the initial error allows us, especially for the more complicated cases, where a too large number of all nonpropagating modes renders a detailed calculation not feasible, to assess the overall error from the calculation which neglects all nonpropagating modes. If δ is small, the contribution of nonpropagating modes to the transmitted acoustic field is small and calculation results for transmitted propagating modes are of satisfactory accuracy. If δ is too large, then calculations give just a rough estimation of mode amplitudes. Only taking into account a very large number of nonpropagating modes would significantly improve the calculation accuracy. Amplitudes of propagating modes reflected from the delamination can not be calculated with satisfactory accuracy without taking into account the contribution of a large number of nonpropagating modes.

Let us now consider the case of a thicker plate, which is characterized by a weaker dispersion of Lamb mode phase velocities. Such an example is shown in Fig. 2(b) for a larger value of the parameter $q_t d$. Phase velocities are given for the first five symmetric and five antisymmetric Lamb modes. The modes S0 and A0 are almost not dispersive in the range of $q_t d \geqslant 30$. Exact solutions can be obtained only taking into account nonpropagating modes. This problem for the case of a thick, multimode plate is very complicated, since it is necessary to take into account a huge number of nonpropagating modes. If the contribution of all nonpropagating modes is neglected, then calculation errors are inevitable. However, even for an incident Rayleigh wave concentrated in the delamination region, the error in most cases stays below 15%. As mentioned before, if all nonpropagating mode contributions are neglected, then it is advised to just neglect the total reflection from the delamination as well, since neglecting propagating reflecting modes does not significantly increase the error further. In the following results, all calculations are based on Eq. (11), thus neglecting reflection and nonpropagating modes. The error is estimated by the δ value. All calculations are carried out for a plate with thickness $d=5$ mm and acoustic wave frequency $f=4$ MHz. In this case $q_t d \approx 36.53$ and there are ten propagating symmetric and ten propagating antisymmetric modes in the region of the plate without the delamination.

A. Mode conversions between Lamb modes

Transmission coefficients of the propagating modes can be defined as follows:

$$T_n = |C_n^{tr}|^2 F_n / \sum_m |C_m^{in}|^2 F_m . \tag{17}$$

Using Eq. (11), diffraction of an incident nth Lamb mode on a delamination leads to the generation of transmitted propagating Lamb modes with amplitudes $C_{p,n}^{tr}$ given by

$$C_{p,n}^{tr} = C_n^{in} \sum_{m,\pm} e^{iq_{m,\pm}L} \frac{(K_{p,m}^{\pm} + I_{p,m}^{\pm})(K_{n,m}^{\pm} + I_{n,m}^{\pm})}{A(p)A^{\pm}(m)} . \tag{18}$$

Typical results for the case of an incident S3 mode are shown in Fig. 4, where the transmission coefficient of the S3 mode, $T_{S3} = |C_{S3}^{tr}|^2 / |C_{S3}^{in}|^2$, and the relative error δ are plotted versus the delamination length L for different values of the delamination depth inside the plate. The T_{S3} values strongly depend on L and undergo considerable oscillations. The reason for these oscillations can be explained as follows. At $z=0$ the incident S3 mode is mainly converted into Lamb modes of the plate with the delamination with the wave numbers closest to the incident mode wave number. This is a

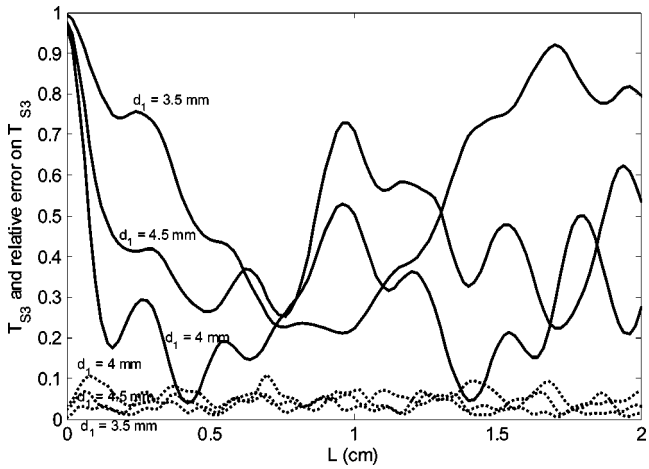


FIG. 4. Dependence of the transmission coefficient of the S3 mode (full lines) and relative error δ (dotted lines) versus the delamination length L for the case of an incident S3 mode at acoustic wave frequency $f=4$ MHz and plate thickness $d=5$ mm.

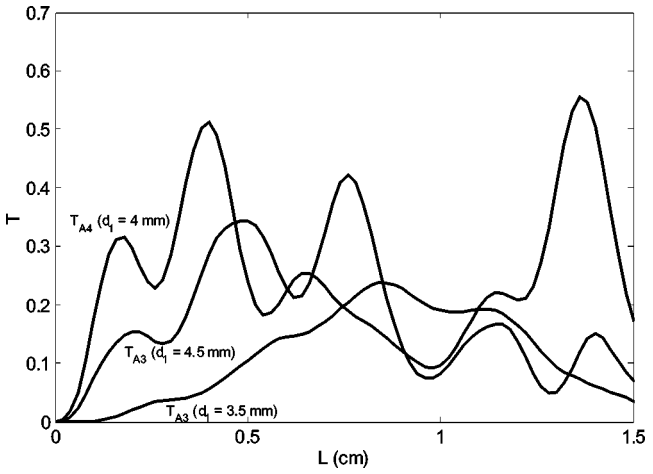


FIG. 5. Dependence of the transmission coefficient of different excited Lamb modes versus the delamination length L for the case of incident mode S3. The calculation parameters are the same as in Fig. 4.

result of a larger value of the overlap integrals $K_{n,m}^{\pm}$ and $I_{n,m}^{\pm}$ for modes with similar wave numbers, when the material properties of the media on both sides of the plane $z=0$ are equal. In this case the proper components of stress tensor and mechanical displacement vector for modes n, m on the both sides of the plane $z=0$ with close wave numbers q_n and q_m have similar dependencies on the y coordinate, leading to larger values of overlap integrals. As an example, the relative values of the acoustic flow of energy $p_{z,M^{\pm}}=P_{z,M^{\pm}}/P_{z,S3}$ for the strongest modes $M^{\pm}$ excited at $z=0$ in the upper $(+)$ and lower $(-)$ parts of the delamination region are given by $p_{z,S2^-}=0.694$, $p_{z,S3^-}=0.0028$, $p_{z,S4^-}=0.0039$, $p_{z,A2^-}=0.0032$, $p_{z,A3^-}=0.01$, $p_{z,S0^+}=0.014$, $p_{z,A0^+}=0.011$, $p_{z,A1^+}=0.24$, and $p_{z,A2^+}=0.0054$ for $d=5$ mm and $d_1=3.5$ mm. It is seen that the incident S3 mode is mainly converted into the $S2^-$ and $A1^+$ modes. In this case the wave numbers of $S2^-$ and $A1^+$ modes turn out to be indeed the closest to the wave number of the incident S3 mode: $q_{S3}\approx 0.803q_t$, $q_{S2,-}\approx 0.794q_t$, and $q_{A1,+}\approx 0.844q_t$.

The modes $S2^-$ and $A1^+$ are excited in phase by the incident S3 mode at $z=0$, but after propagating over the so-called beating length $l_b^{S2^-,A1^+}=\pi/(q_{A1,+}-q_{S2,-})$ these modes become out of phase. Therefore, the interference between $S2^-$ and $A1^+$ modes becomes destructive at $z=l_b^{S2^-,A1^+}$ and leads to a considerable minimum in the dependence of S3 mode transmission coefficient $T_{S3}(z)$. The interference becomes again constructive at $z=2l_b^{S2^-,A1^+}$ and leads to a considerable maximum in the dependence $T_{S3}(z)$. Summarizing, the minimum and maximum of T_{S3} are realized near $L=Nl_b^{S2^-,A1^+}$. The value of $l_b^{S2^-,A1^+}$ is about 0.87 cm for $d_1=3.5$ mm. The minimum and maximum of the S3 mode transmission coefficient are indeed realized near $z=(0.87,\ 1.74)$ cm.

In the same way oscillations of T_{S3} are explained for another delamination depth value. For example, the incident S3 mode is largely converted into the modes $S2^-$ ($q_{S2,-}\approx 0.853q_t$) and $A3^-$ ($q_{A3,-}\approx 0.767q_t$) for $d_1=4$ mm (48.6% of incident energy is converted into the $A3^-$ mode and 17.6% into the $S2^-$ mode). In this case the beating length $l_b^{S2^-,A3^-}=\pi/(q_{S2,-}-q_{A3,-})\approx 0.5$ cm and the minimum and maximum of T_{S3} are realized for $L\approx Nl_b^{S2^-,A3^-}$. In this case other propagating modes are also excited in the region $0<z<L$. This leads to considerable fluctuations of T_{S3}. Therefore the minimum and maximum of T_{S3} are shifted from exact $Nl_b^{S2^-,A3^-}$ values (the same situation is also realized for $d_1=4.5$ mm). The relative error for all these cases is relatively small.

Amplitudes of generated transmitted Lamb modes can become of considerable value for the modes with wave numbers close to the incident mode wave number. This is again explained by the equality of material properties of media in all parts of the structure under consideration when overlap integrals have the largest values for modes with close wave numbers. For $d=5$ mm the modes with the wave numbers closest to the wave number of S3 mode turn out to be A3 ($q_{A3}\approx 0.865q_t$) and A4 ($q_{A4}\approx 0.726q_t$) and calculations show that namely these excited modes are the strongest ones. The dependence of the transmission coefficients $T_{A3,A4}$ for these modes on the delamination length L for different delamination depths is illustrated in Fig. 5. The maxima (minima) of the graphs in Fig. 5 are closely related to the minima (maxima) of the corresponding graphs in Fig. 4. Strong oscillations of $T_{A3,A4}$ values for $d_1=(4,4.5)$ mm are again connected with considerable excitation of several modes (more than 2) in the region $0<z<L$.

Similar dependencies can be plotted versus the delamination depth $d-d_1$ at a fixed value of the delamination length L. An example is given in Fig. 6, where the dependence of the transmission coefficient of the S3 mode on a delamination depth is plotted for the case of an incident S3 mode, for different values of L. The nature of strong oscillations of transmission coefficient is the same as in Figs. 4 and 5. The incident S3 mode is converted to a set of modes at the $z=0$ boundary. These modes propagate with different phase velocities until the $z=L$ boundary. There, each of the modes is converted again to a set of modes which are transmitted in the $z>L$ region. The S3 mode in the $z>L$ region does con-

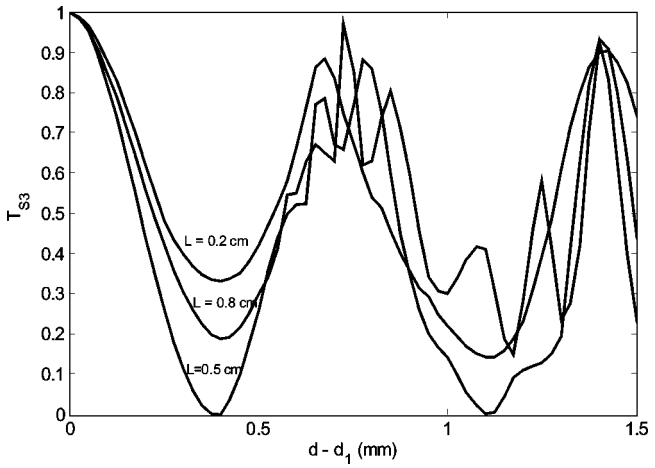


FIG. 6. Dependence of the transmission coefficient of the S3 mode versus the delamination depth $d-d_1$ at acoustic wave frequency $f=4$ MHz and plate thickness $d=5$ mm.

tains a mixture of interfering converted $0<z<L$ modes, whose interference critically depends on their respective velocities and on the propagation path L.

If the time is reversed, and all transmitted Lamb modes normally generated by an nth Lamb mode on a delamination region are incident on the same delamination, then the transmitted acoustic field will be dominated by this nth Lamb mode. A mixing condition for an optimal conversion of a set of incident Lamb modes into one transmitted nth Lamb mode can be written down as follows:

$$C_p^{in}=(C_{p,n}^{tr})^*, \tag{19}$$

where the amplitudes $C_{P,n}^{tr}$ are given by Eq. (18). This condition is equivalent with the well-known effect of phase conjugation in acoustics.[18]

To illustrate this effect the dependence of the transmission coefficient for a generated S3 mode on the delamination length, for different incident Lamb mode combinations and for different delamination depths, is shown in Fig. 7. The used incident mode combination is taken as indicated by

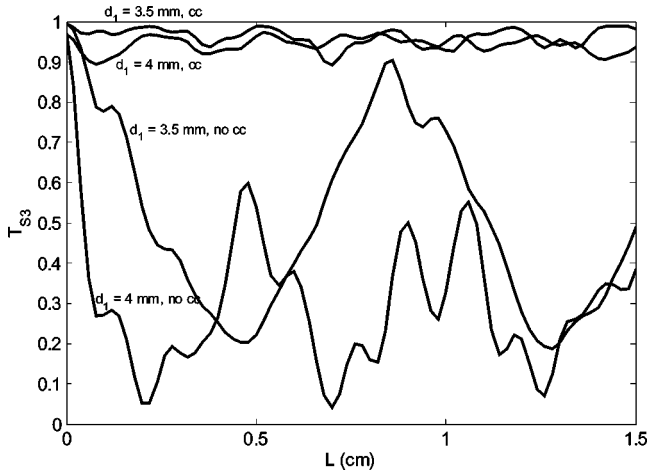


FIG. 7. Dependence of the transmission coefficient of an S3 mode versus the delamination length L for different (complex conjugated and not complex conjugated) incident mode combinations. The calculation parameters are the same as in Fig. 4.

Eqs. (18) and (19), with $C_n^{in}=C_{S3}^{in}$. Two graphs of Fig. 7 have been plotted taking $C_p^{in}=(C_{p,S3}^{tr})^*$. Two more graphs have been plotted taking $C_p^{in}=C_{p,S3}^{tr}$, i.e., without complex conjugation of the mode amplitudes, excited by the S3 mode. Complex conjugation of mode amplitudes, i.e., a reversal of the mode propagation direction, turns out to be effectively necessary in order to convert all incident modes into one S3 mode. The strong conversion of the mode combination into one mode is explained by the effect of phase conjugation, where the initial acoustic field distribution represented for this case by the incident S3 mode is restored. Small oscillations of the S3 mode amplitude in Fig. 7 are explained by calculation errors. These errors are reflected in the calculation of amplitudes of modes for the incident mode combination and also in the calculation of the S3 mode transmission coefficient for this incident mode combination. Total calculation errors are not shown in Fig. 7, but they are close to the values given in Fig. 4. The accuracy of S3 mode restoration by the incident mode combination with mode amplitudes defined by Eqs. (18) and (19) confirms that calculation results for transmitted propagating modes are of satisfactory accuracy when contribution of nonpropagating modes is neglected, if δ is small.

B. Mode conversions between Rayleigh wave and Lamb modes

A Rayleigh wave propagating on a thick plate is not dispersive. This feature makes the Rayleigh wave very interesting for experimental studies in multimode plates. For the ultimate case of a very thick plate, when $q_{S0}=q_{A0}$, the mechanical displacement vector of an incident Rayleigh wave propagating along the plate surfaces $y=0$ and $y=d$ is given by

$$\vec{U}_R^{in}=C_R^{in}(\vec{u}_{tr}^{S0}\pm\vec{u}_{tr}^{A0})Exp^{i(q_Rz-\omega t)}, \tag{20}$$

where $q_{S0}=q_{A0}=q_R$ and "+" is for Rayleigh wave propagating along surface $y=d$.

We consider a delamination situated close to the plate surface $y=d$, and a Rayleigh wave propagating along that surface. In this case the transmission coefficient of the generated Rayleigh wave T_R is defined as follows (here $F_{A0}\cong F_{S0}$):

$$T_R=2|C_R^{tr}|^2F_{S0}/\sum_m |C_m^{in}|^2F_m. \tag{21}$$

The transmission coefficient of the incident Rayleigh wave is equal to $T_R=|C_R^{tr}|^2/|C_R^{in}|^2$.

Typical results are shown in Fig. 8, where T_R and δ are plotted versus the delamination length L, for different values of the delamination depth inside a plate. T_R undergoes considerable oscillations, especially for $d_1=3.75$ mm and $d_1=4$ mm. For these two cases the incident Rayleigh wave energy is mainly concentrated in the region $d_1<y<d$ and almost all Rayleigh wave energy is converted at $z=0$ into $S0^+$ and $A0^+$ modes in the upper part of the delamination region with the maximum energy near the plate surface $y=d$. The field distribution of the mode combination $S0^+ + A0^+$ becomes inverted when z is close to the beating

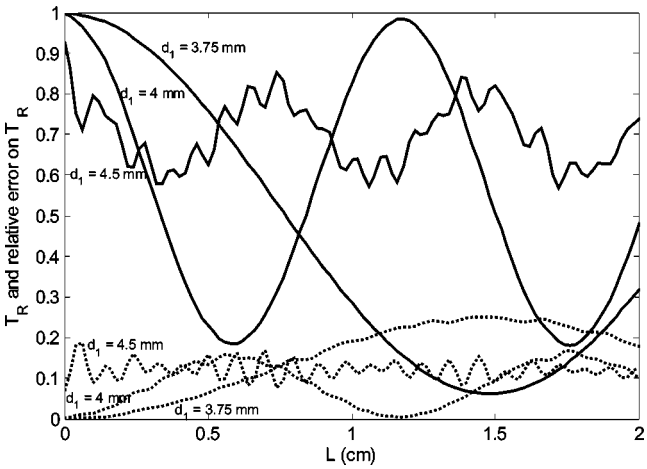


FIG. 8. Dependence of the transmission coefficient of a Rayleigh wave (full lines) and relative error δ (dotted lines) versus the delamination length L for acoustic wave frequency $f=4$ MHz and plate thickness $d=5$ mm.

length between the $S0^+$ and $A0^+$ modes, $l_b^{S0^+,A0^+}=\pi/(q_{A0,+}-q_{S0,+})$. This mode combination represents a wave propagating along the delamination interface $y=d_1$ with the maximum energy concentrated near the delamination. As a result, overlap between this field distribution and the field distribution of a Rayleigh wave propagating along the plate surface $y=d$ becomes minimum at $z=l_b^{S0^+,A0^+}$. The beating length between $S0^+$ and $A0^+$ modes is about 1.48 cm for $d_1=3.75$ mm and 0.586 cm for $d_1=4$ mm. It is seen that the minimum of T_R indeed occurs exactly at $L=l_b^{S0^+,A0^+}$ for $d_1=(3.75,4)$ mm. The situation becomes more complicated for $d_1=4.5$ mm, when comparable parts of the incident Rayleigh wave energy are distributed in the plate above and below the delamination, and a Rayleigh wave propagating along the lower side of the delamination interface, R^-, is also excited [R^- is defined analogically to Eq. (20): $\vec{U}_R^-=C_R^-(\vec{u}_{tr}^{S0^-}+\vec{u}_{tr}^{A0^-})Exp^{i(q_R-z-\omega t)}$ where $S0^-$ and $A0^-$ are the lowest Lamb modes of the lower part of the delamination region]. In this case about 84% and 6% of the incident Rayleigh wave is converted into the $A0^+$ mode and R^- Rayleigh wave, respectively. As a result the effect of beating between the $A0^+$ mode and R^- Rayleigh wave leads to oscillations in T_R with a period close to $2l_b^{A0^+,R^-}\approx 0.72$ cm. However, the considerable difference between the transmitted energies of the $A0^+$ and R^- Rayleigh wave modes results in smaller oscillations then in the previous cases. Ripples on the dependence $T_R(L)$ reveal the existence of another, weak mode. For the case of an incident Rayleigh wave δ is considerably larger than for the incident Lamb mode (compare Figs. 4 and 8). The error is especially large near the pronounced minimum of the transmission coefficient T_R [see graphs for δ at $d_1=(3.75,4)$ mm]. The error increase near such a minimum is connected with a strong concentration of acoustic wave energy near the delamination. In this case the contribution of neglected reflected and nonpropagating modes becomes more important, rendering the approximation in Eq. (11) less reliable. The δ value reaches here about 16% for $d_1=4$ mm and 25% for $d_1=3.75$ mm.

Near the pronounced minimum of the Rayleigh wave

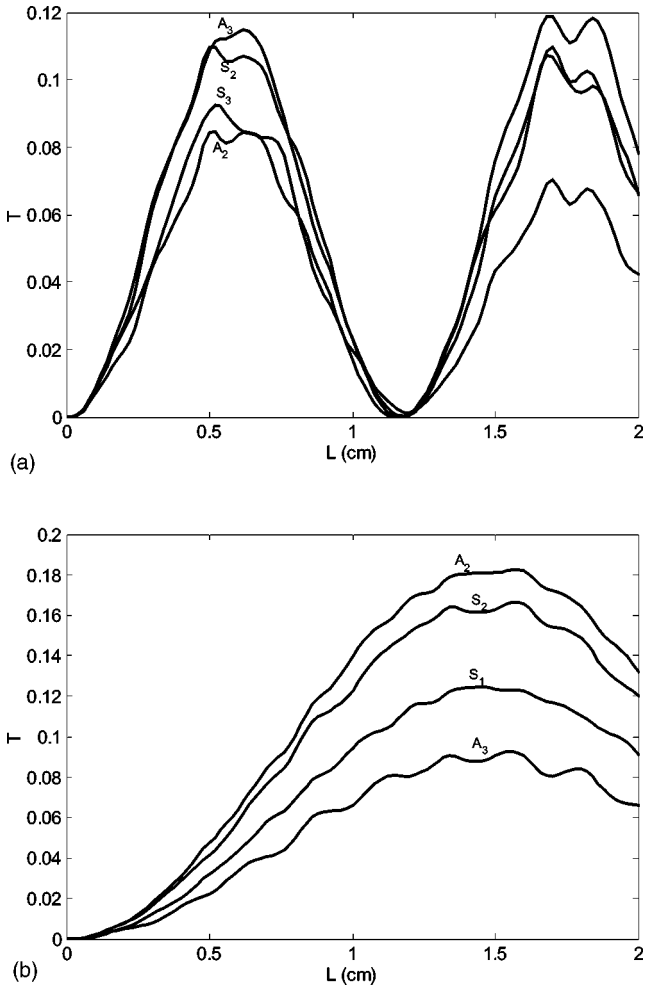


FIG. 9. (a) Dependence of the transmission coefficient of excited Lamb modes versus the delamination length L for the case of an incident Rayleigh wave for $d_1=4$ mm at acoustic wave frequency $f=4$ MHz and plate thickness $d=5$ mm. Red—A3 mode, blue—S2 mode, green—S3 mode, and black—A2 mode. (b) Dependence of the transmission coefficient of excited Lamb modes versus the delamination length L for the case of an incident Rayleigh wave for $d_1=3.75$ mm. The calculation parameters are the same as in (a).

transmission coefficient the incident Rayleigh wave is mainly converted into low-order Lamb modes. The dependences of the transmission coefficients $T_{A2,A3,S2,S3}$ and $T_{A2,S2,S1,A3}$ for the strongest modes on the delamination length L are given in Fig. 9(a) for $d_1=4$ mm and in Fig. 9(b) for $d_1=3.75$ mm. The maxima (minima) in Figs. 9(a) and (b) are analogous to the minima (maxima) of appropriate graphs in Fig. 8. But all transmission coefficients are comparatively small. This is connected with a strong localization of Rayleigh wave energy inside the plate. In that case the overlap integrals between the Rayleigh wave and several Lamb modes become comparable. Compared with results obtained for the case of the incident Lamb mode, here a wider spectrum of transmitted Lamb modes with relatively small amplitudes is excited.

The same kind of dependence can be plotted versus delamination depth $d-d_1$ at a fixed value of delamination length L. In Fig. 10, the dependence of T_R on $d-d_1$ is plotted for the case of incident Rayleigh wave at $L=0.2$ cm. Again the strong oscillations of the T_R value are explained

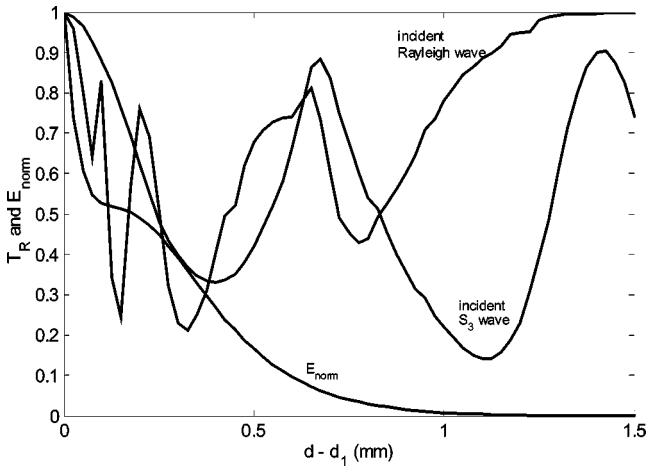


FIG. 10. Dependence of the transmission coefficient of a Rayleigh wave versus the delamination depth $d-d_1$ for the case of the incident Rayleigh wave for $L=0.2$ cm at acoustic wave frequency $f=4$ MHz and plate thickness $d=5$ mm. For comparison the dependence for an incident S3 mode is shown. Also the depth dependence of the normalized Rayleigh wave energy is shown.

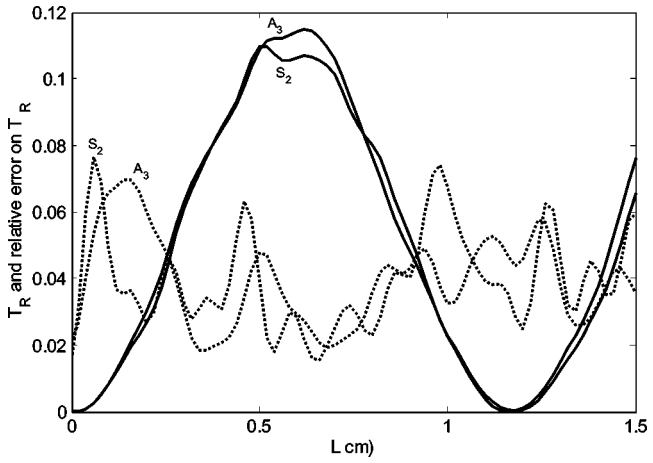


FIG. 11. Dependence of the transmission coefficient T_R (full lines) of a Rayleigh wave and relative error δ (dotted lines) versus the delamination length L for the case of one incident Lamb mode at $d_1=4$ mm. The calculation parameters are the same as in Fig. 8.

by an interference between contributions of different modes excited in the region $0<z<L$. For the sake of comparison in Fig. 10 the analogous dependence of the S3 mode transmission coefficient T_{S3} for the case of an incident mode S3 and the normalized dependence of Rayleigh wave energy density at the delamination position are also shown. The Rayleigh wave transmission coefficient is much more sensitive to the delamination position near the plate surface $y=d$ than the transmission coefficient of an incident mode S3. This is connected with the stronger localization of Rayleigh wave energy near the plate surface $y=d$. The interaction of the Rayleigh wave with the delamination becomes weak for $y<d-2.2\lambda_R$; in this case the transmission coefficient of Rayleigh wave $T_R \geqslant 0.95$ for $L=1$ cm.

The results indicate that investigation of Rayleigh wave propagation along a coating layer can be used to estimate the existence of a delamination in the layer between a coating and a substrate. If the Rayleigh wave frequency is chosen such that almost all Rayleigh wave energy is concentrated inside the coating and the density of Rayleigh wave energy distribution tail touches the substrate, then the presence of a delamination leads to a considerable propagation disturbance.

We now investigate the possibilities for efficient Rayleigh wave excitation from incident Lamb modes. First we consider the case of one incident Lamb mode. The largest amplitude of the generated Rayleigh wave is expected for a low order incident Lamb mode. But even in this case the transmission coefficient of an excited Rayleigh wave does not exceed 15%. In many cases the T_R value lies below the uncertainty level. The dependence of T_R and δ on the delamination length L are shown in Fig. 11 for the most efficient cases of Rayleigh wave excitation from incident S2 and A3 modes at $d_1=4$ mm. T_R values do not exceed 12% and are only considerably larger than the error level near the regions of maximum Rayleigh wave amplitude. The same kind of dependence can be plotted for incident S3 and A2 modes. For these cases T_R values are smaller and calculation errors are larger—up to 10%–12%. This is in the accordance with the result of Lamb mode excitation by an incident Rayleigh wave as presented in Fig. 9. In that case the strongest excited modes are A3 and S2. It is not surprising that namely these modes can realize the strongest conversion into transmitted Rayleigh wave when incident on the same delamination.

To further increase the conversion efficiency of incident Lamb modes into a Rayleigh wave it is necessary to consider a combination of incident Lamb modes. To find such Lamb mode combinations one can use the method described in Sec. III A of the paper. Amplitudes of incident Lamb modes are chosen as $C_p^{in}=(C_{p,R}^{tr})^*$. $C_{p,R}^{tr}$ values are defined as

$$C_{p,R}^{tr}=C_R^{in}\sum_{m,\pm} e^{iq_{m,\pm}L} \times\frac{(K_{p,m}^{\pm}+I_{p,m}^{\pm})(K_{S0,m}^{\pm}+I_{S0,m}^{\pm}+K_{A0,m}^{\pm}+I_{A0,m}^{\pm})}{A(p)A^{\pm}(m)}. \tag{22}$$

$C_{p,R}^{tr}$ are the Lamb modes amplitudes generated on a delamination by the incident Rayleigh wave with amplitude C_R^{in}. The transmitted Rayleigh wave is also contained among all modes generated by the incident Rayleigh wave. If such a full combination is incident on the delamination, then the transmitted wave consists mainly of a transmitted Rayleigh wave, and the acoustic distribution given by the initial Rayleigh wave is restored. If the transmitted Rayleigh wave is excluded from such Lamb modes combinations, the conversion efficiency of Rayleigh wave generation decreases, but can still be very large. To illustrate this effect, the dependence of the transmission coefficient T_R of the Rayleigh wave on the delamination length generated by the incident Lamb mode full combinations, and by reduced combinations without contribution of A0 and S0 modes, is shown in Fig. 12 for different delamination depths. The conversion for the case of the full combination is very large, but even the reduced combination leads to a conversion of about 70% of incident Lamb mode energy into transmitted Rayleigh wave

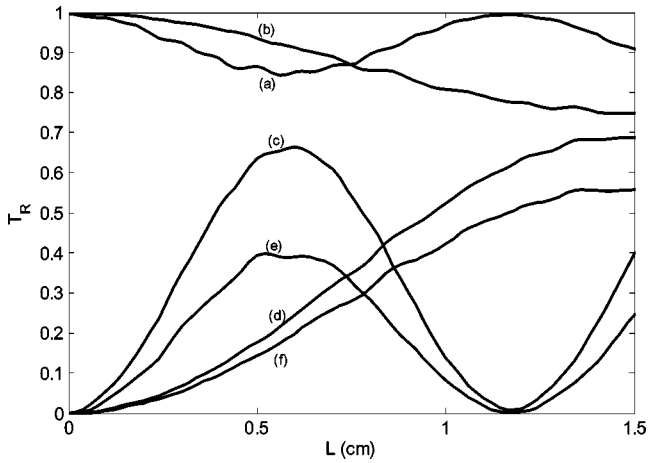


FIG. 12. Dependence of the transmission coefficient T_R of a Rayleigh wave versus the delamination length L for the case of incident mode combinations at acoustic wave frequency $f=4$ MHz and plate thickness $d=5$ mm. (a) Full incident combination at $d_1=4$ mm, (b) full incident combination at $d_1=3.75$ mm, (c) incident combination without contribution of modes A0 and S0 at $d_1=4$ mm, (d) incident combination without contribution of modes A0 and S0 at $d_1=3.75$ mm, (e) incident combination consisting of modes A2, A3, S2, and S3 at $d_1=4$ mm, and (f) incident combination consisting of modes A2, A3, S1, and S2 at $d_1=3.75$ mm.

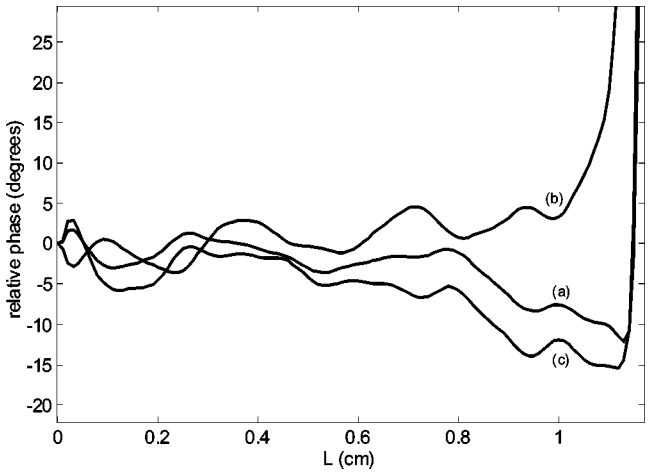


FIG. 13. Dependence of the phases of S2 mode (a), S3 mode (b) and A2 mode (c) versus the delamination length L for $d_1=4$ mm. The phases are normalized with the phase of the A3 mode. The calculation parameters are the same as in Fig. 12.

energy. The maximum of conversion energy is realized at the value of the delamination length, for which the transmission coefficient T_R of the incident Rayleigh wave case is minimum (compare Figs. 8 and 12). The relative error near maximum conversion is about 20%—close to the δ value shown in Fig. 8 near the minimum of Rayleigh wave transmission coefficient. This error leads to inaccurate modes amplitudes for the incident full mode combination and thus to an inexact restoration of the initial Rayleigh wave acoustic field distribution. In Fig. 12 the T_R values for full combinations deviate from 1, the expected value for the case of full restoration. The deviation is about 15% for $d_1=4$ mm and 25% for $d_1=3.75$ mm. This again confirms that accuracy of calculation results for transmitted propagating modes (when contribution of nonpropagating modes is neglected) is defined by the δ value and this accuracy is satisfactory if δ is small.

For the calculation of most of the curves in Fig. 12 all Lamb modes given by Eq. (22) are taken into account. Even for the calculation with the reduced combination (without A0 and S0) there are still 18 propagating Lamb modes for this plate at $f=4$ MHz. One could ask whether the highest modes can be removed from these combinations without large loss of conversion efficiency. The greatest contributions to the reduced combinations are given by modes A3, S2, S3, and A2 for $d_1=4$ mm and by modes A2, S2, S1, and A3 for $d_1=3.75$ mm [see Figs. 9(a), (b)]. The dependence of T_R on L for incident Lamb mode combinations consisting of only these four Lamb modes [with amplitudes defined by Eq. (21)] are shown for comparison in Fig. 12. Still 60% for $d_1=4$ mm and 80% for $d_1=3.75$ mm of the excited Rayleigh wave energy by the reduced mode combinations are due to these four Lamb modes. The L dependence of the phases of the S2, S3, and A2 modes normalized to the phase of the A3 mode are shown in Fig. 13 for $d_1=4$ mm. The phase difference between all the greatest generated Lamb modes is quite small. Neglecting this phase difference results in almost the same amplitude of generated Rayleigh wave by the combination of these four Lamb modes. Analogous results are obtained for modes A2, S2, S1, and A3 at $d_1=3.75$ mm. The results reveal that a considerable part of incident Lamb mode energy can be converted into a Rayleigh wave, if amplitudes of incident Lamb modes are correctly selected.

IV. CONCLUSIONS

In this paper the two-dimensional propagation of Lamb modes through a region containing a finite horizontal delamination was studied. The calculation is based on the modal decomposition method. An infinite system of linear equations for the amplitudes of transmitted and reflected propagating and nonpropagating Lamb modes was derived using boundary conditions and an orthogonality condition between Lamb modes with different wave numbers. The use of orthogonality relations allows us to obtain an equal amount of equations as unknowns, which distinguishes this approach from modal decomposition methods using singular value decomposition. The amplitudes of transmitted propagating Lamb modes are calculated analytically for the case of a multimode plate neglecting contributions of nonpropagating and propagating reflected modes. The error due to the made approximations can be quantified using a measure for unconserved energy, δ.

As a result of diffraction on a delamination, an incident Lamb mode undergoes conversion into all possible transmitted Lamb modes, but this conversion is efficient only if the wave numbers of transmitted and incident Lamb modes are close to each other. Using the method of phase conjugation for calculating a correct mix, incident Lamb mode combination can be efficiently converted into a single transmitted Lamb mode. In this case all incident mode amplitudes have to be chosen to be complex conjugated to Lamb mode amplitudes generated by the nth Lamb mode on the same delamination. Calculation errors for this case usually stay below 10%.

The transmission coefficient of a Rayleigh wave incident on an area with a delamination located near the plate surface is oscillating. The Rayleigh wave is strongly converted into transmitted Lamb modes (mainly into several low order symmetric and antisymmetric Lamb modes) near a pronounced minimum of Rayleigh wave transmission coefficient. Oscillations of the Rayleigh wave transmission coefficient are explained by beating between Lamb modes in the region of plate with the delamination. Lamb mode combinations incident on the delamination region can be efficiently converted into a transmitted Rayleigh wave using the method of phase conjugation. If these combinations consist of, for example, four of the most important Lamb modes, then 50% or more of the energy can be converted into Rayleigh wave energy. For a single Lamb mode incident on a delamination, the generation of transmitted Rayleigh wave is much less efficient. Calculation errors for this case are larger (about 15%–20%). This is related with the strong localization of Rayleigh wave energy near the delamination. The Rayleigh wave is very sensitive to delaminations located near the plate surface and can be used for detecting such delaminations.

ACKNOWLEDGMENTS

The authors would like to thank the Research Council of the K. U. Leuven for the fellowship awarded to GS, which enabled us to perform this research.

APPENDIX

1. Orthogonality Condition

To derive orthogonality condition (2a) we define a vector $P_k^{(n,m)}$ as

$$P_k^{(n,m)} = \sum_{j=y,z} (S_{jk}^n U_j^m - U_j^n S_{jk}^m). \tag{A1}$$

The components of the stress tensor S_{jk}^n and the mechanical displacement vector U_j^n of the nth Lamb mode are defined by Eq. (1). The frequencies of both nth and mth Lamb modes are assumed to be equal. Taking into account stress–strain relations and the equation motion we find

$$\operatorname{div} \vec{P}^{(n,m)} = 0. \tag{A2}$$

Integrating over a cross section of the plate and combining with stress-free boundary conditions on the plate surfaces $y=0$, d one obtains

$$\int_0^d P_z^{(n,m)}\, dy = \text{const}. \tag{A3}$$

Taking into account that $P_z^{(n,m)} \approx \operatorname{Exp}(i(q_n+q_m)z)$, one finds for $q_n \neq -q_m$

$$\int_0^d (\sigma_{yz}^n u_y^m + \sigma_{zz}^n u_z^m - \sigma_{zz}^m u_z^n - \sigma_{yz}^m u_y^n) dy = 0. \tag{A4}$$

The orthogonality condition Eq. (A4) is not so convenient, since modes are orthogonal to themselves. In order to circumvent this, we do the following manipulation. A similar relation for Lamb modes n and m^- with $q_{m^-} = -q_m$ and $q_n \neq q_m$ is given by Eq. (A5):

$$\int_0^d (\sigma_{yz}^n u_y^m - \sigma_{zz}^n u_z^m - \sigma_{zz}^m u_z^n + \sigma_{yz}^m u_y^n) dy = 0. \tag{A5}$$

Adding Eqs. (A4) and (A5) results in Eq. (A6) valid for $q_n \neq \pm q_m$:

$$\int_0^d (\sigma_{yz}^n u_y^m - \sigma_{zz}^m u_z^n) dy = 0. \tag{A6}$$

Because not all unequal modes are orthogonal with respect to (A6), we perform a further manipulation: Finally, Eq. (2a) is obtained by adding Eq. (A6) with the analogous equation where the substitutions $m \leftrightarrow n$ are made. Equation (2a) is now fulfilled for all modes with $q_n \neq q_m$. According to Eqs. (A6) and (2a), forward and backward propagating Lamb modes m and m^- are orthogonal.

2. Lamb Modes

The displacement fields for the symmetric forward propagating nth Lamb mode can be written down as follows:

$$u_{y,tr}^{n,s} = \frac{q_{ly}}{q_t^2} \tan(q_{ly}\tilde{d}) \left(\left(2\frac{q_{n,s}^2}{q_t^2} - 1 \right) \frac{\sin(q_{ly}(y-\tilde{d}))}{\sin(q_{ly}\tilde{d})} - 2\frac{q_{n,s}^2}{q_t^2} \frac{\sin(q_{ty}(y-\tilde{d}))}{\sin(q_{ty}\tilde{d})} \right) e^{i(q_{n,s}z-\omega t)}, \tag{A7}$$

$$u_{z,tr}^{n,s} = i\frac{q_{n,s}}{q_t^2} \left(1 - 2\frac{q_{n,s}^2}{q_t^2} \right) \left(\frac{\cos(q_{ly}(y-\tilde{d}))}{\cos(q_{ly}\tilde{d})} + \left(\frac{q_t^2}{2q_{n,s}^2} - 1 \right) \frac{\cos(q_{ty}(y-\tilde{d}))}{\cos(q_{ty}\tilde{d})} \right) e^{i(q_{n,s}z-\omega t)}, \tag{A8}$$

where $q_{ly}^2 = q_l^2 - q_{n,s}^2$, $q_{ty}^2 = q_t^2 - q_{n,s}^2$, q_l is the wave number of longitudinal waves, and $\tilde{d} = d/2$.

Wave numbers $q_{n,s}$ of symmetric Lamb modes satisfy the characteristic equation (A9):

$$\frac{(q_t^2 - 2q_{n,s}^2)^2}{4q_{ty}q_{ly}q_{n,s}^2} + \frac{\tan(q_{ly}\tilde{d})}{\tan(q_{ty}\tilde{d})} = 0. \tag{A9}$$

The displacement fields for the antisymmetric forward propagating nth Lamb mode can be written down as follows:

$$u_{y,tr}^{n,a} = \frac{q_{ly}}{q_t^2 \tan(q_{ly}\tilde{d})} \left(\left(1 - 2\frac{q_{n,a}^2}{q_t^2} \right) \frac{\cos(q_{ly}(y-\tilde{d}))}{\cos(q_{ly}\tilde{d})} + 2\frac{q_{n,a}^2}{q_t^2} \frac{\cos(q_{ty}(y-\tilde{d}))}{\cos(q_{ty}\tilde{d})} \right) e^{i(q_{n,a}z-\omega t)}, \tag{A10}$$

$$u_{z,tr}^{n,a} = i\frac{q_{n,a}}{q_t^2} \left(1 - 2\frac{q_{n,a}^2}{q_t^2} \right) \left(\frac{\sin(q_{ly}(y-\tilde{d}))}{\sin(q_{ly}\tilde{d})} + \left(\frac{q_t^2}{2q_{n,a}^2} - 1 \right) \frac{\sin(q_{ty}(y-\tilde{d}))}{\sin(q_{ty}\tilde{d})} \right) e^{i(q_{n,a}z-\omega t)}. \tag{A11}$$

Wave numbers $q_{n,a}$ of antisymmetric Lamb modes satisfy the characteristic equation (A12):

$$\frac{(q_t^2 - 2q_{n,a}^2)^2}{4q_{ty}q_{ly}q_{n,a}^2} + \frac{\tan(q_{ty}\tilde{d})}{\tan(q_{ly}\tilde{d})} = 0. \tag{A12}$$

The displacement fields for symmetric and antisymmetric Lamb modes and appropriate characteristic equations in the plate region $0<z<L$, $d_1 \leq y<d$ are given by performing the substitutions $y=y-d_1$, $\tilde{d}=(d-d_1)/2$ in Eqs. (A7)–(A12).

[1] B. A. Auld and M. Tan, ''Symmetrical Lamb wave scattering at a symmetrical pair of thin slots,'' Ultrasonic Symp. Proc. (1977), pp. 61–66.
[2] M. Tan and B. A. Auld, ''Normal mode variational method for two and three dimensional acoustic scattering in an isotropic plate,'' Ultrasonic Symp. Proc. (1980), pp. 857–861.
[3] S. Rokhlin, ''Interaction of Lamb waves with elongated delaminations in thin sheets,'' Int. Adv. Nondestr. Test. **6**, 263–285 (1979).
[4] S. Rokhlin, ''Diffraction of Lamb waves by a finite crack in an elastic layer,'' J. Acoust. Soc. Am. **67**, 1157–1165 (1980).
[5] S. Rokhlin, ''Resonant phenomena of Lamb waves scattering by a finite crack in a solid layer,'' J. Acoust. Soc. Am. **69**, 922–928 (1981).
[6] L. Wang and J. Shen, ''Scattering of elastic waves by a crack in an isotropic plate,'' Ultrasonics **356**, 451–457 (1997).
[7] M. Castaings, E. Le Clezio, and B. Hosten, ''Modal decomposition method for modeling the interaction of Lamb waves with cracks,'' J. Acoust. Soc. Am. **112**, 2567–2582 (2002).
[8] M. J. Love, P. Cawley, J.-Y. Kao, and O. Diligent, ''The low frequency reflection characteristics of the fundamental antisymmetric Lamb wave a_0 from a rectangular notch in a plate,'' J. Acoust. Soc. Am. **112**, 2612–2622 (2002).
[9] N. Guo and P. Cawley, ''The interaction of Lamb waves with delaminations in composite laminates,'' J. Acoust. Soc. Am. **94**, 2240 (1993).
[10] M. J. S. Lowe, R. E. Challis, and C. W. Chan, ''The transmission of Lamb waves across adhesively bonded lap joints,'' J. Acoust. Soc. Am. **107**, 1333 (2000).
[11] E. Moulin, J. Assaad, C. Delebarre, and D. Osmont, ''Modeling of Lamb waves generated by integrated transducers in composite plates using a coupled finite element–normal modes expansion method,'' J. Acoust. Soc. Am. **107**, 87 (2000).
[12] O. M. Mukdadi and S. K. Datta, ''Transient ultrasonic guided waves in layered plates with rectangular cross section,'' J. Appl. Phys. **93**, 9360 (2003).
[13] O. M. Mukdadi, Y. M. Desai, S. K. Datta, A. H. Shah, and A. J. Niklasson, ''Elastic guided waves in a layered plate with rectangular cross section,'' J. Acoust. Soc. Am. **112**, 1766 (2002).
[14] J. Vandeputte, O. Leroy, R. Briers, and G. Shkerdin, ''Influence of planar cracks on reflected and transmitted fields of Gaussian acoustic beams,'' J. Acoust. Soc. Am. **114**, 634 (2003).
[15] B. A. Auld, *Acoustic Fields and Waves in Solids* (Wiley, New York, 1973).
[16] R. Briers, O. Leroy, G. Shkerdin, and Yu. Gulyaev, ''Mode theory as a framework for the investigation of the generation of a Stoneley wave at liquid/solid interfaces,'' J. Acoust. Soc. Am. **95**, 1953–1966 (1994).
[17] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1973).
[18] M. Fink, ''Time reversed acoustics,'' Phys. Today **20**, 34–40 (1997).

Leaky Rayleigh and Scholte waves at the fluid–solid interface subjected to transient point loading

Jinying Zhu and John S. Popovics[a)]
Department of Civil and Environmental Engineering, University of Illinois at Urbana, Illinois 61801

Frank Schubert
Lab EADQ, Fraunhofer Institute for Nondestructive Testing, D-01326 Dresden, Germany

(Received 6 February 2004; revised 9 July 2004; accepted 18 July 2004)

The analysis of acoustic waves generated by a transient normal point load applied on a fluid–solid interface is presented. The closed-form exact solution of the wave motion is obtained by using integral transform techniques. The obtained analytical solution provides necessary theoretical background for optimization of fluid-coupled ultrasonic and acoustic wave detection in experiments. Numerical simulation (elastodynamic finite integration technique) is performed to verify the obtained analytical solution. Detailed descriptions of leaky Rayleigh and Scholte wave solutions are presented. A simplified solution to isolate the contributions of leaky Rayleigh and Scholte waves generated by a transient point load is proposed, and closed-form formulations for displacement and stress components are then presented. The simplified solution is compared to the exact solution for two configurations: water/concrete and air/concrete. The excitation effectiveness of leaky Rayleigh waves for the air/concrete configuration is studied, which has practical significance to air-coupled sensing in civil engineering structures.
[DOI: 10.1121/1.1791718]

PACS numbers: 43.35.Pt, 43.20.El, 43.35.Zc [YHB] Pages: 2101–2110

I. INTRODUCTION

The propagating leaky Rayleigh wave that emanates from a fluid–solid interface has been used as an effective means for surface and subsurface defect detection. With recent improvements in instrumentation, air-coupled transducers have been used for detection of leaky surface (Zhu *et al.*, 2001) and leaky guided waves (Castaings *et al.*, 2001). With the advantage of noncontact sensing, air-coupled transducers provide an opportunity for quick scanning and imaging of large civil engineering structures by detecting the leaky Rayleigh wave. Therefore, detailed study of leaky surface waves for this case is needed.

Extensive studies and applications of leaky surface waves have been reported during the past 40 years. A comprehensive study of Rayleigh waves and leaky Rayleigh waves has been given by Viktorov (1967), where leaky Rayleigh waves at the interfaces of a solid half-space with both a fluid layer and a fluid half-space were investigated in great detail. Leaky Rayleigh waves propagate with a velocity slightly higher than the ordinary Rayleigh wave, and attenuate more intensively with distance due to continuous energy radiation into the fluid. It was initially believed that the leaky Rayleigh wave exists when the fluid wave velocity c_F is smaller than the leaky Rayleigh wave velocity c_{LR}. However, Mozhaev and Weihnacht (2002) showed the actual threshold phase velocity for leaky Rayleigh wave existence was 1.45% lower than the acoustic wave velocity of the fluid. The character and existence conditions of leaky Rayleigh and Scholte waves were also investigated experimentally by Glorieux *et al.* (2001). The propagation of leaky Rayleigh waves under the influences of viscous damping and heat conduction at the fluid–solid interface was studied by Qi (1994), who concluded that the effect of viscosity can be neglected for fluids with Reynolds number larger than 2500. For common fluids such as water and air at normal driving frequencies (<10 MHz), the viscous effect of the fluid can be neglected because the Reynolds numbers are far above the critical value of 2500.

The response of leaky waves owing to transient point loading is of great practical interest to nondestructive evaluation (NDE) researchers, especially in civil engineering, where an impulse hammer or a point impactor applied to the surface of the solid is often used as a transient wave source. Gusev *et al.* (1996) provided detailed theoretical analyses of laser-induced wave motions at the fluid–solid interface, which include Scholte, leaky Rayleigh, and lateral waves. General solutions for interface wave motion were given using a 2-D formulation. 2-D and 3-D analytical solutions for a fluid–solid configuration subjected to implosive line and point sources in the fluid have been given by de Hoop *et al.* (1983, 1984). However, for the case of the ''Lamb'' problem in fluid–solid configuration, where a normal transient point load is applied at the fluid–solid interface, no closed-form exact solution has been reported so far.

In this paper, Laplace and Hankel integral transforms are employed to derive the full analytical solution to the ''Lamb'' problem in a fluid–solid half-space system, where a point load, varying with time as a step function, is applied at the interface. The obtained step response solution is shown in integral format, which can be calculated numerically. Impulse responses are then obtained by differentiating the step responses. Therefore, for any transient impact loading that

[a)]Author to whom correspondence should be addressed. Electronic mail: johnpop@uiuc.edu

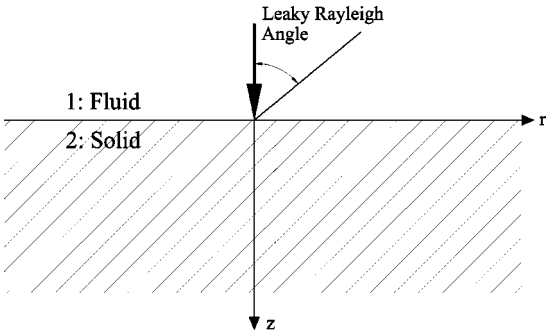


FIG. 1. A transient point load applied at the interface of a fluid–solid half-space system.

has arbitrary temporal variation and spatial distribution, the responses can be obtained by convolving the impulse response in time and space domains. The obtained analytical solutions are verified by EFIT (elastodynamic finite integration technique) numerical simulation, which is a powerful tool for elastodynamic wave field analysis. Then a simplified analytical formulation for the pressure field in the fluid is derived and illustrated.

II. COMPLETE FORMULATION

Consider a fluid–solid half-space system as shown in Fig. 1. The solid half-space is given by $z>0$, and the fluid by $z<0$. The properties of the fluid are given by the Lamé constant λ_1 and mass density ρ_1, and those of the solid by the Lamé constants λ_2 and μ and density ρ_2. The interface between the fluid and solid half-space is subjected to a normal point load of magnitude $QH(t)$, where $H(t)$ is the Heaviside step function. Because the wave motion in the fluid and solid generated by the point load is axially symmetric about the z axis, cylindrical coordinates are employed, where the origin is located at the load point on the interface.

A. Fluid–solid half-space

Introducing displacement potential functions φ_1 in the fluid and φ_2 and ψ in the solid, the governing equations for the fluid and the solid half-spaces are

$$\begin{aligned}
&\frac{\partial^2\varphi_1}{\partial r^2}+\frac{1}{r}\frac{\partial\varphi_1}{\partial r}+\frac{\partial^2\varphi_1}{\partial z^2}=\frac{1}{c_F^2}\ddot{\varphi}_1,\\
&\frac{\partial^2\varphi_2}{\partial r^2}+\frac{1}{r}\frac{\partial\varphi_2}{\partial r}+\frac{\partial^2\varphi_2}{\partial z^2}=\frac{1}{c_P^2}\ddot{\varphi}_2,\\
&\frac{\partial^2\psi}{\partial r^2}+\frac{1}{r}\frac{\partial\psi}{\partial r}+\frac{\partial^2\psi}{\partial z^2}-\frac{\psi}{r^2}=\frac{1}{c_S^2}\ddot{\psi},
\end{aligned} \tag{1}$$

where $c_F^2=\lambda_1/\rho_1$ is the acoustic wave velocity in the fluid, and $c_P^2=(\lambda_2+2\mu)/\rho_2$ and $c_S^2=\mu/\rho_2$ are the P- and S-wave velocities in the solid. The double dots represent a double derivative with respect to time. The displacements are related to potential functions by

$$u_1=\frac{\partial\varphi_1}{\partial r},\quad w_1=\frac{\partial\varphi_1}{\partial z},$$

$$u_2=\frac{\partial\varphi_2}{\partial r}-\frac{\partial\psi}{\partial z},\quad w_2=\frac{\partial\varphi_2}{\partial z}+\frac{1}{r}\frac{\partial(r\psi)}{\partial r}, \tag{2}$$

where the subscripts 1 and 2 represent quantities related to the fluid and solid, respectively.

B. Continuity and initial condition

A point load is applied at the origin as a Dirac delta function in space and varies as a Heaviside step function in time, which can be expressed as $QH(t)\delta(r)/2\pi r$. Because ideal fluid and shear-free interface conditions are assumed, only the normal stress and vertical component of the displacement are continuous at the interface. The continuity conditions at $z=0$ are

$$\begin{aligned}
&w_1=w_2,\\
&\tau_{zz2}=\tau_{zz1}+(-Q)H(t)\frac{\delta(r)}{2\pi r}=-P-QH(t)\frac{\delta(r)}{2\pi r},\\
&\tau_{zr2}=\tau_{zr1}=0,
\end{aligned} \tag{3}$$

where τ_{zz} and τ_{zr} are the normal and shear components of stress, and P the pressure in the fluid.

Assuming the system is at rest prior to $t=0$, we have

$$\begin{aligned}
\varphi_1(r,z,0)&=\ddot{\varphi}_1(r,z,0)\\
&=\varphi_2(r,z,0)=\ddot{\varphi}_2(r,z,0)\\
&=\psi_2(r,z,0)=\ddot{\psi}_2(r,z,0)=0.
\end{aligned} \tag{4}$$

C. Integral transform

One-sided Laplace and Hankel transforms are used to obtain solutions to the equations. The Laplace and nth-order Hankel transforms are defined respectively as

$$\overline{f}(p)=\int_0^\infty f(t)e^{-pt}\,dt,\quad f^{H_n}(\xi)=\int_0^\infty f(r)J_n(\xi r)r\,dr, \tag{5}$$

where p and ξ are variables of the Laplace and Hankel transforms, respectively.

Applying the Laplace transform to Eq. (1) with respect to time t and the zeroth- and first-order Hankel transform with respect to the radial distance r yields

$$\begin{aligned}
&\frac{d^2\overline{\varphi}_1^{H_0}}{dz^2}-\alpha_1^2\overline{\varphi}_1^{H_0}=0,\\
&\frac{d^2\overline{\varphi}_2^{H_0}}{dz^2}-\alpha_2^2\overline{\varphi}_2^{H_0}=0,\quad \frac{d^2\overline{\psi}^{H_1}}{dz^2}-\beta^2\overline{\psi}^{H_1}=0,
\end{aligned} \tag{6}$$

where

$$\alpha_1^2=\xi^2+s_F^2p^2,\quad \alpha_2^2=\xi^2+s_P^2p^2,\quad \beta^2=\xi^2+s_S^2p^2, \tag{7}$$

and s_F, s_P, s_S are P- and S-wave slowness in the fluid and the solid. Only choosing the terms which lead to finite responses for large values of $|z|$, we obtain the solutions to Eq. (6),

$$\bar{\varphi}_1^{H_0}=\Phi_1(\xi,p)e^{\alpha_1 z},$$

$$\bar{\varphi}_2^{H_0}=\Phi_2(\xi,p)e^{-\alpha_2 z},\quad \bar{\psi}^{H_1}=\Psi(\xi,p)e^{-\beta z}, \tag{8}$$

where $\Phi_1(\xi,p)$, $\Phi_2(\xi,p)$, and $\Psi(\xi,p)$ are functions of p and ξ that need to be determined. From Eq. (2) and the displacement–stress relationships, the transformed displacements and stresses (pressure for the fluid) in the fluid are

$$\bar{w}_1^{H_0}=\frac{d\bar{\varphi}_1^{H_0}}{dz},\quad \bar{u}_1^{H_1}=-\xi\bar{\varphi}_1^{H_0},\quad \bar{P}^{H_0}=-\rho_1 p^2\bar{\varphi}_1^{H_0}, \tag{9}$$

and in the solid are

$$\bar{w}_2^{H_0}=\frac{d\bar{\varphi}_2^{H_0}}{dz}+\xi\bar{\psi}^{H_1},\quad \bar{u}_2^{H_1}=-\frac{d\bar{\psi}^{H_1}}{dz}-\xi\bar{\varphi}_2^{H_0},$$

$$\bar{\tau}_{zz2}^{H_0}=\mu\left[(s_S^2p^2+2\xi^2)\bar{\varphi}_2^{H_0}+2\xi\frac{d\bar{\psi}^{H_1}}{dz}\right], \tag{10}$$

$$\bar{\tau}_{zr2}^{H_1}=-\mu\left[2\xi\frac{d\bar{\varphi}_2^{H_0}}{dz}+(s_S^2p^2+2\xi^2)\bar{\psi}^{H_1}\right].$$

Applying the same integral transforms to the continuity conditions Eq. (3), then substituting Eqs. (9) and (10) into it, generates a group of linear equations in terms of $\Phi_1(\xi,p)$, $\Phi_2(\xi,p)$, and $\Psi(\xi,p)$. Solving the equations yields

$$\Phi_1=\frac{Q}{2\pi\mu p}\frac{\alpha_2}{\alpha_1}\frac{s_S^2p^2}{D_H(\xi,p)},\quad \Phi_2=-\frac{Q}{2\pi\mu p}\frac{s_S^2p^2+2\xi^2}{D_H(\xi,p)},$$

$$\Psi=-\frac{Q}{2\pi\mu p}\frac{2\alpha_2\xi}{D_H(\xi,p)}, \tag{11}$$

where

$$D_H(\xi,p)=(s_S^2p^2+2\xi^2)^2-4\xi^2\alpha_2\beta+\frac{\rho_1}{\rho_2}p^4s_S^4\frac{\alpha_2}{\alpha_1}. \tag{12}$$

Substituting Eq. (11) into Eqs. (9) and (10), we obtain the displacements and pressure in the fluid,

$$\bar{w}_1^{H_0}=\frac{Q}{2\pi}\frac{1}{\mu}\frac{\alpha_2 s_S^2p}{D_H(\xi,p)}e^{\alpha_1 z},$$

$$\bar{u}_1^{H_1}=-\frac{Q}{2\pi}\frac{1}{\mu}\frac{\alpha_2}{\alpha_1}\frac{s_S^2p\xi}{D_H(\xi,p)}e^{\alpha_1 z}, \tag{13}$$

$$\bar{P}^{H_0}=-\frac{Q}{2\pi}\frac{\rho_1}{\rho_2}\frac{\alpha_2}{\alpha_1}\frac{s_S^4p^3}{D_H(\xi,p)}e^{\alpha_1 z},$$

and the displacements and stresses in the solid

$$\bar{u}_2^{H_1}=\frac{Q}{2\pi}\frac{1}{\mu p}[(s_S^2p^2+2\xi^2)e^{-\alpha_2 z}-2\alpha_2\beta e^{-\beta z}]\frac{\xi}{D_H(\xi,p)},$$

$$\bar{w}_2^{H_0}=\frac{Q}{2\pi}\frac{1}{\mu p}[(s_S^2p^2+2\xi^2)e^{-\alpha_2 z}-2\xi^2e^{-\beta z}]\frac{\alpha_2}{D_H(\xi,p)}, \tag{14}$$

$$\bar{\tau}_{zz2}^{H_0}=-\frac{Q}{2\pi}\frac{1}{p}[(s_S^2p^2+2\xi^2)^2e^{-\alpha_2 z}-4\alpha_2\beta\xi^2e^{-\beta z}]\frac{1}{D_H(\xi,p)},$$

$$\bar{\tau}_{zr2}^{H_1}=\frac{Q}{2\pi}\frac{1}{p}[-e^{-\alpha_2 z}+e^{-\beta z}]\frac{2\alpha_2\xi(s_S^2p^2+2\xi^2)}{D_H(\xi,p)}.$$

III. CHARACTERISTIC EQUATION FOR LEAKY RAYLEIGH AND SCHOLTE WAVES

$D_H(\xi,p)=0$ is the characteristic equation for leaky Rayleigh waves at the interface of fluid–solid half-spaces. Introducing the substitution $\gamma=i\xi/ps_S$ and notations $q=s_L/s_S$ and $u=s_F/s_S$, the equation $D_H=0$ can be changed to the following in terms of γ:

$$(1-2\gamma^2)^2-4\gamma^2\sqrt{\gamma^2-1}\sqrt{\gamma^2-q^2}+\frac{\rho_1}{\rho_2}\frac{\sqrt{\gamma^2-q^2}}{\sqrt{\gamma^2-u^2}}=0. \tag{15}$$

Equation (15) is the same as that given by Viktorov (1967), which differs from the regular Rayleigh equation in the third term due to the pressure of the fluid. Equation (15) produces eight Riemann sheets owing to the square roots. The roots have the physical meaning of normalized slowness with respect to s_S. When leaky Rayleigh waves exist, we can obtain six roots, which include two pairs of complex conjugate pairs corresponding to the leaky Rayleigh wave $[\mathrm{Re}(\gamma_R^2)<u^2]$, and two opposite real roots corresponding to the Scholte wave $(\gamma_{sch}^2>u^2)$.

The two pairs of complex conjugate roots corresponding to the leaky Rayleigh wave take the form of $\pm[\mathrm{Re}(\gamma)\pm i\,\mathrm{Im}(\gamma)]$, in which only two roots are acceptable. Because p is the Laplace transform variable, $\mathrm{Re}(p)$ must be negative to have a finite time domain response. Therefore, only the roots which lead to $\mathrm{Re}(p)<0$ are acceptable, where $p=i\xi/\gamma s_S$. A quick analysis shows that the two acceptable roots take the form of $\gamma_{R1,R2}=\pm\mathrm{Re}(\gamma)-i\,\mathrm{Im}(\gamma)$. Here, we denote them as γ_{R1}, γ_{R2}, and note $\gamma_{R2}=-\bar{\gamma}_{R1}$, where the top bar indicates complex conjugate.

Analysis also shows that $|\mathrm{Re}(\gamma_{R1,R2})|<s_R/s_S$, where s_R is the slowness of Rayleigh wave in free surface solid half-space. This result confirms that the slowness of the leaky Rayleigh wave is smaller than s_R, i.e., in a fluid–solid half-space system, the leaky Rayleigh wave will travel faster than the pure Rayleigh wave in the same free surface solid. This conclusion is reasonable considering the influence of fluid pressure applied on solid surface. The imaginary part of γ_R accounts for the energy leaked into fluid. The larger the value of $|\mathrm{Im}(\gamma_{R1})|$, the more energy leaks into the fluid.

The Scholte wave velocity can be obtained from the real roots $\pm\gamma_{sch}$, where $\gamma_{sch}>u$ and $s_{sch}=\gamma_{sch}s_S$. Scholte waves always exist for any fluid–solid combination case. For the lighter fluids lying on stiffer solids case, s_{sch} is only slightly smaller than the acoustic wave velocity of the fluid, and the Scholte wave contribution is relatively small. The property of Scholte waves will be discussed in more detail in Sec. VII.

IV. SOLUTION OF PRESSURE IN THE FLUID

The solution of fluid pressure in the space–time domain is derived because pressure in the fluid is the most often measured quantity in practical application. By taking the inverse zeroth-order Hankel transform to Eq. (13) and introducing the substitution $\xi = p\eta$, we have

$$\overline{P} = -\frac{Qs_S^4}{2\pi}\frac{\rho_1}{\rho_2}\int_0^\infty \frac{\sqrt{\eta^2+s_P^2}}{\sqrt{\eta^2+s_F^2}}\frac{p}{D(\eta)} \times e^{pz\sqrt{\eta^2+s_F^2}} J_0(p\eta r)\,\eta\,d\eta, \tag{16}$$

where $D(\eta)$ is defined as

$$D(\eta) = (2\eta^2+s_S^2)^2 - 4\eta^2\sqrt{\eta^2+s_P^2}\sqrt{\eta^2+s_S^2} + \frac{\rho_1}{\rho_2}\frac{\sqrt{\eta^2+s_P^2}}{\sqrt{\eta^2+s_F^2}}s_S^4. \tag{17}$$

The inverse Laplace transform is evaluated by the Cagniard–de Hoop method, as described by Achenbach (1973). Introducing the following representation of $J_0(x)$,

$$J_0(x) = \frac{2}{\pi}\mathrm{Im}\int_1^\infty \frac{e^{ixs}}{\sqrt{s^2-1}}ds, \tag{18}$$

and substituting it into Eq. (16) yields

$$\overline{P} = -\frac{Qs_S^4}{\pi^2}\frac{\rho_1}{\rho_2}\mathrm{Im}\int_0^\infty \frac{\sqrt{\eta^2+s_P^2}}{\sqrt{\eta^2+s_F^2}}\frac{\eta}{D(\eta)}d\eta \times \int_1^\infty p\frac{e^{ip\eta rs+pz\sqrt{\eta^2+s_F^2}}}{\sqrt{s^2-1}}ds. \tag{19}$$

Introducing the substitution $t = -i\eta rs - z\sqrt{\eta^2+s_F^2}$ and $\eta = iv$ yields

$$\overline{P} = \frac{Qs_S^4}{\pi^2}\frac{\rho_1}{\rho_2}\mathrm{Im}\int_0^{-i\infty}\frac{\sqrt{s_P^2-v^2}}{\sqrt{s_F^2-v^2}}\frac{v}{D(iv)}dv \times \int_1^\infty p\frac{e^{-pt}}{\sqrt{s^2-1}}ds, \tag{20}$$

where

$$L^{-1}\left(\int_1^\infty p\frac{e^{-pt}}{\sqrt{s^2-1}}ds\right) = \frac{d}{dt}\left(\frac{H(t+z\sqrt{s_F^2-v^2}-vr)}{\sqrt{(t+z\sqrt{s_F^2-v^2})^2-v^2r^2}}\right) \tag{21}$$

and L represents the Laplace transform. Therefore, the inverse Laplace transform of Eq. (20) is

$$P(t) = \frac{Qs_S^4}{\pi^2}\frac{\rho_1}{\rho_2}\frac{d}{dt}\left(\mathrm{Im}\int_0^{-i\infty}\frac{\sqrt{s_P^2-v^2}}{\sqrt{s_F^2-v^2}}\frac{v}{D(iv)} \times \frac{H(t+z\sqrt{s_F^2-v^2}-vr)}{\sqrt{(t+z\sqrt{s_F^2-v^2})^2-v^2r^2}}dv\right) = \frac{Qs_S^4}{\pi^2}\frac{\rho_1}{\rho_2}\frac{d}{dt}G^P(t), \tag{22}$$

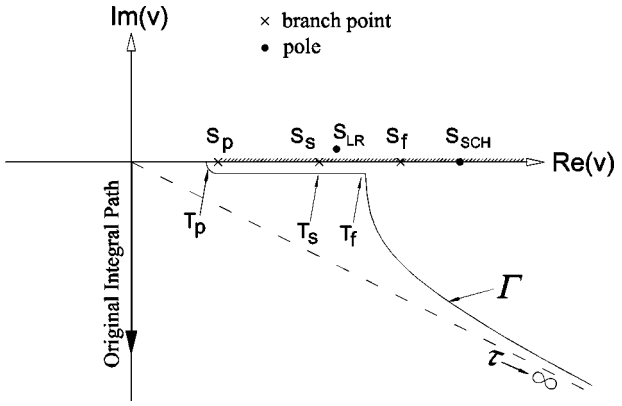


FIG. 2. Complex v plane and the modified integral path. The indices denote: P→leaky P wave, S→leaky S wave, LR→leaky R wave, f→fluid acoustic wave, SCH→Scholte wave.

where $G^P(t)$ is defined as the Green's function for pressure. The variable v is related to γ in Eq. (15) by $v = -\gamma s_S$. According to the analysis in the previous section, there are two poles in the right half complex v plane, which correspond to leaky Rayleigh and Scholte wave arrivals. The integral path along the negative imaginary axis can be deformed to path Γ as shown in Fig. 2. The new integration path Γ is defined by the equation $\tau = vr - z\sqrt{s_F^2-v^2}$, which can be solved for v to yield

$$v(\tau) = \frac{\tau r + z\sqrt{s_F^2R^2-\tau^2}}{R^2}, \quad \tau \text{ is real and } 0 \leqslant \tau \leqslant \infty, \tag{23}$$

where $R = \sqrt{r^2+z^2}$. The benefit of deforming the original integration path to Γ is obvious, because there is no pole along the new path. The Green's function integration along the new path is

$$G^P(t) = \left(\mathrm{Im}\int_{t_p}^t \frac{\sqrt{s_P^2-v(\tau)^2}}{\sqrt{s_F^2-v(\tau)^2}}\frac{v(\tau)}{D[iv(\tau)]} \times \frac{r - \tau z/\sqrt{s_F^2R^2-\tau^2}}{R^2\sqrt{[t-\tau+v(\tau)r]^2-v(\tau)^2r^2}}d\tau\right), \tag{24}$$

where $t_p = rs_P + |z|\sqrt{s_F^2-s_P^2}$ is the leaky P wave arrival time. $G^P(t) = 0$ when $t < t_p$.

However, when the receiver is located on the interface, i.e., $z = 0$, the integral path Γ is along the real axis, and passes by the Scholte pole. In this case, the principal value of Eq. (24) must be taken.

The integrand in Eq. (24) has a square root singularity at end point $\tau = t$, which increases the difficulty of numerical integration. This problem can be solved by further introducing the following transformation,

$$\tau = t_1\cos^2(\theta) + t_2\sin^2(\theta), \tag{25}$$

where t_1, t_2 are the lower and upper limits of the integration. Then the integration interval $t_1 \leqslant \tau \leqslant t_2$ is mapped to the fixed interval $0 \leqslant \theta \leqslant \pi/2$, while

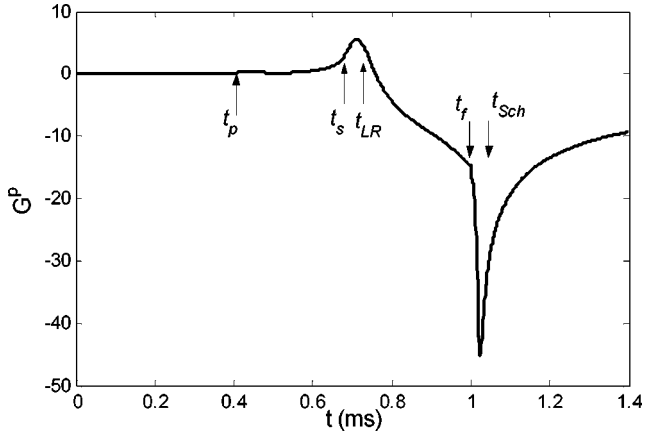


FIG. 3. Green's function $G^P(t)$ for pressure in the fluid at position $r=1.5$ m, $z=-0.05$ m. Material parameters for the fluid $\rho_1=1000$ kg/m^3, $c_F=1500$ m/s; for the solid, $\rho_2=2400$ kg/m^3, $c_P=4000$ m/s, $\upsilon=0.25$.

$$d\tau=2(t_2-t_1)\sin(\theta)\cos(\theta),$$

$$\sqrt{[t_2-\tau+v(\tau)r]^2-v(\tau)^2r^2}$$
$$=\sqrt{(t_2-t_1)}\cos(\theta)\sqrt{(t_2-t_1)\cos^2(\theta)+2v(\tau)r}. \quad (26)$$

Another square root singularity at point $\tau=t_f=s_FR$ can be processed similarly. A similar technique was also used by de Hoop (1984).

The vertical component of displacement in the fluid is also obtained in the same way:

$$w(t)=-\frac{Q}{\pi^2}\frac{s_S^4}{\rho_2}\mathrm{Im}\int_{t_p}^{t}\frac{\sqrt{s_P^2-v(\tau)^2}v(\tau)}{D[iv(\tau)]R^2}$$
$$\times\frac{r-\tau z/\sqrt{s_F^2R^2-\tau^2}}{\sqrt{[t-\tau+v(\tau)r]^2-v(\tau)^2r^2}}d\tau. \quad (27)$$

Figure 3 shows the Green's function $G^P(t)$ at a near-surface position $r=1.5$ m, $z=-0.05$ m in the fluid. The material configuration simulates a concrete/water system. The arrival times of all wave types are marked. It is noticed that the slope is discontinuous at positions $t=t_p$, t_s, and t_f, which correspond to the leaky P-, leaky S-, and fluid acoustic waves in water. From the mathematics viewpoint, these discontinuities come from the branch points along the integral path Γ. The poles corresponding to the leaky Rayleigh and Scholte wave arrivals contribute to the large smooth peaks. When the receiving position is very close to the interface, the integral path Γ in Fig. 2 will bend to the real axis, and a sharp peak will appear nearby the Scholte wave arrival in Fig. 3.

The impulse response to a point loading on the interface between a fluid and a solid half-space can be obtained from the corresponding step response solution by taking differentiation with respect to time. Then for any transient loading that has arbitrary temporal variation and spatial distribution, the response can be obtained by convolving the impulse response in both time and space domains.

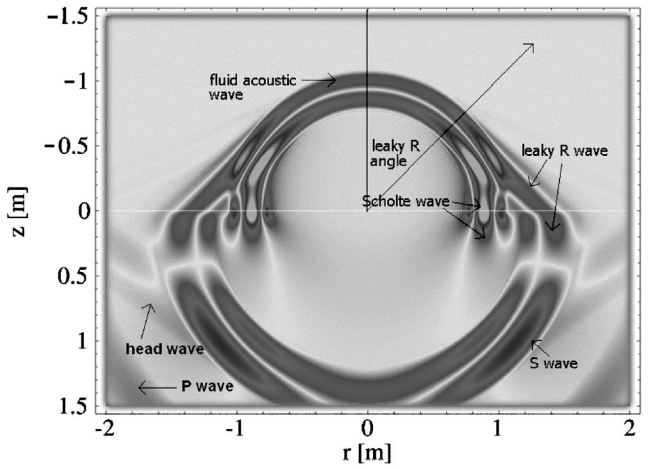


FIG. 4. Absolute value of pressure (in fluid) and stress τ_{zz} (in solid) field snapshot at $t=0.72$ ms given by EFIT analysis. The interface is subject to a transient point load that varies with time as $f(t)=\sin^2(\pi t/T)$ with $T=200$ μs.

V. VERIFICATION BY NUMERICAL SIMULATION

To verify the obtained analytical solutions, numerical analyses were performed to simulate the response for fluid–solid half-space cases. The elastodynamic finite integration technique (EFIT) is a numerical time-domain scheme to model elastic wave propagation in homogeneous and heterogeneous, dissipative and nondissipative, as well as isotropic and anisotropic elastic media (Fellinger *et al.*, 1995). EFIT uses a velocity-stress formalism on a staggered spatial and temporal grid complex. The starting point of EFIT is the integral form of the linear governing equations, i.e., the Cauchy equation of motion, and the equation of deformation rate. EFIT performs integrations over certain control volumes V, and the surfaces of these cells S, assuming constant velocity and stress within V and on each S. This method requires staggered grids and leads to a very stable and efficient numerical code, which also allows easy and flexible treatment of various boundary conditions. In recent years, EFIT has been successfully used for a wide variety of applications, especially in the field of nondestructive testing (Schubert and Marklein, 2003).

In the present case of a transient point load at a fluid–solid interface, we used a special axisymmetric EFIT code in cylindrical coordinates (Schubert *et al.*, 1998). The water/concrete case shown in Fig. 3 was studied. The material parameters are for water, $\rho_1=1000$ kg/m^3, $c_F=1500$ m/s, and for concrete $\rho_2=2400$ kg/m^3, P wave velocity $c_P=4000$ m/s, and Poisson's ratio $\nu=0.25$. The vertical transient point load varies with time as function $f(t)=\sin^2(\pi t/T)$, where the force duration is $T=200$ μs. A grid spacing of $\Delta r=\Delta z=2.5$ mm and a time step of 0.44 μs are used in order to guarantee stability as well as sufficient discretization of the shortest wavelengths. The dimensions of the model are 2 m in radial and 3 m in axial direction, resulting in 800×1200 grid cells. At the outer boundaries of the model, highly effective absorbing boundary conditions based on the perfectly matched layer (PML) are used in order to suppress interfering reflections (Liu, 1999).

Figure 4 shows the cross-sectional snapshot image of

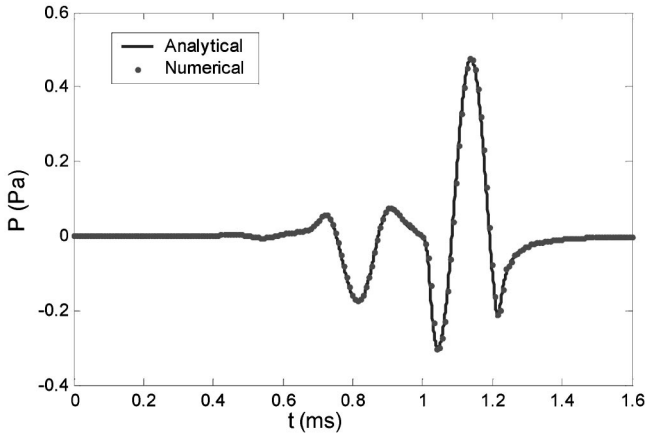


FIG. 5. Comparison of the analytical and the numerical solutions for pressure in the fluid at $r=1.5$ m, $z=-0.05$ m. The interface is subject to a point load that varies with time as $f(t)=\sin^2(\pi t/T)$ with $T=200\ \mu$s. Material parameters for the fluid $\rho_1=1000\ \mathrm{kg/m^3}$, $c_F=1500$ m/s; for the solid $\rho_2=2400\ \mathrm{kg/m^3}$, $c_P=4000$ m/s, $v=0.25$.

pressure field (absolute value) as generated by EFIT, at time $t=0.72$ ms. Only the region with $r>0$ was calculated, but for better illustration the reversed region with $r<0$ is also shown here. The half-circle in the upper half-plane represents the acoustic wave front in water, and the inclined lines represent the leaky Rayleigh wave fronts, which are tangent to the half circle at the leaky angle direction. In this case, the leaky Rayleigh angle determined by Snell's law is $\theta=43.8°$, measured from the vertical axis. The leaky Rayleigh wave front is separable from the subsequent Scholte and fluid acoustic wave fronts at larger values of radial distance r. The 3-D shape of the combined leaky Rayleigh and fluid acoustic wave fronts looks like a domed cone. The leaky Rayleigh wave in concrete is also seen, which behaves similarly to the ordinary Rayleigh wave, and attenuates exponentially with increasing depth in the solid. In the near-interface region, the Scholte wave effect is strong in both the fluid and the solid.

Figure 5 shows the analytical and numerical time domain near-surface response of the pressure in water, at position $r=1.5$ m, $z=-0.05$ m. Very good agreement between analytical and numerical responses is observed.

VI. SIMPLIFIED LEAKY RAYLEIGH WAVE RESPONSE

In Sec. IV, the complete analytical solution of pressure and displacement in the fluid are obtained. However, the integral form solution is not always convenient to use. In this and the next sections, the authors provide a simplified closed-form solution to the same problem by considering only Rayleigh and Scholte pole contributions. This simplification is acceptable when measurements are taken at a large distance from the source.

According to the previous analysis, the wave field excited by a normal point load at the interface includes contributions from leaky P, S, Rayleigh, fluid acoustic, and Scholte waves. Analysis shows that, at large horizontal distances from the source, disturbances near the interface are dominated by leaky Rayleigh and Scholte wave contributions, which can be obtained from the residues at corresponding poles. The similar idea was already used to investigate Rayleigh wave effects in a free surface half-space by Chao *et al.* (1961). Achenbach (1973) reproduced Chao's results in detail. The simplified solution provides an easy and quick way to estimate leaky Rayleigh and Scholte wave effects.

A. Displacements and stresses in the fluid

Applying the inverse Hankel and Laplace transforms to the transformed solution in Eq. (13), the pressure in fluid can be expressed as

$$P(r,z,t)=-\frac{Q}{2\pi}\frac{s_S^4}{2\pi i}\frac{\rho_1}{\rho_2}\int_0^\infty J_0(\xi r)\xi \times\int_{\epsilon-i\infty}^{\epsilon+i\infty}\frac{\alpha_2}{\alpha_1}\frac{p^3}{D_H(\xi,p)}e^{\alpha_1 z+pt}\,dp\,d\xi. \tag{28}$$

Considering the integrand term, the contributions from the leaky Rayleigh poles are

$$I_1=\frac{1}{2\pi i}\int_0^\infty J_0(\xi r)\xi\int_{\epsilon-i\infty}^{\epsilon+i\infty}\frac{\alpha_2}{\alpha_1}\frac{p^3}{D_H(\xi,p)}e^{\alpha_1 z+pt}\,dp\,d\xi =\int_0^\infty J_0(\xi r) \times\xi\left[\frac{\alpha_2}{\alpha_1}\frac{p^3}{\partial D_H(\xi,p)/\partial p}e^{\alpha_1 z+pt}\right]_{p=i\xi/s_S\gamma_{R1},\,i\xi/s_S\gamma_{R2}}d\xi, \tag{29}$$

where the expression $[(\alpha_2/\alpha_1)\times[p^3/\partial D_H(\xi,p)/\partial p]]_{p=i\xi/s_S\gamma_{R1},\,i\xi/s_S\gamma_{R2}}$ represents a pair of complex conjugate constants, denoted as A_1 and $\overline{A_1}$.

Introducing

$$m_{1,2}=-\frac{i}{\gamma_{R1,R2}}\left[\frac{z}{r}\sqrt{u^2-\gamma_{R1,R2}^2}+\frac{t}{s_S r}\right] \tag{30}$$

generates

$$[\alpha_1 z+pt]_{p=i\xi/s_S\gamma_{R1,R2}}=-\xi r m_{1,2}, \tag{31}$$

where $m_1=\overline{m_2}$. Thus, Eq. (29) can be expressed as

$$I_1=A_1\int_0^\infty J_0(\xi r)\xi e^{-\xi m_1 r}d\xi+\overline{A_1}\int_0^\infty J_0(\xi r)\xi e^{-\xi m_2 r}\,d\xi =2\,\mathrm{Re}\int_0^\infty A_1 e^{-\xi m_1 r}J_0(\xi r)\xi\,d\xi =\frac{2}{r^2}\mathrm{Re}\left[\frac{A_1 m_1}{(1+m_1^2)^{3/2}}\right],\quad \mathrm{Re}(m_1)>0, \tag{32}$$

where we use the zeroth-order Hankel transform formula

$$\int_0^\infty J_0(\xi r)\xi e^{-a\xi}\,d\xi=\frac{a}{(r^2+a^2)^{3/2}},\quad \mathrm{Re}(a)>0. \tag{33}$$

From Eq. (28), the pressure P is expressed as

$$P=-\frac{Q}{\pi}\frac{\rho_1}{\rho_2}\frac{1}{r^2}s_S^4\,\mathrm{Re}\left[\frac{A_1 m_1}{(1+m_1^2)^{3/2}}\right]. \tag{34}$$

Similarly, by introducing

$$A_2=\xi\left[\frac{\alpha_2 p}{\partial D_H(\xi,p)/\partial p}\right]_{p=i\xi/s_S\gamma_{R1}},$$
$$A_3=\left[\frac{\alpha_2}{\alpha_1}\frac{p\xi^2}{\partial D_H(\xi,p)/\partial p}\right]_{p=i\xi/s_S\gamma_{R1}}, \tag{35}$$

the vertical and radial components of the displacement in the fluid are obtained:

$$w_1=\frac{Q}{\pi}\frac{s_S^4}{\rho_2}\frac{1}{r}\mathrm{Re}\left[\frac{A_2}{(1+m_1^2)^{1/2}}\right],$$
$$u_1=-\frac{Q}{\pi}\frac{s_S^2}{\mu}\frac{1}{r}\mathrm{Re}\left(\left[1-\frac{m_1}{(1+m_1^2)^{1/2}}\right]A_3\right), \tag{36}$$
$$\mathrm{Re}(m_1)>0.$$

Care should be taken with the square root when calculating $m_{1,2}$ from Eq. (30). To have bounded results, only the branch that gives $\mathrm{Re}(m_1)>0$ should be selected.

B. Displacement and stress in the solid half-space

The response of the leaky Rayleigh wave in the solid can be obtained in a similar way. Introducing

$$n_{p1,2}=\frac{1}{\gamma_{R1,R2}}\left[\pm\frac{z}{r}\sqrt{\gamma_{R1,R2}^2-q^2}-i\tau\right],$$
$$n_{s1,2}=\frac{1}{\gamma_{R1,R2}}\left[\pm\frac{z}{r}\sqrt{\gamma_{R1,R2}^2-1}-i\tau\right] \tag{37}$$

yields

$$[-\alpha_2 z+pt]_{p=i\xi/s_S\gamma_{R1,R2}}=-\xi\frac{r}{\gamma_{R1,R2}}\left[\pm\sqrt{\gamma_{R1,R2}^2-q^2}\frac{z}{r}-i\tau\right]=-\xi r n_{p1,2},$$
$$[-\beta z+pt]_{p=i\xi/s_S\gamma_{R1,R2}}=-\xi\frac{r}{\gamma_{R1,R2}}\left[\pm\sqrt{\gamma_{R1,R2}^2-1}\frac{z}{r}-i\tau\right]=-\xi r n_{s1,2}. \tag{38}$$

The real part of Eq. (38) must be negative to have bounded responses, therefore the real parts of n_{p1}, n_{p2}, n_{s1}, n_{s2} must be positive. Using the similar argument for dealing with $m_{1,2}$, only those results of n_{p1}, n_{p2}, n_{s1}, n_{s2} that have positive real parts are acceptable. In addition, n_{p1}, n_{p2} and n_{s1}, n_{s2} should be complex conjugate pairs.

Applying the inverse Hankel and Laplace transforms to Eq. (14) and calculating the residues at Rayleigh poles yields the displacements and stresses in the solid,

$$w_2=\frac{Q}{\pi}\frac{s_S^2}{\rho_2}\frac{1}{r}\mathrm{Re}\left[\frac{B_1}{\sqrt{1+n_{p1}^2}}-\frac{B_2}{\sqrt{1+n_{s1}^2}}\right],$$
$$u_2=\frac{Q}{\pi}\frac{s_S^2}{\rho_2}\frac{1}{r}\mathrm{Re}\left[B_3\left(1-\frac{n_{p1}}{\sqrt{1+n_{p1}^2}}\right)-B_4\left(1-\frac{n_{s1}}{\sqrt{1+n_{s1}^2}}\right)\right], \tag{39}$$
$$\tau_{zz2}=-\frac{Q}{\pi}\frac{1}{r^2}\mathrm{Re}\left[\frac{B_5 n_{p1}}{(1+n_{p1}^2)^{3/2}}-\frac{B_6 n_{s1}}{(1+n_{s1}^2)^{3/2}}\right],$$
$$\tau_{zr2}=\frac{Q}{\pi}\frac{1}{r^2}\mathrm{Re}\left[\frac{-B_7}{(1+n_{p1}^2)^{3/2}}+\frac{B_7}{(1+n_{s1}^2)^{3/2}}\right],$$

where the coefficients are

$$B_1=\left[\frac{s_S^2p^2+2\xi^2}{p\partial D_H(\xi,p)/\partial p}\alpha_2\xi\right]_{p=i\xi/s_S\gamma_{R1}},$$
$$B_2=\left[\frac{2\xi^3\alpha_2}{p\partial D_H(\xi,p)/\partial p}\right]_{p=i\xi/s_S\gamma_{R1}},$$
$$B_3=\left[\frac{(s_S^2p^2+2\xi^2)}{p\partial D_H(\xi,p)/\partial p}\xi^2\right]_{p=i\xi/s_S\gamma_{R1}}, \tag{40}$$
$$B_4=\left[\frac{2\xi^2\alpha_2\beta}{p\partial D_H(\xi,p)/\partial p}\right]_{p=i\xi/s_S\gamma_{R1}},$$
$$B_5=\left[\frac{(s_S^2p^2+2\xi^2)^2}{p\partial D_H(\xi,p)/\partial p}\right]_{p=i\xi/s_S\gamma_{R1}},$$
$$B_6=2B_4,\quad B_7=2B_1.$$

The expressions in Eq. (39) are only valid in the region where $\mathrm{Re}(n_{p1,p2})>0$ and $\mathrm{Re}(n_{s1,s2})>0$. Other stress components in the solid can be derived from the following stress–displacement relations:

$$\tau_{rr2}=(\lambda_2+2\mu)\frac{\partial u_2}{\partial r}+\lambda_2\left(\frac{u_2}{r}+\frac{\partial w_2}{\partial z}\right),$$
$$\tau_{\theta\theta2}=(\lambda_2+2\mu)\frac{u_2}{r}+\lambda_2\left(\frac{\partial u_2}{\partial r}+\frac{\partial w_2}{\partial z}\right). \tag{41}$$

C. Attenuation and dispersion of leaky Rayleigh waves

In addition to the geometric decay due to the effect of point loading, which varies as $1/\sqrt{r}$ along the interface for the Rayleigh wave, there is another type of attenuation caused by continuous radiation (leakage) of energy into the fluid. In frequency domain, the solutions are exponential functions of $(-\xi r)$, where ξ has the physical meaning of wavenumber. According to Eq. (32), higher frequency (larger ξ) contents give more attenuation during propagation. Therefore the waveform generated by a transient loading becomes wider with increasing distance, i.e., it shows dispersion property due to leakage-induced attenuation, although the phase velocity of leaky Rayleigh waves does not vary with frequency.

VII. SCHOLTE WAVE RESPONSE

The real roots of the Scholte equation correspond to Scholte wave propagation along the interface. The Scholte wave solutions can be obtained by calculating the residues at the poles $\gamma=\pm\gamma_{sch}$. For common cases of light fluids lying

on stiff solids, the root $|\gamma_{sch}| = s_{sch}/s_S$ is slightly larger than $u = s_F/s_S$. Therefore, Eqs. (31) and (38) are changed to

$$
\begin{aligned}
&[\alpha_1 z + pt]_{p=i\xi/\pm s_S\gamma_{sch}} = -\xi r m_{1,2}^{sch},\\
&[-\alpha_2 z + pt]_{p=i\xi/\pm s_S\gamma_{sch}} = -\xi r n_{p1,2}^{sch},\\
&[-\beta z + pt]_{p=i\xi/\pm s_S\gamma_{sch}} = -\xi r n_{s1,2}^{sch},
\end{aligned}
\tag{42}
$$

where

$$
\begin{aligned}
m_{1,2}^{sch} &= -\frac{1}{\gamma_{sch}}\left(\sqrt{\gamma_{sch}^2 - u^2}\frac{z}{r} \pm i\frac{t}{s_S r}\right),\\
n_{p1,2}^{sch} &= \frac{1}{\gamma_{sch}}\left(\sqrt{\gamma_{sch}^2 - q^2}\frac{z}{r} \pm i\frac{t}{s_S r}\right),\\
n_{s1,2}^{sch} &= \frac{1}{\gamma_{sch}}\left(\sqrt{\gamma_{sch}^2 - 1}\frac{z}{r} \pm i\frac{t}{s_S r}\right).
\end{aligned}
\tag{43}
$$

The solutions for the leaky Rayleigh wave are also valid for the Scholte wave by substituting m^{sch}, n_p^{sch}, and n_s^{sch} for m, n_p, and n_s. The first terms in the expressions of Eq. (43) represent the real parts, which are positive and result in the decay in z direction. It can be seen that the Scholte wave decays exponentially in z direction in both the fluid and the solid. For lighter fluid cases, i.e., where the acoustic impedance of the fluid is less than that of the solid, $\sqrt{\gamma_{sc}^2 - u^2}$ is much smaller than $\sqrt{\gamma_{sc}^2 - q^2}$ and $\sqrt{\gamma_{sc}^2 - 1}$. This indicates that the Scholte wave attenuates much more quickly in the solid than in the fluid. Therefore, in contrast to the leaky Rayleigh wave, most of Scholte wave energy is localized in the fluid (Gusev *et al.*, 1996). The Scholte wave generation efficiency increases with increasing acoustic impedance of the fluid. For example, it is much easier to generate Scholte waves in the water/concrete configuration than the air/concrete configuration. In fact, almost no Scholte wave effect can be observed in the air/concrete case. With increasing impedance of the fluid, Scholte waves have deeper penetration depth in the solid. This property provides the possibility for NDT application of Scholte waves, which was studied experimentally by Glorieux *et al.* (2001). Because there is no leakage during Scholte wave propagation along the radial direction, the decay in the radial direction is only attributed to the geometrical effect. In 2-D cases, the Scholte wave travels without attenuation along the propagation direction (Glorieux *et al.*, 2001).

VIII. COMPARISON OF THE EXACT AND SIMPLIFIED SOLUTIONS

Figure 6 shows the comparison of the exact and simplified analytical solutions for fluid pressure for the water/concrete case. In Fig. 6(a), when the receiver position is close to the interface ($r = 1.5$ m, $z = -0.05$ m), good agreement is observed around the leaky Rayleigh and Scholte wave arrival times. The small yet noticeable differences before the leaky Rayleigh and Scholte wave arrivals are due to the absence of leaky body waves and fluid acoustic waves in the simplified solution. When the receiver is away from the interface, as shown in Fig. 6(b) for receiver position $r = 1.5$ m, $z = -0.5$ m, the degree of agreement between the simplified and exact analytical solutions decreases. Generally speaking, the simplified solution provides better estimation of pressure for larger r and smaller $|z|$ cases, where the contribution of body waves is negligible.

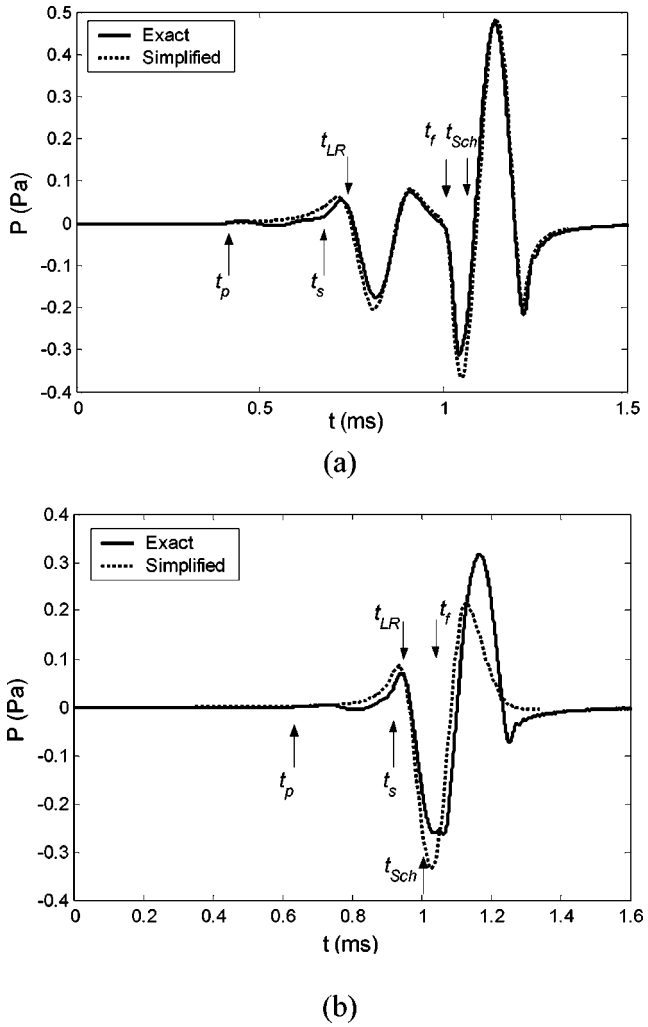


FIG. 6. Comparison of the exact and simplified solutions for water/concrete case. Pressure in the fluid at position (a) $r = 1.5$ m, $z = -0.05$ m, and (b) $r = 1.5$ m, $z = -0.5$ m. The interface is subject to a point load that varies with time as $f(t) = \sin^2(\pi t/T)$ with $T = 200$ μs. Material parameters for the fluid $\rho_1 = 1000$ kg/m^3, $c_F = 1500$ m/s; and for the solid $\rho_2 = 2400$ kg/m^3, $c_P = 4000$ m/s, $v = 0.25$.

Figures 7(a) and (b) show the comparison of the exact and simplified solutions for the air/concrete case. The receiver positions are $r = 1.0$ m, $z = -0.05$ m and $r = 1.0$ m, $z = -0.5$ m, respectively. Good agreement is observed near the leaky Rayleigh wave arrival time for both positions, while obvious differences can be seen near the Scholte and fluid acoustic wave arrival times. The reason is that the leaky Rayleigh wave is well separated from the fluid acoustic wave, and Scholte wave contribution is very small compared to the acoustic wave contribution in the fluid for air/concrete case, even in the near-interface region. Therefore, for the air/concrete configuration, the pressure field in the fluid is dominated by leaky Rayleigh and fluid acoustic waves. In air-coupled sensing, the leaky Rayleigh wave is usually the component in which we are interested. The acoustic wave contribution in the fluid can be separated by increasing mea-

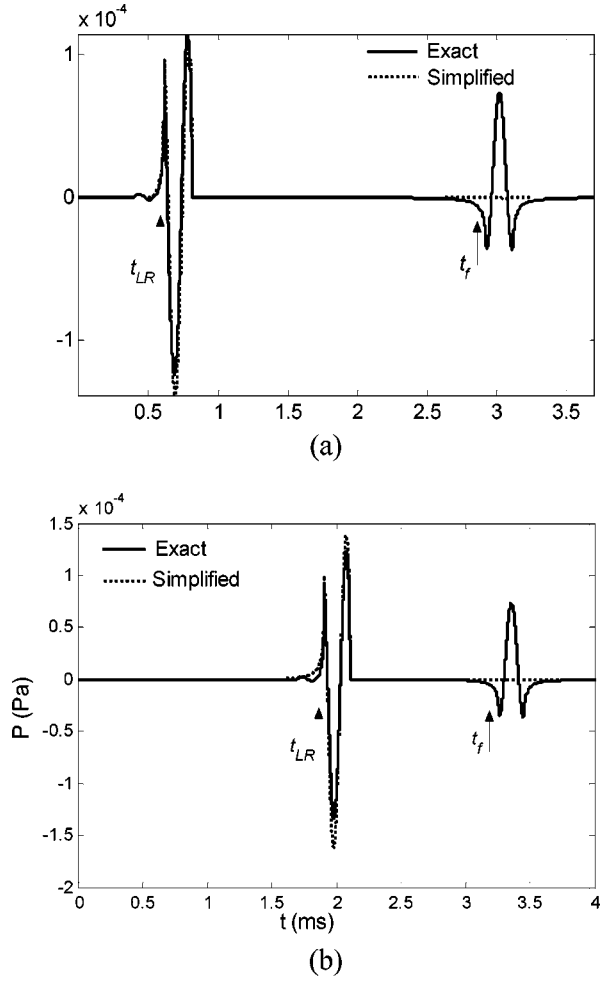


FIG. 7. Comparison of the exact and simplified solutions for air/concrete case. Pressure in the fluid at position (a) $r=1.0$ m, $z=-0.05$ m, and (b) $r=1.0$ m, $z=-0.5$ m. The interface is subject to a point load that varies with time as $f(t)=\sin^2(\pi t/T)$ with $T=200$ μs. Material parameters for the fluid $\rho_1=1.21$ kg/m^3, $c_F=343$ m/s; for the solid $\rho_2=2400$ kg/m^3, $c_P=4000$ m/s, $\nu=0.25$.

suring distance r between receivers and the source, or eliminated by employing the directional property of the air-coupled sensor (Zhu *et al.*, 2001).

The excitation effectiveness of leaky Rayleigh waves induced by an impact point load can be inferred from Fig. 7. For a typical impulse force $f(t)=\sin^2(\pi t/T)$ with a modest peak value of 1 kN and duration $T=200$ μs, the output pressure of the leaky Rayleigh wave is 0.1–0.15 Pa, which is approximately equivalent to a sound pressure level of 75 dB. Such a pressure is large enough to be detected readily by an air-coupled sensor, even when material attenuation effects are considered. The excitation effectiveness of leaky Rayleigh waves is dependent on the impact force duration. Shorter force durations give higher output pressure of leaky Rayleigh waves. For example, when the duration is decreased to $T=50$ μs, the output peak pressure of the leaky Rayleigh wave will increase to 1.2 Pa (95 dB). Most impactors used for concrete testing will generate transient forces with duration between 50 and 200 μs.

IX. CONCLUSIONS

The exact analytical solution to the ''Lamb'' problem in a fluid/solid half-space system is derived by the Cagniard–de Hoop method. Simplified formulations are also derived, which provide an easy and quick way to estimate leaky Rayleigh and Scholte wave contributions. The following conclusions can be drawn based on the analysis:

(1) A transient point load applied to the interface is an effective way to generate leaky Rayleigh waves in the fluid. For air-coupled wave detection in concrete, the excitation effectiveness of leaky Rayleigh waves is around 0.1–1.0 Pa/kN, depending on the impact force duration.
(2) For the light fluid/heavy solid case, the leaky Rayleigh wave is separable from Scholte and acoustic waves in the fluid when distance r is large enough, where r depends on velocities of leaky Rayleigh, Scholte, and acoustic waves, vertical distance $|z|$, and force duration. For the air–concrete configuration shown in Fig. 7(a), where $|z|=0.05$ m and $r>0.2$ m, the difference in arrival time between leaky Rayleigh and acoustic waves is >362 μs. Therefore the received signals will be dominated by leaky Rayleigh waves, which provide important material information of the underlying solid.
(3) Simplified solutions are obtained when contributions from leaky Rayleigh waves and Scholte waves poles only are considered. Equations (34)–(36) and (39) give the solution to responses in the fluid and solid, respectively. The simplified solution accurately simulates the transient pressure field response for the air/concrete case when the fluid acoustic wave contribution is removed or separated.
(4) The Scholte wave contribution is prominent in the near-interface region for the water/concrete case. Because most of the energy of Scholte waves is localized in the fluid, however, Scholte wave properties are not very sensitive to the variation of the underlying solid materials, which limits the NDE application of Scholte waves for the common light fluid/heavy solid cases.

ACKNOWLEDGMENTS

This work was carried out in the course of research sponsored by the National Science Foundation under Grant No. 0223819. The authors also wish to thank Dr. Nelson Hsu from National Institute of Standards and Technology for his helpful suggestions.

Achenbach, J. D. (**1973**). *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, The Netherlands), Chap. 7, pp. 262–321.
Castaings, M., and Hosten, B. (**2001**). ''Lamb and SH waves generated and detected by air-coupled ultrasonic transducers in composite material plates,'' NDT & E Int. **34**, 249–258.
Chao, C. C. (**1961**). ''Surface waves in an elastic half space,'' J. Appl. Mech. **28**, 300–301.
de Hoop, A. T., and van der Hijden, J. (**1983**). ''Generation of acoustic waves by an impulsive line source in a fluid/solid configuration with a plane boundary,'' J. Acoust. Soc. Am. **74**, 333–342.
de Hoop, A. T., and van der Hijden, J. (**1984**). ''Generation of acoustic waves by an impulsive point source in a fluid/solid configuration with a plane boundary,'' J. Acoust. Soc. Am. **75**, 1709–1715.
Fellinger, P., Marklein, R., Langenberg, K. J., and Klaholz, S. (**1995**). ''Numerical modelling of elastic wave propagation and scattering with EFIT—Elastodynamic finite integration technique,'' Wave Motion **21**, 47–66.
Glorieux, C., and Van de Rostyne, K. (**2001**). ''On the character of acoustic waves at the interface between hard and soft solids and liquids,'' J. Acoust. Soc. Am. **110**, 1299–1306.

Gusev, V., Desmet, C., Lauriks, W., Glorieux, C., and Thoen, J. (**1996**). ''Theory of Scholte, leaky Rayleigh, and lateral wave excitation via the laser-induced thermoelastic effect,'' J. Acoust. Soc. Am. **100**, 1514–1528.
Liu, Q. H. (**1999**). ''Perfectly matched layers for elastic waves in cylindrical and spherical coordinates,'' J. Acoust. Soc. Am. **105**, 2075–2084.
Mozhaev, V. G., and Weihnacht, M. (**2002**). ''Subsonic leaky Rayleigh waves at liquid-solid interfaces,'' Ultrasonics **40**, 927–933.
Qi, Q. (**1994**). ''Attenuated leaky Rayleigh waves,'' J. Acoust. Soc. Am. **95**, 3222–3230.
Schubert, F., and Marklein, R. (**2003**). ''Numerical Computation of Ultrasonic Wave Propagation in Concrete using the Elastodynamic Finite Integration Technique (EFIT),'' in *Proc. IEEE Ultrasonics Symp.*, Munich, Germany, 8–11 October, 2002, Article 5G-5, 778–783 (on CD).
Schubert, F., Peiffer, A., Koehler, B., and Sanderson, T. (**1998**). ''The elastodynamic finite integration technique for waves in cylindrical geometries,'' J. Acoust. Soc. Am. **104**, 2604–2614.
Viktorov, I. A. (**1967**). *Rayleigh and Lamb Waves* (Plenum, New York).
Zhu, J., and Popovics, J. S. (**2001**). ''Non-contact detection of surface waves in concrete using an air-coupled sensor,'' in *Review of Progress in Quantitative Nondestructive Evaluation* (American Institute of Physics, Melville, NY), Vol. 20B, 1261–1268.

Development of local volume displacement sensors for vibrating plates

Marcellin Zahui[a] and Ronald Wendt
Mechanical Engineering Department, University of North Dakota, Grand Forks, North Dakota 58202

(Received 28 October 2003; revised 22 June 2004; accepted 9 July 2004)

Development of local volume displacement sensors is presented. This development supports the implementation of noise control techniques that are based on minimization of local volume displacements, velocities, or accelerations of a vibrating structure. In this paper, a general methodology for the development of local volume displacement sensors for vibrating plates using polyvynilidene fluoride (PVDF) is presented. This methodology was verified experimentally for a clamped plate. The local volume displacement measured using a PVDF sensor matched the local volume displacement found using multiple accelerometer measurements. The resulting sensor spans the entire x-length of the plate and covers the width of the area of interest. The sensor is composed of several strips of PVDF whose width varies with x quadratically on the local area and linearly elsewhere. Design issues for a clamped plate, a simply supported plate and a plate with arbitrary boundary conditions are discussed along with a presentation of some sample sensor shapes. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1786851]

PACS numbers: 43.50.Ki, 43.50.Yw [KAC] Pages: 2111–2117

I. INTRODUCTION

It has been widely accepted that minimization of volume displacement results in lower sound radiation. Publications by Guigou *et al.*,[1] Gardonio *et al.*,[2] Johnson and Elliott,[3] and Charette *et al.*[4] introduced and explored the concept of the active control of sound radiated from structures by sensing and controlling the total volume displacement of the structure. The implementation of this technique was based on novel development of distributed piezoelectric laminates originally proposed by Lee and Moon[5] for control and sensing of bending vibration of flexible beams and plates. Using concepts similar to modal filtering, Lee and Moon described a design for polyvynilidene fluoride (PVDF) modal sensors/actuators for some types of beams and plates.

The methodology pursued here is different from the work of previous authors in that it relies on sensing and minimizing the local volume displacement (instead of total volume displacement) over a broad band of frequencies.

Papers by Mason *et al.*[6] and Naghshineh and Mason[7,8] gave a detailed discussion of actively controlling the sound radiated from vibrating structures based on sensing and minimizing the local volume displacement, velocity, or acceleration (assuming $e^{i\omega t}$ time dependence) of a vibrating structure. Noise reduction is achieved by distributing an array of these noise-control devices over the surface of the radiating structure. Each device is controlled independently of the other devices and consists of a motion sensor, a controller, and a loudspeaker. The loudspeaker is driven such that it reduces the volume displacement of the radiating structure within its close proximity. Since each device is controlled independently, the controller can be very simple and made of low cost components.

Measurement of local volume displacement of a structure necessitates the division of the radiating surface of the vibrating structure into many smaller areas. The volume displacements of each of these smaller areas need to be sensed before any cancellation is possible. Naghshineh and Mason[7,8] proposed one control device per each half wavelength of plate vibration. Thus, because active noise control primarily targets low frequency applications, noise radiating plates are often divided into four or six areas. The accurate measurement of the volume displacement of theses areas will require an array of discrete sensors and additional computation in the control loop. Thus, distributed sensors have been investigated as local motion sensor. Following the path successfully adopted by researchers mentioned in the previous paragraph, PVDF was selected as the local motion sensor.

The initial implementation of this concept was tested in a uniformly vibrating baffled piston.[8] A nonuniformly vibrating baffled beam was chosen next as a test bed. Development of a local volume displacement sensor for a vibrating beam proved to be a challenging task as shown by Zahui *et al.*[9,10] PVDF is a strain sensitive material, and, as such, it provides local strain measurements. However, relating these local strain measurements to a displacement related quantity (e.g., volume displacement) requires the extension of the sensor to a nonvibrating reference boundary which makes the implementation of the sensor somewhat impractical for the cases where the structure is divided into many sections (this issue addressed in the conclusion). In the preliminary implementation of the sensor, modal filtering formulation presented by Guigou *et al.*[1] was followed and adapted to the development of local volume displacement sensors. Next, the method developed by Rex[11] and Gardonio *et al.*[2] for development of a quadratic-shaped PVDF sensor was extended to the local sensing application and active control of sound radiated from a clamped beam divided into two sections was demonstrated. The control scheme was found to work within a narrow frequency range. The present goal is to develop these sensors

[a]Electronic mail: marcellin_zahui@mail.und.nodak.edu

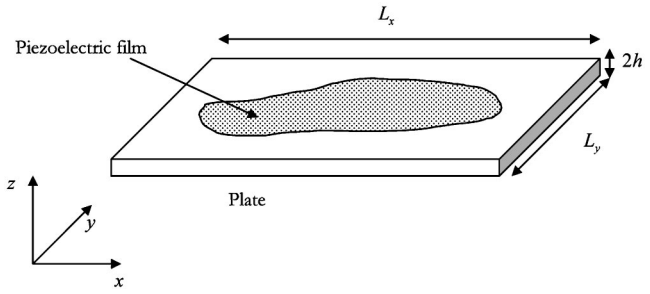


FIG. 1. Plate with bonded piezoelectric film.

for the case of a vibrating plate (i.e., advancing from a 1-D case to a 2-D case).

This paper describes the development of a general formulation for design of local volume displacement sensors using PVDF for a nonuniformly vibrating rectangular plate. The formulation presented by Gardonio *et al.*[2] for a total volume displacement sensor and reviewed in the next section has been extended to the case of a local volume displacement sensor and presented in Sec. III. The sensor designed for a clamped plate was experimentally verified as described in Sec. IV.

II. TOTAL VOLUME DISPLACEMENT SENSOR DEVELOPMENT

As mentioned previously, publications by Guigou *et al.*,[1] Gardonio *et al.*,[2] Johnson and Elliott,[3] and Charette *et al.*[4] have discussed the concept of the active control of sound radiated from structures by sensing and controlling the total volume displacement of the structure. These researchers made use of distributed sensors made of polyvinylidene fluoride (PVDF) that are shaped to provide a measurement of the total volume displacement of a vibrating beam or plate. This section presents a brief review of the development of total volume displacement PVDF sensors as presented by Gardonio *et al.*[2] For the plate of Fig. 1, consider a PVDF film which is attached to the surface of the plate and covers an area described by the function $F(x,y)$. The total charge output q of the film is given by[5]

$$q=\int_0^{Ly}\int_0^{Lx}(-h)F(x,y)[e_{31}w_{xx}+e_{32}w_{yy}]dxdy, \qquad (1)$$

where Lx and Ly are the dimensions of the plate in the x and y directions respectively, e_{31} and e_{32} are the PVDF sensor stress/charge coefficients, w_{xx} and w_{yy} are the second partial derivatives with respect to x and y, respectively, of the plate surface displacement field $w(x,y)$ (i.e., displacement in the z direction) such that

$$w_{xx}=\frac{\partial^2 w(x,y)}{\partial x^2} \quad \text{and} \quad w_{yy}=\frac{\partial^2 w(x,y)}{\partial y^2}$$

and h is given by the following equation:

$$h=\frac{h_p+h_s}{2}, \qquad (2)$$

where h_p and h_s define the plate and sensor thicknesses, respectively.

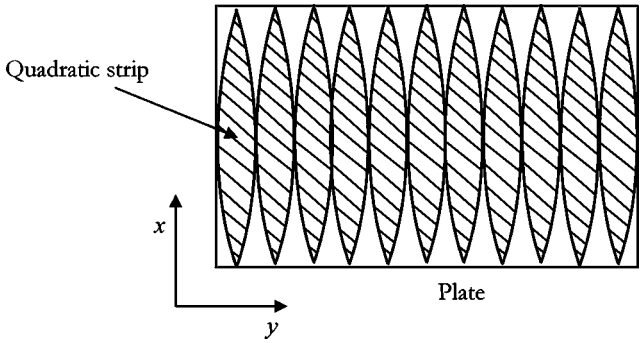


FIG. 2. Total volume displacement sensor proposed by Gardonio *et al.*[2]

The volume displacement D of the plate can be defined as the integral of the displacement over the plate:

$$D=\int_0^{Ly}\int_0^{Lx} w(x,y)\,dx\,dy. \qquad (3)$$

Equation (1) can be separated into two charges q_x and q_y such that

$$q_x=-h\int_0^{Ly}\int_0^{Lx} e_{31}w_{xx}F(x,y)dxdy \qquad (4a)$$

and

$$q_y=-h\int_0^{Ly}\int_0^{Lx} e_{32}w_{yy}F(x,y)dxdy. \qquad (4b)$$

q_x represents the charge output due to bending in the x direction and q_y the charge output due to bending in the y direction. The total charge output is q_x+q_y. Integrating by parts Eq. (4a) along L_x and Eq. (4b) along L_y yields

$$q_x=-e_{31}h\int_0^{Ly}\left\{[w_xF]_0^{Lx}-[F_xw]_0^{Lx}+\int_0^{Lx}F_{xx}w\,dx\right\}dy, \qquad (5a)$$

$$q_y=-e_{32}h\int_0^{Lx}\left\{[w_yF]_0^{Ly}-[F_yw]_0^{Ly}+\int_0^{Ly}F_{yy}w dy\right\}dx, \qquad (5b)$$

where the subscripts in w_x, F_x, F_{xx}, F_y, and F_{yy} represent the partial derivatives such that

$$w_x=\frac{\partial w}{\partial x}, \quad F_x=\frac{\partial F}{\partial x}, \quad F_{xx}=\frac{\partial^2 w}{\partial x^2},$$

$$F_y=\frac{\partial F}{\partial y}, \quad \text{and} \quad F_{yy}=\frac{\partial^2 F}{\partial y^2}.$$

The shape function $F(x,y)$ is defined to be quadratic in the x direction such that $F(x,y)$ is independent of y:

$$F(x,y)=ax^2+bx+c. \qquad (6)$$

Equations (3), (5), and (6) are used along with the plate boundary conditions to define the sensor shape and location on the plate. Gardonio *et al.*[2] developed the sensor shape shown in Fig. 2 for a clamped plate. They suggested techniques to develop a similar sensor for simply supported

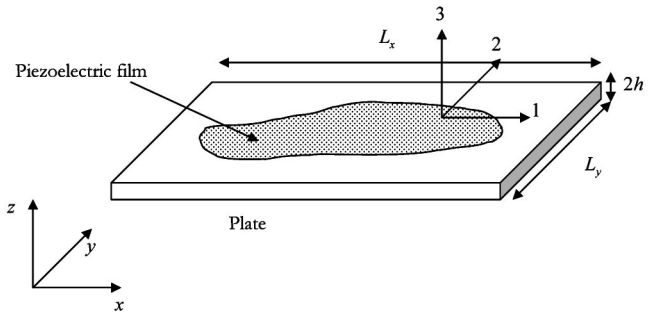


FIG. 3. PVDF orientation relative to the xyz axis.

plates and plates with arbitrary boundary conditions. Their methodology is extended to the development of a local volume displacement sensor for plates and is presented in the next section.

III. LOCAL VOLUME DISPLACEMENT SENSOR DEVELOPMENT

A. Basic equations

The goal in this paper is to design a sensor that accurately measures the volume displacement over any subarea of a plate. The proposed design adopts the formulation proposed by Rex[11] and Gardonio *et al.*[2] and presented above to develop a local volume displacement sensor for plates. It is important to note again that all the work performed by other researchers in development of a plate volume displacement sensor using PVDF has, to date, been focused primarily on measurement of total structural volume displacement (instead of local structural volume displacement, as proposed here).

Figure 3 shows the orientation of the plate and piezoelectric film used in the following derivation. In order to relate the output charge of the PVDF sensor to the volume displacement of area A_2 (see Fig. 4), Eqs. (1) and (3) are modified as follows:

$$q_L^1 = -h\int_{y_1}^{y_2}\int_0^{Lx} F(x,y)[e_{31}w_{xx}+e_{32}w_{yy}]dxdy, \tag{7}$$

$$D_L = \int_{y_1}^{y_2}\int_{x_1}^{x_2} w(x,y)dxdy. \tag{8}$$

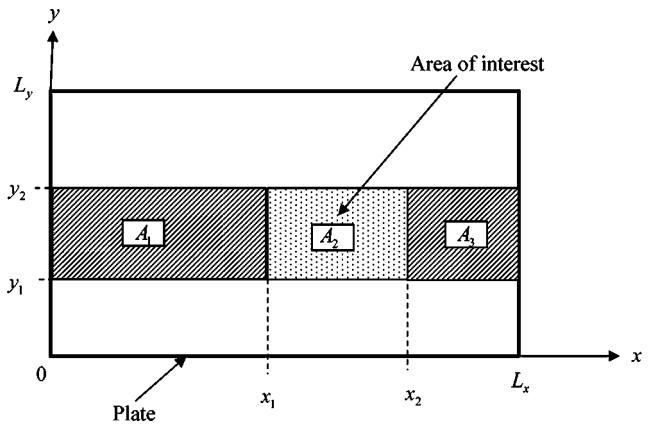


FIG. 4. Arbitrary local area considered (A_2).

The subscript L in q_L^1 and D_L refers to local sensor output charge and local volume displacement, respectively, whereas the superscript of q_L^1 is used to differentiate between the two sensors being used (discussed later). As it can be seen in Eq. (7), the sensor spans the entire length of the plate in the x direction and has a width covering only a portion of the y direction [$F(x,y)$ is independent of y]. The physical explanation for the extension of the sensor to at least one of the plate boundaries lies in the fact that the PVDF transducer is a strain sensor which is a local quantity. The goal here is to measure volume displacement as defined by an integration of plate displacement over the plate area of interest. As such, to successfully implement PVDF as a local volume displacement sensor, it has to be extended to at least one zero-displacement plate boundary. Here the boundaries where $x=0$ and $x=L_x$ have been selected.

Equation (7) can be separated into two charges q_{Lx}^1 and q_{Ly}^1 such that

$$q_L^1 = q_{Lx}^1 + q_{Ly}^1, \tag{9}$$

where

$$q_{Lx}^1 = -h\int_{y1}^{y2}\int_0^{Lx} e_{31}w_{xx}F(x,y)dxdy \tag{10a}$$

and

$$q_{Ly}^1 = -h\int_{y1}^{y2}\int_0^{Lx} e_{32}w_{yy}F(x,y)dxdy. \tag{10b}$$

q_{Lx}^1 represents the charge output due to bending in the x direction and q_{Ly}^1 represents the charge output due to bending in the y direction. The total charge output is $q_{Lx}^1+q_{Ly}^1$. Integrating by parts Eq. (10a) along L_x and Eq. (10b) along L_y yields

$$q_{Lx}^1 = -e_{31}h\int_{y_1}^{y_2}\left\{[w_xF]_0^{Lx}-[F_xw]_0^{Lx}+\int_0^{Lx}F_{xx}wdx\right\}dy, \tag{11a}$$

$$q_{Ly}^1 = -e_{32}h\int_0^{L_x}\left\{[w_yF]_{y_1}^{y_2}-[F_yw]_{y_1}^{y_2}+\int_{y_1}^{y_2}F_{yy}wdy\right\}dx. \tag{11b}$$

Before defining the shape of the sensor, we will assume that the local volume displacement sensor consists of three distinct segments respectively over A_1, A_2, and A_3 (see Fig. 4) while maintaining the continuity of sensor shape and slope at $x=x_1$ and $x=x_2$. The charge generated by such a sensor can be written as the sum of charges generated by individual segments as

$$q_{Lx}^1 = q_1^x + q_2^x + q_3^x, \tag{12a}$$

$$q_{Ly}^1 = q_1^y + q_2^y + q_3^y, \tag{12b}$$

where the subscripts 1, 2, and 3 refer to the sensor segments on A_1, A_2, and A_3 respectively. Using Eqs. (11a) and (11b), the output charge from individual sensor can be written as

$$q_1^x = -e_{31}h\int_{y_1}^{y_2}\left\{[w_xF_1]_0^{x_1}-[F_{1x}w]_0^{x_1}+\int_0^{x_1}F_{1xx}wdx\right\}dy, \tag{13a}$$

$$q_2^x = -e_{31}h\int_{y_1}^{y_2}\left\{[w_xF_2]_{x_1}^{x_2} - [F_{2x}w]_{x_1}^{x_2} + \int_{x_1}^{x_2} F_{2xx}w\,dx\right\}dy, \tag{13b}$$

$$q_3^x = -e_{31}h\int_{y_1}^{y_2}\left\{[w_xF_3]_{x_2}^{Lx} - [F_{3x}w]_{x_x}^{Lx} + \int_{x_2}^{Lx} F_{3xx}w\,dx\right\}dy, \tag{13c}$$

$$q_1^y = -e_{32}h\int_{0}^{L_x}\left\{[w_yF_1]_{y_1}^{y_2} - [F_{1y}w]_{y_1}^{y_2} + \int_{y_1}^{y_2} F_{1yy}w\,dy\right\}dx, \tag{13d}$$

$$q_2^y = -e_{32}h\int_{0}^{L_x}\left\{[w_yF_2]_{y_1}^{y_2} - [F_{2y}w]_{y_1}^{y_2} + \int_{y_1}^{y_2} F_{2yy}w\,dy\right\}dx, \tag{13e}$$

$$q_3^y = -e_{32}h\int_{0}^{L_x}\left\{[w_yF_3]_{y_1}^{y_2} - [F_{3y}w]_{y_1}^{y_2} + \int_{y_1}^{y_2} F_{3yy}w\,dy\right\}dx, \tag{13f}$$

where $F_1(x,y)$, $F_2(x,y)$, and $F_3(x,y)$ describe the sensor shape over the plate areas A_1, A_2, and A_3, respectively. Using Eqs. (13a)–(13f) and assuming a linear shape for $F_1(x,y)$ and $F_3(x,y)$ (such that $F_{1xx}=F_{3xx}=F_{1yy}=F_{3yy}=0$), along with constraining the PVDF sensor shapes to maintain the continuity of sensor shape and its slope at the segment boundaries $x=x_1$ and $x=x_2$, we have

$$F_1(x_1,y) = F_2(x_1,y), \tag{14a}$$

$$F_2(x_2,y) = F_3(x_2,y), \tag{14b}$$

$$F_{1x}(x_1,y) = F_{2x}(x_1,y), \tag{14c}$$

$$F_{2x}(x_2,y) = F_{3x}(x_2,y). \tag{14d}$$

As mentioned earlier the sensor sensitivity is assumed constant in the y direction, therefore $F_{1y}=F_{2y}=F_{3y}=0$. One can show that Eqs. (12a) and (12b) yield

$$q_{Lx}^1 = -e_{31}h\int_{y_1}^{y_2}\Big\{[w_xF_1]_0^{x_1} - [F_{1x}w]_0^{x_1} + [w_xF_2]_{x_1}^{x_2} - [F_{2x}w]_{x_1}^{x_2} + [w_xF_3]_{x_2}^{Lx} - [F_{3x}w]_{x_2}^{Lx} + \int_{x_1}^{x_2} F_{2xx}w\,dx\Big\}dy, \tag{15a}$$

$$q_{Ly}^1 = -e_{32}h\int_{0}^{L_x}\{[w_yF_1]_{y_1}^{y_2} + [w_yF_2]_{y_1}^{y_2} + [w_yF_3]_{y_1}^{y_2}\}dx. \tag{15b}$$

The objective is to have the output charge of the PVDF sensor be equal to the local volume displacement of the plate. This can be achieved by setting $D_L = q_{Lx}^1 + q_{Ly}^1$. Equation (15a) can be written as follows:

$$q_{Lx}^1 = -e_{31}h\int_{y_1}^{y_2}\{[w_xF_1]_0^{x_1} - [F_{1x}w]_0^{x_1} + [w_xF_2]_{x_1}^{x_2} - [F_{2x}w]_{x_1}^{x_2} + [w_xF_3]_{x_2}^{Lx} - [F_{3x}w]_{x_2}^{Lx}\}dy + \int_{y_1}^{y_2}\int_{x_1}^{x_2} F_{2xx}w\,dx\,dy. \tag{16}$$

If we assume a quadratic shape function for $F_2(x,y)=a_2x^2+b_2x+c_2$ such that $F_{2xx}=2a_2$, then to accurately measure the local plate volume displacement over an arbitrary area, the above equations should satisfy the following conditions:

$$\int_{y_1}^{y_2}\{[w_xF_1]_0^{x_1} - [F_{1x}w]_0^{x_1} + [w_xF_2]_{x_1}^{x_2} - [F_{2x}w]_{x_1}^{x_2} + [w_xF_3]_{x_2}^{Lx} - [F_{3x}w]_{x_2}^{Lx}\}dy = 0,$$

$$q_{Ly}^1 = -e_{32}h\int_{0}^{L_x}\{[w_yF_1]_{y_1}^{y_2} + [w_yF_2]_{y_1}^{y_2} + [w_yF_3]_{y_1}^{y_2}\}dx = 0, \tag{17}$$

$$q_{Lx}^1 = -2a_2e_{31}h\int_{y_1}^{y_2}\int_{x_1}^{x_2} w\,dx\,dy,$$

which can be simplified using Eq. (14) as follows:

$$\int_{y_1}^{y_2}\{[-w_x(0,y)F_1(0,y)] + [F_{1x}(0,y)w(0,y)] + [w_x(L_x,y)F_3(L_x,y)] - [F_{3x}(L_x,y)w(L_x,y)]\}dy = 0,$$

$$q_{Ly}^1 = -e_{32}h\int_{0}^{L_x}\{[w_yF_1]_{y_1}^{y_2} + [w_yF_2]_{y_1}^{y_2} + [w_yF_3]_{y_1}^{y_2}\}dx = 0, \tag{18}$$

$$q_{Lx}^1 = -2a_2e_{31}h\int_{y_1}^{y_2}\int_{x_1}^{x_2} w\,dx\,dy.$$

It is clear that when all the conditions are satisfied, the output charge of the sensor will be proportional to the volume displacement of the area A_2. Furthermore, the shape of the sensor will be defined by the following equations:

$$F_1(x,y) = a_1x, \quad F_2(x,y) = a_2x^2 + b_2x + c_2,$$
$$F_3(x,y) = a_3x + b_3. \tag{19}$$

The next section of this paper will finalize the design of the local volume displacement sensor by considering plates with various boundary conditions, i.e., simply supported, clamped, and arbitrary boundary conditions.

B. Boundary conditions consideration

For any plate with displacement equal to zero at the boundary, the first condition in Eq. (18) is always satisfied, that is, $F_1(0,y), w(0,y), F_3(L_x,y)$, and $w(L_x,y)$ are all equal to zero at the boundaries. However, the second condition is not satisfied. The limits y_1 and y_2 are arbitrary locations on the plate, therefore the terms in the integrant are not necessarily equal to zero, but can be eliminated by using a matched pair of local volume displacement sensors. The objective is to eliminate the sensitivity of the sensor in the y direction. These two sensors must be identical[2,3] with the exception that the piezoelectric axes are swapped such that the *1* axis of the piezoelectric corresponds to the y axis of the plate (Fig. 5). Now, based on this figure, the output charge of the second sensor can be written as

$$q_L^2 = -h\int_{y_1}^{y_2}\int_{0}^{Lx} F(x,y)[e_{32}w_{xx} + e_{31}w_{yy}]dx\,dy. \tag{20}$$

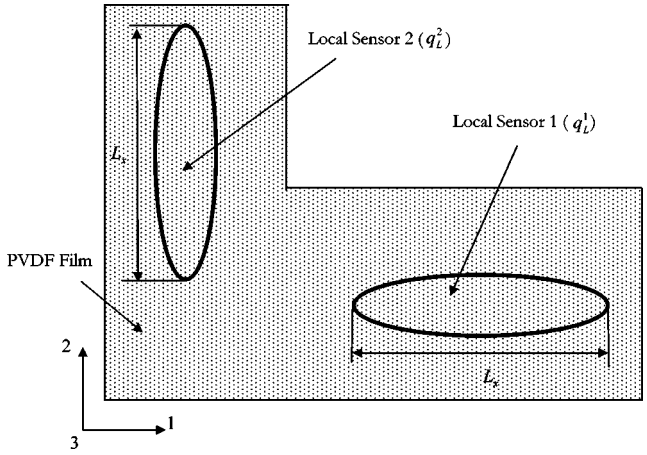


FIG. 5. Sensor orientation as cut from a sheet of PVDF film.

Notice the superscript 2 of the charge q_L^2 in order to differentiate it from the first sensor output charge q_L^1. Also the coefficient of w_{xx} changed from e_{31} to e_{32} and that of w_{yy} changed from e_{32} to e_{31}. In the above equation, w_{xx} is multiplied by e_{31}/e_{31} and w_{yy} is multiplied by e_{32}/e_{32} to produce

$$q_L^2 = -h\int_{y_1}^{y_2}\int_0^{Lx}\frac{e_{32}e_{31}}{e_{31}}w_{xx}(x,y)F(x,y)dxdy - h\int_{y_1}^{y_2}\int_0^{Lx}\frac{e_{31}e_{32}}{e_{32}}w_{yy}(x,y)F(x,y)dxdy. \tag{21}$$

Comparing Eq. (21) with Eqs. (10a) and (10b), it can then be said that

$$q_L^2 = q_{Lx}^1\frac{e_{32}}{e_{31}} + q_{Ly}^1\frac{e_{31}}{e_{32}}. \tag{22}$$

Now that we have two sensors, the total charge Q produced is the sum of the charges produced by each sensor. In other words,

$$Q = q_L^1 + q_L^2. \tag{23}$$

Using Eqs. (9) and (22),

$$Q = q_{Lx}^1\left(1+\frac{e_{32}}{e_{31}}\right) + q_{Ly}^1\left(1+\frac{e_{31}}{e_{32}}\right). \tag{24}$$

Rearranging Eq. (24) along with Eq. (22) yields

$$q_{Lx}^1 = \frac{Q - q_L^2(1+e_{32}/e_{31})}{[1-(e_{32}/e_{31})^2]}. \tag{25}$$

When the PVDF film stress/charge coefficients e_{31} and e_{32} are known, the output charge q_{Lx}^1 can be obtained by measuring the output charges q_L^1 and q_L^2 from the two sensors. And using Eq. (17), the local volume displacement of the plate can be written as

$$D_L = \frac{-e_{31}(Q - q_L^2(1+e_{32}/e_{31}))}{[(e_{31})^2-(e_{32})^2](h)(2a_2)}. \tag{26}$$

Equation (26) fulfills the purpose of this research to relate sensor charge to local volume displacement. Examination of Eq. (26) shows that the only two variables present are q_L^2 and Q, where q_L^2 refers to the charge read from the second sensor and Q refers to the total charge produced by the two sensors $(q_L^1+q_L^2)$. This means that by reading the output of the two sensors, the local volume displacement of the plate cam be determined.

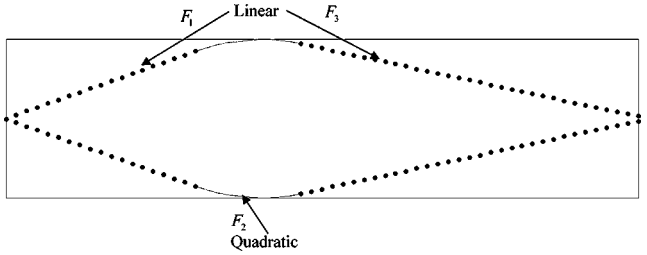


FIG. 6. Local volume displacement sensor for area A_2. The dotted line represents the linear functions F_1 and F_3, and the continuous line the quadratic function F_2.

The above described sensors only apply to plates for which the movement of the edges is equal to zero. For plates with arbitrary boundary conditions (the movement of the edges is not necessary zero), the examination of Eq. (18) shows that additional sensors will be required to measure the displacement at the boundaries but also away from the boundaries to determine the slopes. Then some type of electronic will be needed to compute the integral of the first equation of condition Eq. (18). This will result in a very complex system. Thus the sensor implementation will address the case of plates for which the displacement at the boundaries is equal to zero, mainly simply supported and clamped plates.

C. Sensor implementation

The sensor design theory presented above shows output charge for the two sensors is proportional to local volume displacement of the plate due to bending. Also the sensor is designed such that the sensitivity along the y direction is removed. Finally, it is shown that two sensors with quadratic sensitivity in the x direction are needed. In order to complete the design of the sensor, the coefficients of the shape function shown in Eq. (19) must be determined. The local area defined in Fig. 4 will be used to determine the coefficients of the shape function. From Eq. (8) and the last equation of Eq. (18) the coefficient a_2 can be written as

$$a_2 = \frac{-1}{2e_{31}h}. \tag{27}$$

The rest of the coefficients are determined using Eqs. (14) and (19). The resulting coefficients are

$$c_2 = a_2x_1^2, \quad a_3 = \frac{a_2}{Lx}(x_2^2-x_1^2), \quad b_2 = a_3 - 2a_2x_2, \quad b_3 = -a_3Lx, \quad a_1 = 2a_2x_2 + b_2. \tag{28}$$

The corresponding sensor shape is shown in Fig. 6.

Gardonio *et al.*[2] have suggested various methods to implement two-dimensional volume displacement sensors with quadratic weighting function using piezoelectric film. In this paper their concept of spatial weighting is adopted and

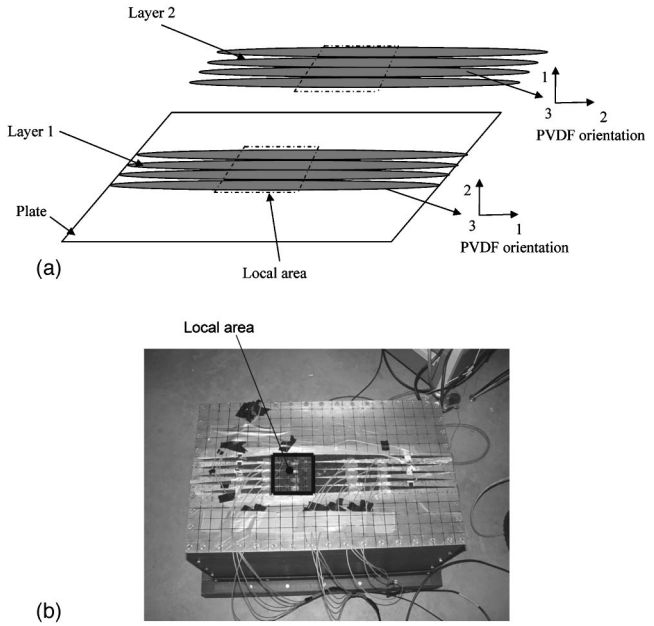


FIG. 7. (a) PVDF local volume displacement sensor implementation. (b) Experimental setup.

extended to local volume displacement implementation. Several quadratically shaped strips of PVDF are placed over the area of interest and connected in parallel as shown in Figs. 7(a) and (b).

A word needs to be said on the physical implementation of these sensors. The two sensors could be bonded on top of each other and the output charge of the sensor at the top could be compensated such that the combined plate-sensor thickness is $h=(h_p+2h_s)/2$. Alternatively, the second sensor could be bonded on the opposite side of the plate such that the output charge is multiplied by (-1) to compensate for the existence of tension where the design equation predicts compression and vice versa.

It is important to note that in order to reduce the effect of bending in the y direction, the strip must have a sufficiently small width compared to the plate structural wavelength of the highest frequency in the signal of interest. This case will most likely be satisfied in active noise application where the interest is on low frequency sound.

IV. SENSOR VERIFICATION

To verify the above conclusions about sensor development, a plate with clamped boundaries was fabricated. The plate was made of 4.76-mm-thick, 609.6-mm-long, and 381-mm-wide aluminum and clamped on each edge over 15 mm. Therefore, the surface of the plate free to vibrate had dimensions 579.6×351 mm^2. The sensors were made of PVDF strips. Ten strips, with geometry shown in Fig. 6, were cut out of a PVDF sheet. Care was taken so that the proper PVDF axis was parallel to the x axis of the plate. Each strip was 20 mm wide, 609.6 mm long, and was bonded to the plate such that its quadratically shaped section coincides with the local area of the plate with coordinates $(x_1,x_2)=(180,280)$ mm and $(y_1,y_2)=(140,240)$ mm (Fig. 4). The geometry and physical constants of the plate and the sensors

TABLE I. Geometry and physical constants for the plate.

Parameter	Value	Units
Length	609.6	mm
Width	381	mm
Thickness	4.76	mm
Mass density	2700	Kg/m^3
Young's modulus	70E9	N/m^2
Poisson ratio	0.33	...
(x_1,x_2)	(180,280)	mm
(y_1,y_2)	(140,240)	mm

are shown in Tables I and II, respectively. The value of the stress/charge coefficient e_{32} was measured using a simply supported beam setup. A volume displacement sensor for a vibrating beam was fabricated[10] such that its axis 2 was along the beam axis. The measurements from the sensor and an accelerometer were used to calculate e_{32}.

As mentioned before, two layers of PVDF strips were used in the experiment. Five PVDF strips per layer, or a total of ten strips, were bonded to the surface of the plate (Fig. 7). The volume displacements of the local area were measured using the sensors. These measurements were then validated by computing the volume displacement of the local area using accelerometer measurements.

A single accelerometer was used to measure the acceleration of a series of 144 points over the local area of the plate subjected to a unit force over a broad band of frequencies from 0 to 700 Hz. As shown by Zahui *et al.*,[10] the weight of the accelerometer used (PCB-352C23) had little or no effect on the vibration of the plate. Thus, assuming the motion to be harmonic at each frequency, the plate displacements were calculated. Next, the volume displacement of the plate local area was estimated by integrating the displacements.

Figure 8 shows a comparison of the local volume displacements obtained using the accelerometer measurements and the PVDF sensor. The PVDF measured the plate local volume displacement accurately up to 400 Hz (capturing the first two modes of the plate). The sensor accuracy is greatly diminished above 400 Hz. However, for low frequency application, such as active noise control, the sensor will accurately measure the local volume displacement of the plate. A low pass filter will be needed to cut off the high frequency content of the sensor signal.

One possible reason for the inaccuracy at high frequencies could be the fact that the sensitivity of the sensor in the y direction cannot be neglected anymore. In fact, the first two modes of the plate represent the (1,1) and (2,1) modes whereas the third is a (1,2) mode such that the bending in the

TABLE II. Physical constants for the PVDF film.

Parameter	Value	Units
Thickness	28	μm
Mass density	1780	Kg/m^3
Young's modulus	2.3E9	N/m^2
Stress/charge constants, e_{31}	0.310	V/m/N/m^2
Stress/charge constants, e_{32}	0.025	V/m/N/m^2
Stress/charge constants, e_{36}	...	...

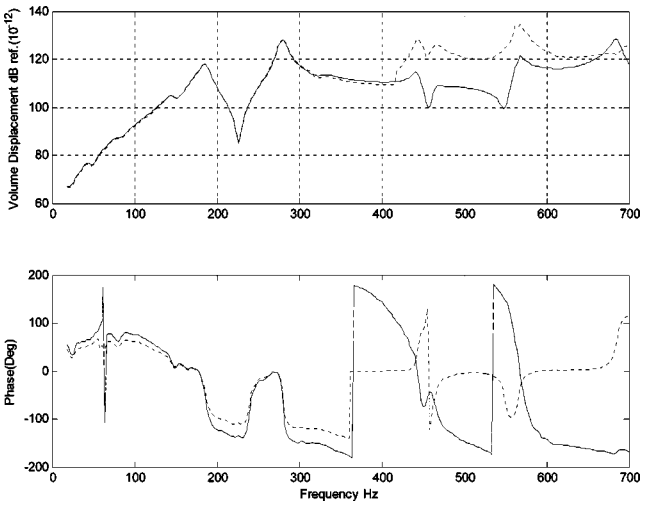


FIG. 8. Phase and magnitude of the local volume displacement of a subarea of a clamped plate measured using an accelerometer (dashed) and a PVDF sensor bonded to the plate surface (solid).

y direction is becoming important starting with the third mode. Another reason for the inaccuracy of the sensor at high frequencies has been given by Gardonio *et al.*[2] as originating from in plane vibrations which were not taken into account in Eq. (1). However, in the present experiment the effect of in-plane axial stresses are negligible as shown by Timoshenko[12] and Ugural.[13] The in-plane axial stresses can be shown to be insignificant by considering the elongation of the plate due to its curvature. In this experiment the difference between the length of the deflection curve and the length of the plate is in the order of 0.005 μm. Thus the main source of sensor error at high frequency is in the increase of sensitivity in the *y* direction at higher frequency and the experimental errors including sensor fabrication errors, boundary conditions errors, etc.

V. CONCLUSION

This paper presented a general methodology for the development of a two-dimensional local volume displacement sensor for vibrating plates using polyvynilidene fluoride (PVDF). Starting with a technique developed by others for the design of total volume displacement sensors, a modification for local volume displacement was presented.

The design yields two sensors which are quadratic in the *x* direction (i.e., the sensor has a quadratic sensitivity in the *x* direction only). The physical layout of the sensors was achieved through spatial quadratic and linear weighting. Several strips of PVDF film are laid over the area of interest. Each strip has a quadratic shape over that area of the plate whose volume displacement is measured, and it has a linear shape that connects the sensor to the boundaries. The width of these strips is significantly small compared to the structural wavelength in order to reduce the sensitivity in the *y* direction.

Next it was shown that a sensor with quadratic and linear shape could be used to accurately measure the local volume displacement of a vibrating plate with clamped or simply supported boundary condition. Finally, a local volume displacement sensor was fabricated and bonded to the clamped plate. Measurements taken using the PVDF sensor were found to be accurate below 400 Hz.

Finally, the local volume displacement sensors described in this paper present practical application issues for plates divided into more than two local areas.[10] For two local areas, one could bond the sensors to the top and bottom of the plate. As described by Naghshineh and Mason, the use of two cancellation devices will only be effective in reducing noise through the modes (1,1), (2,1), and (1,2). Thus the smart panel with two cancellation devices must be designed such that the frequency of interest falls within these modes. However, for application requiring more than two local areas, spatial multiplexing method, initially suggested by Lee and Moon[5] and detailed in Zahui *et al.*,[10] could be used.

[1] C. Guigou, F. Charette, and A. Berry, "Active control of sound by minimization of volume velocity on finite beam," Proc. of the 3rd International Congress on Air- and Structure-Borne Sound and Vibration, Montreal, Canada (1994), pp. 1507–1514.
[2] P. Gardonio, Y.-S. Lee, and S. J. Elliott, "Analysis and measurement of a matched volume velocity sensor and uniform force actuator for active structural acoustic control," velocity sensors for active control," J. Acoust. Soc. Am. **110**, 3025–3031 (2001).
[3] M. E. Johnson and S. J. Elliott, "Active control of sound radiation using volume velocity cancellation," J. Acoust. Soc. Am. **98**, 2174–2186 (1995).
[4] F. Charette, A. Berry, and C. Guigou, "Active control of sound radiation from a plate using a polyvinylidene fluoride volume displacement sensor," J. Acoust. Soc. Am. **103**, 1493–1503 (1998).
[5] C. K. Lee and F. C. Moon, "Modal sensors/actuators," Trans. ASME **57**, 434–441 (1990).
[6] V. B. Mason, K. Naghshineh, and G. K. Toth, "Broadband, wide-area active control of sound radiated from vibrating structures using local surface-mounted radiation suppression devices," Proc. of Noise-Con (1994), pp. 467–472.
[7] K. Naghshineh and V. B. Mason, "Active control of sound radiated from vibrating structures using surface-mounted radiation suppression devices," Proc. of the 3rd International Congress on Air- and Structure-Borne Sound and Vibration, Montreal, Canada (1994), pp. 1411–1419.
[8] K. Naghshineh and V. B. Mason, "Reduction of sound radiated from vibrating structures via active control of local volume velocity," J. Appl. Acoust. **47**, 27–46 (1996).
[9] M. Zahui, J. Kamman, and K. Naghshineh, "Narrow band Active Control of Sound Radiated from a Baffled Beam Using Local Volume Displacement Minimization," J. Appl. Acoust. **62**, 47–64 (2001).
[10] M. Zahui, J. Kamman, and K. Naghshineh, "Theoretical Development and Experimental Validation of Local Volume Displacement Sensors for a Vibrating Beam," J. Vibr. Acoust. **123**, 110–118 (2001).
[11] J. Rex, "The use of Integrated Transducers for the Active Control of Structural Vibration," M.Sc thesis, University of Southampton, 1991.
[12] S. Timoshenko, *Strength of Materials, Parts II, Advanced Theory and Problems* (Krieger, New York, 1958).
[13] S. Ugural, *Stresses in Plates and Shells* (McGraw–Hill, New York, 1981).

Hybrid passive–active absorption using microperforated panels

Pedro Cobo,[a] Jaime Pfretzschner, María Cuesta, and David K. Anthony
Instituto de Acústica, CSIC, Serrano 144, 28006 Madrid, Spain

(Received 9 January 2004; revised 7 June 2004; accepted 9 July 2004)

A theoretical and experimental study of a proposed hybrid passive–active plane-wave system to provide broadband acoustic absorption is presented. The passive absorber consists of a microperforated panel (MPP), used in place of conventional fibrous materials, in front of an air layer. The active system uses an active transducer (a loudspeaker), an error sensor (a microphone), and an adaptive controller. MPPs are thinner than fibrous materials and provide a better solution in hygienic environments. For two such systems, the dependence of performance on MPP parameters is studied for two control strategies: impedance matching and pressure release. The first condition is found to be better for cases where the acoustic impedance of the system approaches that of air. Otherwise, the pressure-release condition is better, and a wideband solution is the implementation of the active control system up to a frequency where the natural band of the passive system provides higher absorption. Therefore, the use of a low-pass filter in the error signal is implemented to afford hybrid passive–active wideband absorption. The control effort of active system is also considered. Experimental verification shows good agreement with theory, and an average absorption coefficient of 0.82 in the frequency bandwidth 100 to 1600 Hz was achieved. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1786831]

PACS numbers: 43.55.Ev, 43.50.Ki [KAC] Pages: 2118–2125

I. INTRODUCTION

Passive absorption has been traditionally used to improve the acoustic environments that inhabitants are exposed to daily. The weight and/or size requirements for absorption at low frequencies are often prohibitive. The more recent technology of active noise control can be used effectively at low frequencies, and therefore it is practical to conceive of hybrid systems that combine passive and active techniques to provide absorption at both low and high frequencies.

Passive absorption is often achieved using a two-layer system consisting of an air layer sandwiched between a porous or fibrous layer and a rigid wall. Depending on the constitutive properties of the porous layer and the thickness of the porous and air layers, such a system is capable of providing high absorption at medium and high frequencies, with maxima at those frequencies for which the air-layer thickness is an odd-integer multiple of the quarter-wavelength. Guicking and Lorenz (1984) first described the active equivalent of this $\lambda/4$ *resonance absorber* by replacing the rigid wall with a loudspeaker driven such that zero sound pressure is achieved at a microphone positioned on the rear face of the porous layer. They reported absorption coefficients in the range 0.6–0.7 over the frequency band 200–500 Hz using an analog feedback controller. They needed a second microphone in front of the porous layer due to the use of analog circuitry. Furtoss *et al.* (1997) proposed the alternative strategy of minimizing the pressure at only one microphone at the rear face of the porous layer (*pressure release* condition), thus resulting in a simpler active control system. This hybrid passive–active method controlled the input impedance to the system so as to maximize the low-frequency absorption of the porous layer.

Beyene and Burdisso (1997) proposed canceling the reflected component of the standing wave in the air layer to further increase the active absorption of the hybrid passive–active system (*impedance matching* condition). The implementation of such a condition, as described first by Guicking and Karcher (1984), required two closely spaced microphones together with a deconvolution circuit to separate both the incident and reflected plane waves in the air cavity. Using such a hybrid passive–active system, they reported a high absorption coefficient in the range of 0.8–1.0 in the frequency band 100–2000 Hz. Cobo *et al.* (2003) carried out a theoretical and experimental comparison of both conditions (pressure release and impedance matching), concluding that the pressure-release condition is capable of providing higher values of absorption than the impedance-matching condition when the flow resistance of the porous layer equals the acoustic impedance of air. On the other hand, the impedance-matching condition yields more absorption when the flow resistance of the porous layer is smaller than about 0.7 times the acoustic resistance of air. Using a 10-cm-thick two-layer system (3-cm-thick porous layer plus a 7-cm air cavity), they obtained an absorption coefficient higher than 0.95 at frequencies between 90 and 850 Hz using the pressure-release condition.

The total thickness of the hybrid system can be further decreased by using microperforated panels (MPP) as the absorbent material. An MPP consists of a thin sheet panel perforated with many orifices of submillimeter size distributed along its surface (Maa, 1987, 1998). Such MPPs have high acoustic resistance and low acoustic mass reactance necessary for broadband sound absorption without further additional porous material. Therefore, this absorber has remark-

[a]Electronic mail: iacpc24@ia.cetef.csic.es

able practical benefits since it is lightweight and, with an appropriate material selection, can be inexpensive. It also produces fewer health-related concerns than fibrous materials (Kang and Fuchs, 1999). Furthermore, the MPP constitutes the exposed surface of the absorptive device, as it can be made of many kinds of materials (fine sheets of metal, plastic, methacrylate, or glass, for example).

The main objective of this paper is to analyze the capabilities of a hybrid passive–active system to provide wideband absorption using MPP as the passive absorbent. Preliminary experiments with such a system showed promising results (Pfretzschner *et al.*, 2003). Following the procedure reported in a previous paper by Cobo *et al.* (2003), a model of propagating plane waves in a tube allows the absorption coefficient to be evaluated using both the pressure-release and the impedance-matching conditions. This model will also permit the comparison of the control effort required by each control condition. In Sec. II the equations of the microperforated panels (MPP) are reviewed (Maa, 1987, 1998). The control conditions of a hybrid system using an MPP, as well as the control effort, are analyzed in Sec. III. The description of the experimental setup and the experimental results are given in Sec. IV.

II. MICROPERFORATED PANELS

MPP absorbers consist of panels perforated with submillimeter orifices in front of an air cavity. Appropriately choosing the parameters of the MPP, absorption in certain frequency bands can be obtained using cavities of moderate thickness without the need of additional fibrous materials. The absence of conventional fibrous materials allows the design of ''clean'' absorbent systems. In hospitals, food, pharmaceutical, and microelectronics industries, MPP are better suited than absorbent fibrous materials for silencers in ventilation, heating, and air-conditioning systems (Wu, 1997).

An MPP can be made with distribution of narrow short tubes with an interhole distance that is small compared to the wavelength of the sound to be absorbed. The propagation through one of these tubes was first studied by Lord Rayleigh and later by Crandall (see Maa, 1987, 1998). Appropriately modifying the Crandall solution to include the end correction, Maa (1987) derived the following equation for the specific acoustic impedance of the MPP when normalized by the air characteristic acoustic impedance, Z_0:

$$z_m=\frac{Z_m}{Z_0}=j\frac{\omega\rho t}{pZ_0}\left[1-\frac{2J_1(x\sqrt{-j})}{x\sqrt{-j}J_0(x\sqrt{-j})}\right]^{-1}+\frac{\sqrt{2}\eta x}{pZ_0d}+j\frac{0.85\omega\rho d}{pZ_0}, \tag{1}$$

where η is the coefficient of viscosity of the air (1.789×10^{-5} kg/m·s), ω is the angular frequency, ρ (1.21 kg/m^3) is the density of the air, t is the panel thickness, d is the orifice diameter, p is the perforation ratio of the total surface (i.e., perforated and unperforated) area of the MPP, and J_0 and J_1 are Bessel functions of the first kind of orders zero and one, respectively. The *perforate* constant, x, is defined as

$$x=\frac{d}{\sqrt{\frac{4\eta}{\rho\omega}}}. \tag{2}$$

This is the ratio of the orifice diameter, d, to the viscous boundary layer thickness of the air in the orifice. In the following, normalized impedances will always be with respect to Z_0.

For a wide range of x values, the MPP has a specific acoustic resistance close to that of the air, with a slightly smaller specific mass reactance. In order to obtain good absorption, the magnitude of the reactance component must be much smaller than the air resistance. This can be achieved by adding an air cavity of thickness D between the MPP and the back wall. The normalized specific acoustic impedance of the air cavity is

$$z_2^p=\frac{Z_2}{Z_0}=-j\cot(kD). \tag{3}$$

(Note that z_2^p is used here to distinguish the generalized cavity impedance z_2 in cases only applicable to the passive case.) The normalized specific input acoustic impedance of the complete system (denoted MPP+cavity), z_1, is the series specific acoustic impedance of the MPP, z_m, and cavity, z_2 (Dupont *et al.*, 2003)

$$z_1=z_m+z_2, \tag{4}$$

where $z_2=z_2^p$ here for the consideration of the passive case. The coefficient of reflection of the whole system is

$$R=\frac{Z_1-Z_0}{Z_1+Z_0}=\frac{z_1-1}{z_1+1}. \tag{5}$$

Finally, the absorption coefficient will be

$$\alpha=1-|R|^2. \tag{6}$$

Equations (1)–(6) allow the absorption coefficient of the MPP+cavity system to be calculated as a function of the acoustic parameter f, the MPP parameters d, t, p, and x, and the cavity parameter D. Mechel (1988) showed the convenience of plotting the absorption coefficient as a function of nondimensional combinations of the design parameters (which he called *design charts*). The first nondimensional variable used here is the perforate constant, x, which is proportional to $d/\sqrt{\lambda}$. Assuming $d=t$, and considering Eqs. (1) and (2), z_m can be defined as

$$z_m=\frac{\eta}{pZ_0d}\times\left\{\sqrt{2}x+j4x^2\left[0.85+\left[1-\frac{2J_1(x\sqrt{-j})}{x\sqrt{-j}J_0(x\sqrt{-j})}\right]^{-1}\right]\right\}. \tag{7}$$

Defining the second nondimensional variable as

$$s=\frac{\eta}{Z_0d}, \tag{8}$$

Eq. (7) becomes

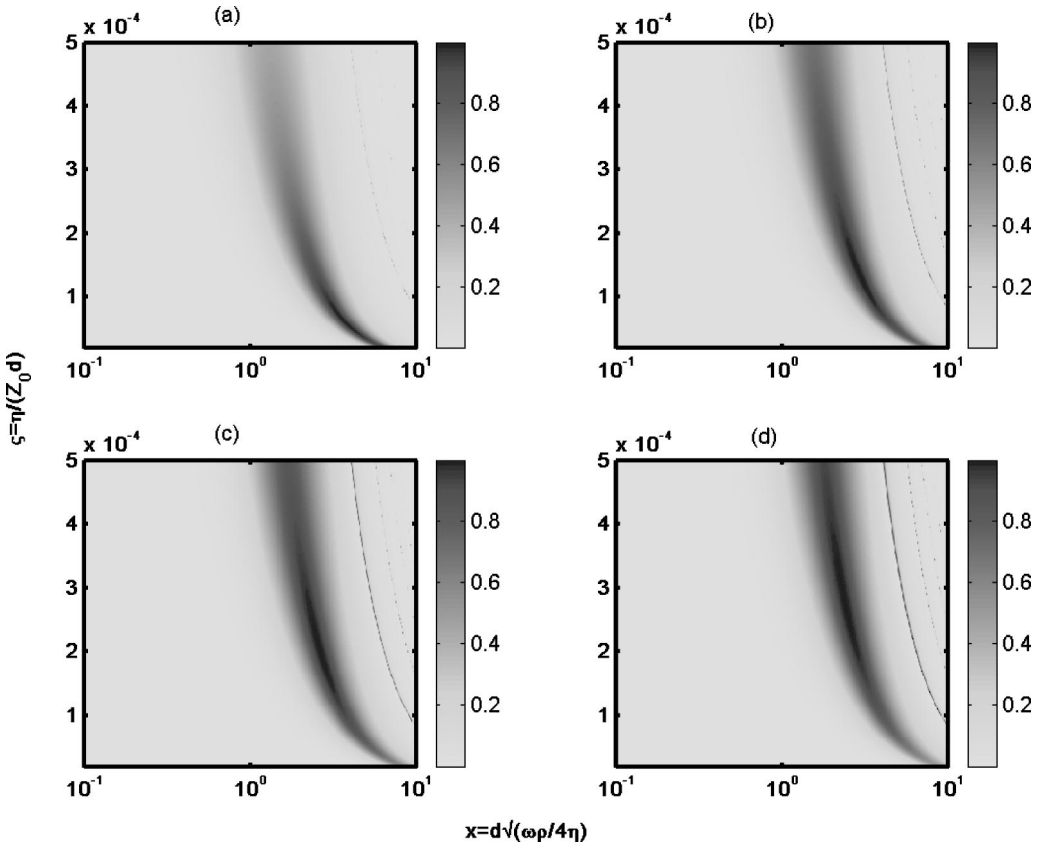


FIG. 1. Absorption charts of an MPP with different values of perforation ratio, p, for cavity thickness–orifice diameter ratio $h=100$. (a) $p=0.0025$; (b) $p=0.005$; (c) $p=0.0075$; and (d) $p=0.01$.

$$z_m=\frac{\varsigma}{p}\left\{\sqrt{2}x+j4x^2\left[0.85+\left[1-\frac{2J_1(x\sqrt{-j})}{x\sqrt{-j}J_0(x\sqrt{-j})}\right]^{-1}\right]\right\}. \tag{9}$$

The normalized specific acoustic impedance of the cavity will be

$$z_2^P=\frac{Z_2}{Z_0}=-j\cot\left(\frac{\omega\rho D}{Z_0}\right)=-j\cot\left(\frac{4\eta Dx^2}{Z_0d^2}\right). \tag{10}$$

And, defining $h=D/d$, the ratio of the cavity thickness to the orifice diameter, this impedance can be expressed as

$$z_2^P=-j\cot(4h\varsigma x^2). \tag{11}$$

Equations (4)–(7) and (11) allow design absorption charts of an MPP to be constructed as a function of the nondimensional variables x and ς, with p and h as parameters. Figure 1 illustrates the effect of the perforation ratio, p, in the absorption chart, for constant $h=D/d$. When the perforation ratio increases, the region of maximum absorption moves upwards (corresponding to lower values of d) and slightly to the left (lower frequencies). Figure 2 shows the effect of the parameter h in the absorption chart for a constant perforation ratio. As h increases, the absorption contours move to the left and the ''band'' of high absorption is widened.

Design charts illustrate well the effect of the active control in the low-frequency region of the MPP absorbers. The reconstruction of the frequency curve of sound absorption for an individual MPP absorber can be carried out easily from design charts. First, the desired absorption curve can be chosen from the corresponding design chart, as a compromise between maximum absorption at resonance and bandwidth. This will determine the values of the orifice diameter, d, the panel thickness, t (chosen commonly to be equal to d), the cavity thickness, D, and the perforation ratio, p. Once d is fixed, the variable x of the absorption curve can be converted to frequency. For the procedure to calculate the MPP parameters (d,t,p,D) starting from the desired frequency characteristics (resonance frequency and bandwidth), the reader is referred to Maa (1998).

III. HYBRID PASSIVE–ACTIVE ABSORPTION

Figure 3 shows the scheme of a hybrid passive–active system using an MPP as the passive absorber. The active component of the hybrid system is a piston in place of the back wall that has velocity V_p. A primary source on the left radiates plane waves downstream. These incident plane waves have amplitude A_i, and the reflected plane waves have amplitude B_i, where i identifies the wave components on each side of the absorbent material. The input-normalized specific acoustic impedance of the system, z_1, is still governed by the combination of z_m and z_2, as defined in Eq. (4). Assuming a time dependence $e^{j\omega t}$, the normalized specific acoustic impedance of the cavity becomes

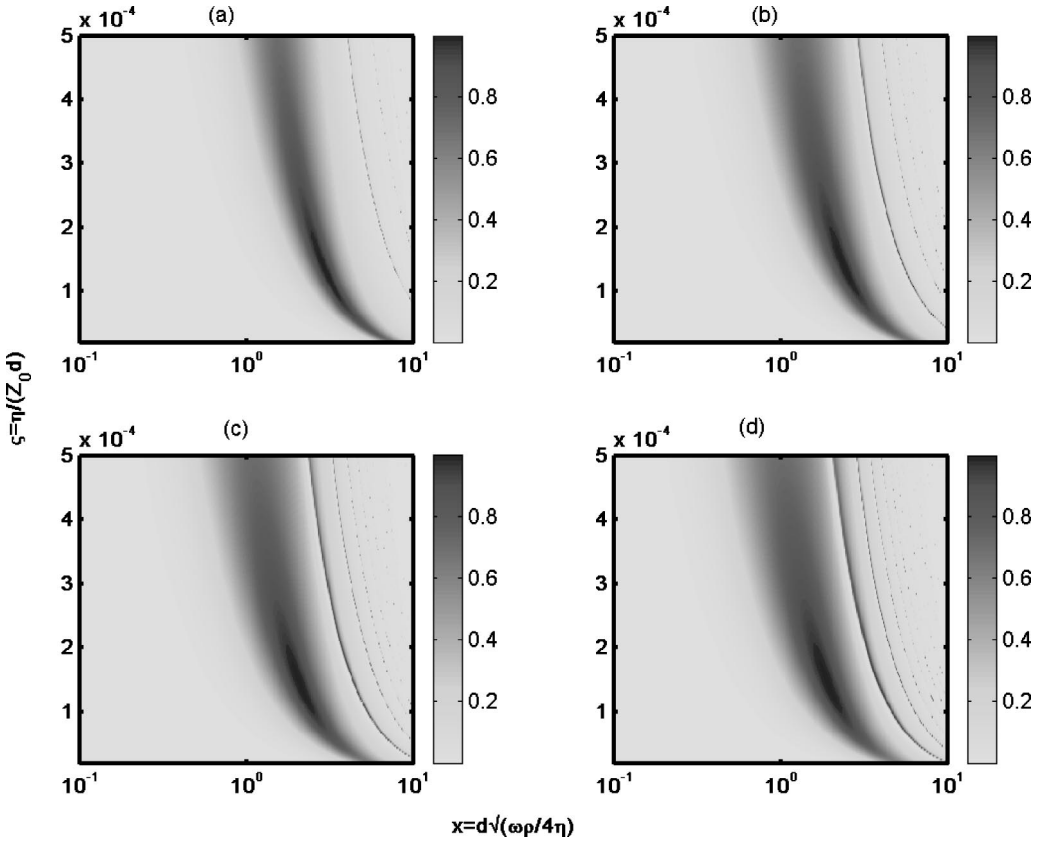


FIG. 2. Absorption charts for an MPP with different values of cavity thickness–orifice diameter ratio, h, for perforation ratio $p=0.005$. (a) $h=100$; (b) $h=200$; (c) $h=300$; and (d) $h=400$.

$$z_2=\frac{P_2}{Z_0V_2}=\frac{A_2+B_2}{A_2-B_2}, \tag{12}$$

where P_2 and V_2 are the acoustic pressure and particle velocity, respectively, at the cavity side of the MPP. From the continuity of the particle velocity at the interface $y=D$

$$A_2e^{-jkD}-B_2e^{jkD}=Z_0V_p, \tag{13a}$$

or

$$A_2=(Z_0V_p+B_2e^{jkD})e^{jkD}. \tag{13b}$$

Replacing this value of A_2 in Eq. (12), the normalized specific acoustic impedance at the cavity side of the MPP becomes

$$z_2=\frac{Z_2}{Z_0}=\frac{Z_0V_p+2B_2\cos(kD)}{Z_0V_p+2jB_2\sin(kD)}. \tag{14}$$

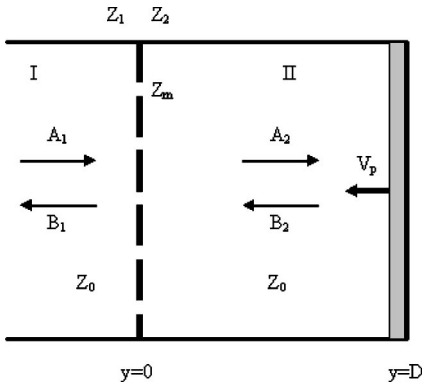


FIG. 3. Scheme of the hybrid passive–active absorber, using an MPP and a piston for the passive and active parts.

Using Eq. (14), the following control conditions can be examined:

(i) In the passive case, $V_p=0$ and z_2 in Eq. (14) becomes equal to z_2^p in Eq. (3).
(ii) In the active case, using the pressure-release condition, $P_2=0$, and $z_2=0$.
(iii) In the active case, using the impedance-matching condition, $B_2=0$, and $z_2=1$.

Figures 4, 5, and 6 show the absorption charts of an MPP with $(p,h)=(0.005,100)$ in the passive, active by pressure-release, and active by impedance-matching cases, respectively. It can be seen that both active control conditions provide absorption at low frequencies (corresponding to small values of the perforate constant x). The maximum value of absorption for the pressure-release case occurs for values of ς similar to those for the passive case. Nevertheless, the impedance-matching condition provides the maximum absorption for lower values of ς. This difference in behavior can be explained by the fact that the active control using the pressure-release condition drives the specific acoustic impedance to zero on the cavity side of the MPP. Thus, under this condition the simulated impedance is that of the MPP against a non-naturally occurring *virtual surface* with a reflection coefficient of -1.

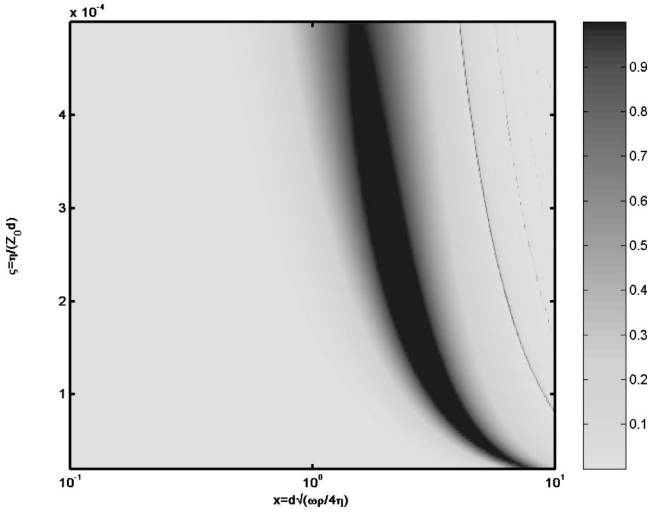


FIG. 4. Passive absorption chart of an MPP with values $(p,h)=(0.005,100)$.

On the other hand, using the impedance-matching condition, the specific acoustic impedance on the cavity side of the MPP is adapted to Z_0. Therefore, the lower the MPP impedance, Z_m, the better the impedance of the MPP +cavity system matches that of air, that is, the larger the diameter of the perforations, the better the performance of the hybrid absorber using the impedance-matching condition.

Two hybrid passive–active systems (system A and system B) are first modeled in order to study the dependence of the system performance on the four parameters that define the system. The values of these parameters for the two systems are shown in Table I. System A has the thinner MPP with smaller orifice size and perforation ratio, and the deeper cavity, compared to system B. The predicted performance of system A is shown in Fig. 7(a). It is designed to provide maximum passive absorption at the frequency band around 800 Hz. Since the cotangent function in Eq. (10) is many branched, the absorption band shown in Fig. 7(a) repeats itself when the argument is increased by multiples of π

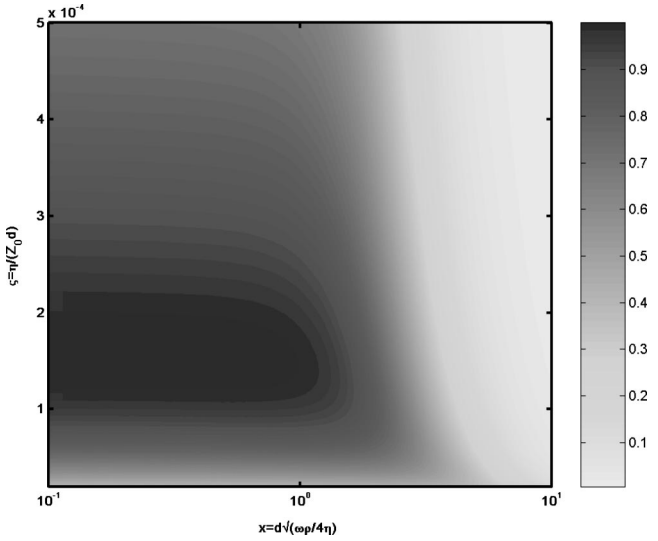


FIG. 5. Active pressure-release absorption chart of an MPP with values $(p,h)=(0.005,100)$.

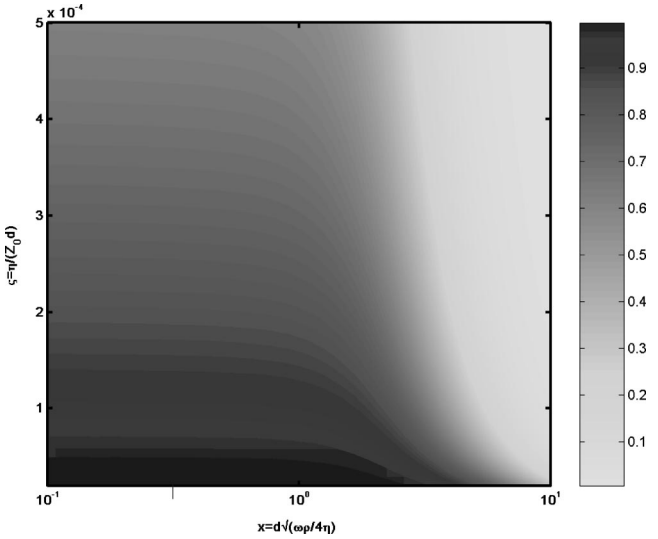


FIG. 6. Active impedance-matching absorption chart of an MPP with values $(p,h)=(0.005,100)$.

(Maa, 1998). It is seen that an active system using the pressure-release condition should be able to provide values of absorption close to 90% at frequencies below 500 Hz. As suggested in Fig. 7(a), active control again outperforms passive above 1300 Hz. However, as discussed by Maa (1998), the gaps between absorption bands of the passive curve would be filled up in diffuse field provided that each band is broad enough. Furthermore, due to the inherent problems of running active control at higher frequencies, the operation of the active system should be backed off at the lowest frequency where the active and passive curves coincide.

The predicted performance of system B is shown in Fig. 7(b). This system is not capable of providing significant levels of passive or pressure-release active absorption. However, active control using the impedance-matching condition also delivers high absorption in the low-frequency band which gradually rolls off with frequency. This case is similar to that studied by Guicking and Karcher (1984), who canceled the reflected wave in front of the active system, without an absorbent layer. The case discussed here has the advantage with respect to that of Guicking and Karcher of protecting the sensor–actuator pair from hostile environmental conditions by the use of a perforated sheet.

It is noted that practical embodiments of the impedance-matching approach would typically be deeper than the pressure-release approach due to the required microphone spacing to accurately decompose the cavity acoustic field at low frequencies.

The control effort is an important practical consideration when using an active control system. This is proportional to

TABLE I. Parameters defining the three hybrid passive–active absorption systems considered.

Parameter		System A	System B	System C
Orifice diameter (m)	d	0.14×10^{-3}	3.0×10^{-3}	0.13×10^{-3}
MPP thickness (m)	t	0.14×10^{-3}	3.0×10^{-3}	0.13×10^{-3}
Perforation ratio	p	0.0053	0.1000	0.0050
Air cavity thickness (m)	D	0.064	0.050	0.045

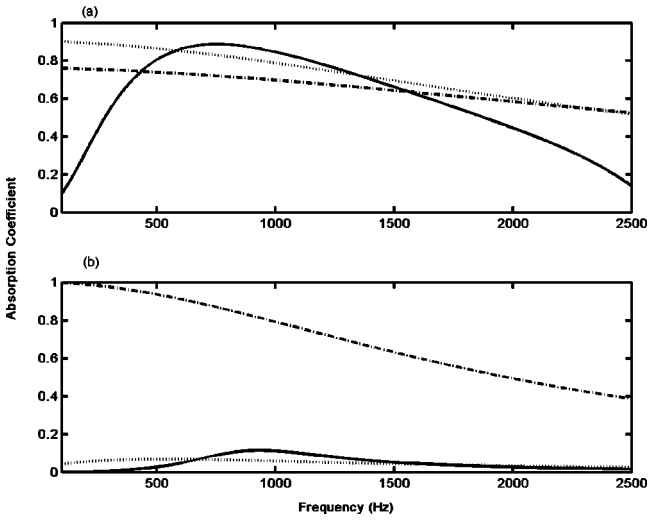


FIG. 7. Passive absorption (———), active absorption by pressure release (---) and impedance matching (-·-) conditions, of hybrid passive–active systems A (a) and B (b).

the squared magnitude of the control signal (Elliott, 2001), which in this case is the actuator velocity. Assuming a time dependence $e^{j\omega t}$, the sound pressure and the particle velocity on each side of the MPP (see Fig. 3) are given by

$$P_I(y)=A_1e^{-jky}+B_1e^{jky}$$
$$V_I(y)=\frac{A_1e^{-jky}-B_1e^{jky}}{Z_0}, \tag{15}$$

and

$$P_{II}(y)=A_2e^{-jky}+B_2e^{jky}$$
$$V_{II}(y)=\frac{A_2e^{-jky}-B_2e^{jky}}{Z_0}. \tag{16}$$

Due to the large wavelengths involved in this model, the cross section along the tube (including also the MPP) is assumed to be uniform. Thus, this model inherently neglects any evanescent waves near nonuniformities along the tube, and therefore, point pressures and particle velocities are employed in Eqs. (15)–(16).

Recalling that the condition $t=d$ was assumed above, the MPP can be considered thin enough so that the condition of continuity of the particle velocity at the interface $y=0$ can be applied

$$A_1-B_1=A_2-B_2. \tag{17}$$

The pressure loss through the orifices is due to the impedance of the MPP. Therefore

$$P_I-P_{II}=Z_mV_{II}, \tag{18a}$$

or more explicitly

$$A_1+B_1-(A_2+B_2)=\frac{Z_m}{Z_0}(A_2-B_2)=z_m(A_2-B_2). \tag{18b}$$

From the particle velocity continuity condition on the interface $y=D$

$$A_2e^{-jkD}-B_2e^{jkD}=Z_0V_p. \tag{19}$$

Equations (17)–(19) can also be written in matrix form

$$\begin{pmatrix} 1 & 1 & -1 \\ -1 & (1+z_m) & (1-z_m) \\ 0 & e^{-jkD} & -e^{jkD} \end{pmatrix}\begin{pmatrix} B_1 \\ A_2 \\ B_2 \end{pmatrix}=\begin{pmatrix} A_1 \\ A_1 \\ Z_0V_p \end{pmatrix}. \tag{20}$$

The solution of this equation system by the Cramer method affords

$$B_1=\frac{A_1[z_m\sin(kD)+e^{-jkD}]+Z_0V_p}{z_m\sin(kD)+e^{jkD}}, \tag{21}$$

$$A_2=\frac{A_1e^{jkD}-\frac{z_m}{2}Z_0V_p}{z_m\cos(kD)+e^{jkD}}, \tag{22}$$

and

$$B_2=\frac{A_1e^{-jkD}-\left(1+\frac{z_m}{2}\right)Z_0V_p}{z_m\cos(kD)+e^{jkD}}. \tag{23}$$

Thus, from the condition of pressure release $(P_{II})_{y=0}=A_2+B_2=0$

$$V_p=\frac{A_1\cos(kD)}{(1+z_m)Z_0}, \tag{24}$$

whereas in the condition of impedance-matching $(B_2=0)$

$$V_p=\frac{A_1e^{-jkD}}{\left(1+\frac{z_m}{2}\right)Z_0}. \tag{25}$$

Taking into account that the amplitude particle velocity of the incident wave is A_1/Z_0, the actuator velocity normalized by the particle velocity of the input wave is

$$\frac{V_p}{A_1/Z_0}=\begin{cases}\dfrac{\cos(kD)}{1+z_m} & \text{pressure-release condition} \\ \dfrac{e^{-jkD}}{1+\frac{z_m}{2}} & \text{impedance-matching condition.}\end{cases} \tag{26}$$

It is noted that

(i) In both control conditions the actuator velocity is always smaller than, or equal to, the particle velocity of incident wave to the MPP. In other words, the magnitude of the expression in Eq. (26) is always smaller than, or equal to, unity.

(ii) At low frequencies, the actuator velocity tends to the particle velocity of the incident wave when the impedance of the MPP tends to zero.

(iii) Under the pressure-release condition, zero pressure naturally occurs when a node aligns with the MPP, and so the actuator does not have to respond to these frequencies where one-quarter of the wavelength (or odd multiple) equals the cavity depth.

Figures 8(a) and (b) show the control effort for systems A and B as a function of the frequency, using the pressure-

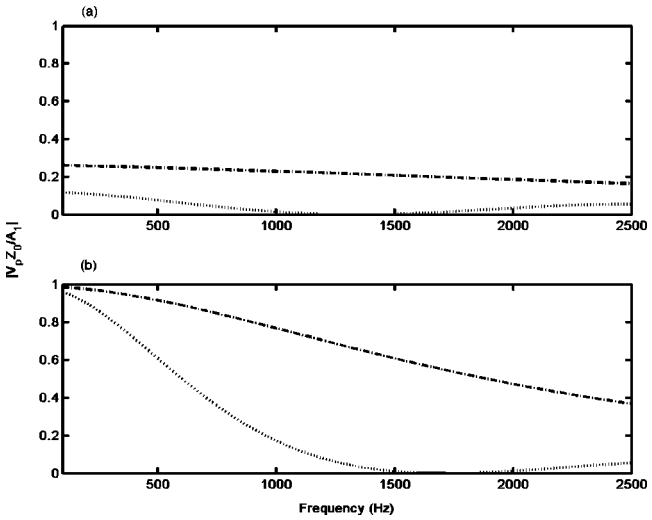


FIG. 8. Control effort for the active system using pressure release (---) and impedance matching (-·-) conditions, of hybrid passive–active systems A (a) and B (b).

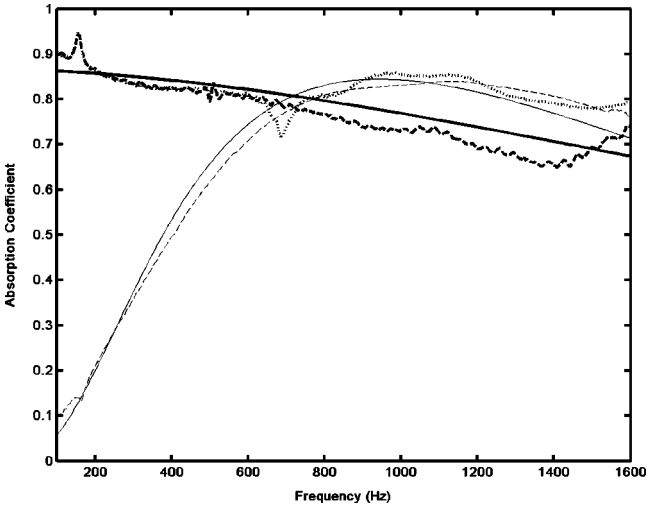


FIG. 9. Absorption coefficient against frequency of the experimental hybrid passive–active system C using the pressure-release control condition. Key: (———) passive theoretical; (---) passive experimental; (———) active theoretical; (- -) active experimental; (- - -) hybrid passive–active experimental.

release and impedance-matching conditions. For system A, the pressure-release condition provides more active absorption than the impedance-matching condition and requires approximately half the control effort throughout the whole frequency band of interest. For system B, the impedance-matching condition yields values of absorption higher than 80% up to about 1000 Hz [Fig. 7(b)], but the actuator must move with a velocity closer to the particle velocity of the incident wave and therefore requires a higher level of control effort.

IV. EXPERIMENTAL SYSTEM AND RESULTS

An experimental system of a hybrid passive–active system with parameters similar to those for system A, described in Sec. III, was implemented. The exact parameters for MPP and the air cavity of the system are shown in Table I as system C. The pressure-release condition was used for the active control element as this was predicted to provide better performance than impedance matching for such a system. All absorption measurements were carried out in a standard impedance tube with the same setup described by Cobo *et al.* (2003). The primary broadband noise was generated by a loudspeaker located at one end of the tube. The MPP absorber was located at the other end. The active system consisted of a secondary loudspeaker with a flat diaphragm, an error sensor (a 1/4-in. microphone) located just behind the MPP, and a digital controller. The absorption coefficient was measured by means of the transfer function method using two microphones close to the center of the tube, separated by either 9 or 18 cm (Cobo *et al.*, 2003). The dimensions of the tube (10-cm diameter and 1-m length) and the separation of the two measurement microphones limit the valid frequency bandwidth from 95 to 1715 Hz (ISO 10534-2).

First, the passive absorption was measured by placing the MPP in a custom holder ensuring that it is plane parallel and the required distance D from the rigid end cap at the end of the tube. D was set at 4.5 cm. Furthermore, care was taken to avoid bending of the flexible MPP.

Next, the active system was implemented. In this case, the loudspeaker replaced the end cap. The air cavity thickness (4.5 cm) was measured from the back of the MPP to the surface of the flat diaphragm of the loudspeaker. A commercial configurable controller, based on the Texas Instruments TMS320 C40 DSP, was used. It implements both the filtered-x LMS (adaptive FIR) and filtered-u LMS (adaptive IIR) algorithms, and includes also bandpass filters for the reference and error signals. The user can configure the number of taps of the control and identification filters, the convergence factor, and the cutoff frequencies of the bandpass filters. The reference signal fed to the controller was the same white-noise signal used to drive the primary source. The optimum FIR filter was generated using the filtered-x LMS algorithm so as to achieve the maximum reduction of the square pressure at the error microphone. The number of filter taps and the convergence factor were also adjusted to obtain the maximum error signal reduction. The resulting absorption coefficient was then measured for the combined passive–active system. Figure 9 shows the comparison of the performance of the passive, active, and hybrid passive–active systems with that predicted by the model for system C for frequencies from 100 to 1600 Hz. Adaptive filters with 50 taps each were used for both control and identification filters. In general, good agreement is found. The experimental system provides a slightly lower passive absorption coefficient than that predicted by the model of system C at frequencies between 250 and 1080 Hz. On the other hand, the experimental curve exceeds the theoretical one below 250 Hz and above 1090 Hz. This is thought to be mainly due to the small differences between the model and experimental parameters. In practice, it is difficult to realize an exact match due to the fact that these parameters are indirectly specified from other parameters in the model.

With the active control system operational, slightly less absorption than that predicted by the model for system C is achieved in the frequency range from 200 to 1500 Hz. Below

200 Hz and above 1500 Hz the absorption measured surpasses the theoretical one. In addition to the reasons stated above, this is also probably due to the lack of perfect cancellation of the pressure at the microphone in practice. It is seen that in both the predicted and experimental performance of the active system a very good value of absorption is achieved which slowly falls with frequency. However, above a certain frequency (about 720 Hz in this case) the passive system performs better. Thus, a hybrid system could be implemented that incorporates a high-order low-pass filter in the error signal of the active system so that the system effectively operates passively above this frequency. In this case, the error signal was low-pass filtered with a 14th-order Butterworth filter. The hybrid passive–active system achieves an average absorption coefficient of 0.82 in the frequency bandwidth 100 to 1600 Hz.

V. CONCLUSIONS

This paper has reported an investigation of the feasibility of hybrid passive–active absorbers to provide broad frequency band absorption. In this type of absorber, a passive layer is placed in front of an air layer backed by an actively controlled actuator. A microperforated panel (MPP) was used in place of conventional fibrous materials for the passive layer. MPPs have advantages in being preferable in applications where hygiene is an issue, and allow compact hybrid absorbers to be constructed. For the actively controlled absorption two control conditions were tested: pressure release and impedance matching. The first acts to cancel the pressure on the rear surface of the MPP, which is equivalent to a boundary condition with a (non-naturally occurring) reflection coefficient of -1. The second uses two closely spaced microphones in the air cavity to separate the reflected and incident acoustic plane waves, and acts to minimize the reflected component.

Using a plane-wave propagation model, the performance of two systems with different MPP parameters was predicted for the hybrid passive–active system. It was found that the choice of the best control condition depended on the impedance of the absorber system, and when this approaches that of air, the impedance-matching condition is better. The control effort of the systems using each condition was also studied. The control effort for the impedance-matching condition was found to be higher than that for the pressure-release condition. The pressure-release condition requires zero control effort when the quarter-wavelength (or odd multiple) corresponds to the cavity thickness. Moreover, the impedance matching condition would inherently require deeper air cavities than the pressure-release approach for good low-frequency performance.

Even though the actively controlled system achieves high values of absorption across the frequency range considered, above a cutoff frequency the passive element itself affords higher absorption. Thus, a hybrid solution is suggested using a low-pass filter in the error signal of the control system, so that the active control system effectively ceases to function beyond a frequency where the passive absorption alone performs better.

Experimental results were then finally presented for a system similar to the theoretical one studied using the pressure-release control condition. Good agreement was found between the predicted and experimental results. The feasibility of designing hybrid passive–active absorbers with high absorption and reduced size, as compared with the conventional porous materials, has been demonstrated.

ACKNOWLEDGMENTS

We are grateful to the Spanish Ministry of Science and Technology for funding this research (Grant No. DPI2001-1613-C02-01). We are also grateful for the comprehensive and detailed suggestions and comments of the reviewers, which have improved the quality of this paper.

Beyene, S., and Burdisso, R. A. (**1997**). ''A new hybrid passive/active noise absorption system,'' J. Acoust. Soc. Am. **101**, 1512–1515.

Cobo, P., Fernández, A., and Doutres, O. (**2003**). ''Low frequency absorption using a two-layer system with active control of input impedance,'' J. Acoust. Soc. Am. **114**, 3211–3216.

Dupont, T., Pavic, G., and Laulagnet, B. (**2003**). ''Acoustic properties of lightweight micro-perforated plate systems,'' Acust. Acta Acust. **89**, 201–212.

Elliott, S. J. (**2001**). *Signal Processing for Active Control* (Academic, London), Chap. 4, p. 184.

Furstoss, M., Thenail, D., and Galland, M. A. (**1997**). ''Surface impedance control for sound absorption: Direct and hybrid passive/active strategies,'' J. Sound Vib. **203**, 219–236.

Guicking, D., and Lorenz, E. (**1984**). ''An active sound absorber with porous plate,'' J. Vib. Acoust. Stress and Reliab. Des. **106**, 389–392.

Guicking, D., and Karcher, K. (**1984**). ''Active impedance control for one-dimensional sound,'' J. Vib. Acoust. Stress and Reliab. Des. **106**, 393–396.

Kang, J., and Fuchs, H. V. (**1999**). ''Predicting the absorption of open weave textiles and microperforated membranes backed by an air space,'' J. Sound Vib. **220**, 905–920.

Maa, D. Y. (**1987**). ''Microperforated-panel wideband absorbers,'' Noise Control Eng. J. **29**, 77–84.

Maa, D. Y. (**1998**). ''Potential of microperforated panel absorber,'' J. Acoust. Soc. Am. **104**, 2861–2866.

Mechel, F. P. (**1988**). ''Design charts for sound absorber layers,'' J. Acoust. Soc. Am. **83**, 1002–1013.

Pfretzschner, J., Cobo, P., Bote, D., Fernandez, A., and Doutres, O. (**2003**). ''Active control and microperforated panels,'' Proc. Euronoise (Naples).

Wu, M. Q. (**1997**). ''Microperforated panels for duct silencing,'' Noise Control Eng. J. **45**, 69–77.

Acoustic imaging in enclosed spaces: Analysis of room geometry modifications on the impulse response

M. Kuster,[a] D. de Vries, E. M. Hulsebos, and A. Gisolf
Laboratory of Acoustic Imaging and Sound Control, Delft University of Technology, 2600 GA Delft, The Netherlands

(Received 28 November 2003; revised 19 May 2004; accepted 2 July 2004)

Sound propagation in enclosed spaces is characterized by reflections at the boundaries of the enclosure. Reflections can be wanted in the case when they support the direct sound or give a feeling of envelopment or they can be unwanted when they lead to echoes and colouration. When measuring multiple impulse responses in an enclosed space along an array the reflections can be mapped to the reflecting objects. Similar to seismic exploration, medical diagnostics, and underwater acoustics, an image of the reflecting objects is obtained in terms of reflected energy. The imaging process is based on inverse wave field extrapolation with the Kirchhoff–Helmholtz and Rayleigh integrals. The inverse of the imaging process recreates the measured impulse responses from the image and it allows one to remove or alter reflecting objects in the image and investigate their influence on the wave field in the enclosed space in a physically correct way. This can be verified by reimaging the altered wave field. Preliminary results from listening tests for the perceptual evaluation are presented. They indicate that the influence of a reflecting object can only be perceived in its close proximity.
[DOI: 10.1121/1.1785591]

PACS numbers: 43.55.Gx [MK] Pages: 2126–2137

I. INTRODUCTION

Traditional approaches to room acoustic modeling suffer from difficulty in the amount of details required to model. Usually, most objects are modeled as plane surfaces with a specified reflection and diffusion coefficient. With the method presented here it is possible to determine the influence of a small object on the acoustics and thus determine to what detail an enclosed space should be modeled. Further, in the case of complex structures such as rows of seats, one can perform an acoustic image of the seats and then ''paste'' that into the modeling software.

Another suggested application is the treatment of strong unwanted (early) reflections. The acoustic image of those will help to identify the reflecting object such that corrective measures can be taken.

The proposition of the work presented here is to map the reflections in the impulse response to the reflecting object resulting in an image of the enclosed space. It will be shown that it is then possible to remove an object from the image and quantify its influence on the acoustics in the enclosed space. The approach is essentially new though some effort to the same end has been made previously.[1]

It is expected that the proposed method will give a better understanding of the importance of individual objects on the acoustics of an enclosed space.

The principle of acoustic reflection imaging is well known in seismics where it is used to characterize the earth's subsurface.[2,3] The fundamental is the measurement of impulse responses along receiver arrays; more recently this has also been introduced in room acoustics.[4]

The acoustics of an enclosed space for a particular source and receiver position is completely described by the room impulse response (RIR) in the time domain or the room transfer function in the frequency domain. Mathematically the room can be interpreted as a linear time-invariant system leading to the following expression,

$$p(t)=\int_{-\infty}^{\infty} h(\tau)s(t-\tau)\mathrm{d}\tau, \tag{1}$$

where $p(t)$ is the pressure, $s(t)$ is the source function and $h(t)$ is the impulse response of the room. The problem is that $h(t)$ is space-variant such that there are infinitely many RIRs in an enclosed space, one for each source–receiver configuration. Traditionally, the acoustics of a hall was characterized, objectively and subjectively, by a small number of representative source and receiver positions.[5,6]

In Fig. 1 two RIRs are shown measured a distance of 0.5 m apart in a simple shoe-box-shaped enclosure. What can be seen is that within the first 22 ms the differences are rather small and for later times the structure becomes more complicated and there is no more similarity. When measuring not just a few impulse responses at arbitrary positions but many and along a straight line the offset-travel time registration shown in Fig. 2 is obtained. With this array measurement method discrete wavefronts caused by reflections can be identified and it is obvious that the individual impulse responses are correlated. The two RIRs of Fig. 1 are marked with arrows.

Array technology is widely used in many fields of science; the most well-known applications in acoustics are in the areas of seismic exploration and medical diagnostics. Maynard *et al.*[7] introduced a method to visualize the near-field radiation characteristics of a sound source. In more re-

[a]Electronic mail: martin@akst.tn.tudelft.nl

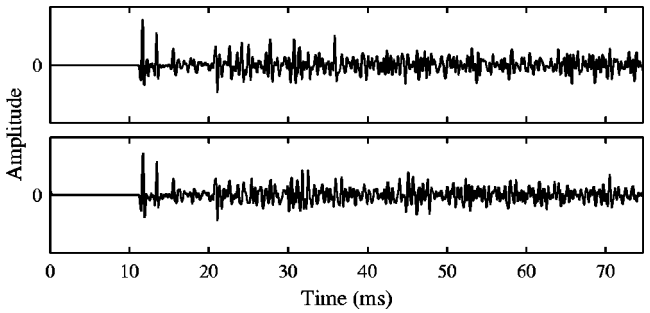


FIG. 1. Room impulse response (RIR) at two positions.

cent years Berkhout[8] proposed a method to reproduce sound fields with natural temporal and spatial properties from loudspeaker arrays.

II. THEORY

The Kirchhoff–Helmholtz integral states that the pressure at a point A in a source-free region can be expressed in terms of the pressure and normal component of the particle velocity on an enclosing surface S.[9] In the case that the surface S degenerates into an infinitely large plane the Kirchhoff–Helmholtz integral allows one to calculate the pressure for all points in the direction *away* from the source and this is known as forward wave field extrapolation. Strictly speaking, only knowledge of pressure or particle velocity is required, leading to the Rayleigh integrals. In the context of room acoustics where the observation plane S is embedded in the medium and waves consequently arrive from all directions, usage of both pressure and particle velocity allows one to distinguish between front and back incidence.

With the aid of inverse wave field extrapolation (or back propagation) it is possible to calculate the pressure at a point *A closer* to the source.[10] Due to the evanescent waves this operation is potentially unstable. To overcome this, the inverse wave field extrapolation operator is usually chosen such that it exponentially damps the evanescent waves. A natural choice is to use the complex conjugate of the forward operator which results in simple time reversal in the time domain.

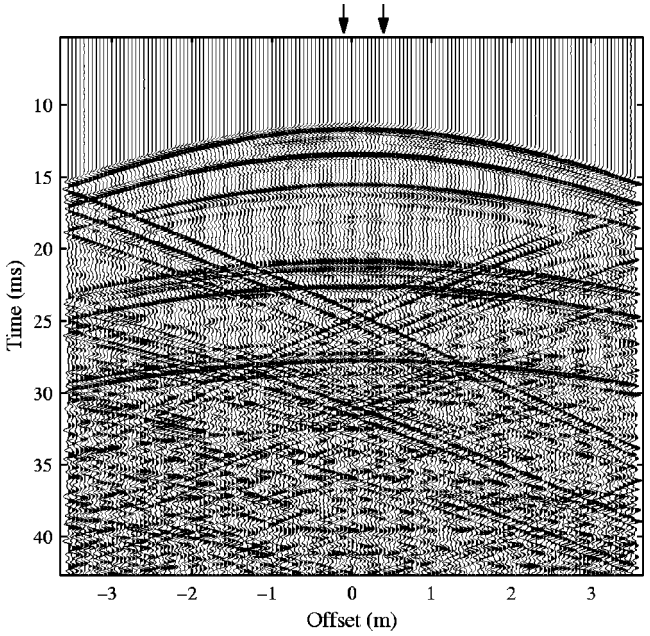


FIG. 2. Multiple room impulse responses measured along a linear array of microphones. The two RIRs of Fig. 1 are marked with arrows.

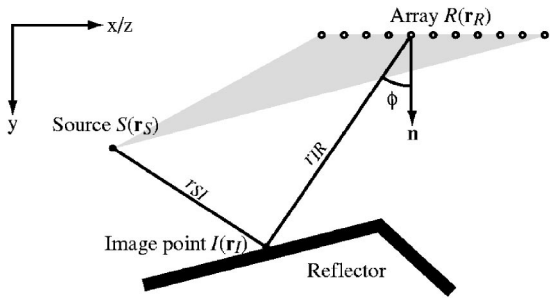


FIG. 3. Geometry for the imaging with source, image point, and receiver array position. The normal to the array points in the positive y direction. Imaging is not possible inside the gray-shaded area.

A. Imaging principle

The fundamental of acoustic reflection imaging is to treat each reflecting object as a secondary diffraction source and perform an inverse wave field extrapolation towards its position. Based on an expression for inverse wave field extrapolation with the Kirchhoff–Helmholtz integral in the time domain, the formulation reads mathematically as follows:[11]

$$\langle p_{\mathrm{Im}}(\mathbf{r}_I)\rangle = \int\int \mathrm{d}r_{R_x}\,\mathrm{d}r_{R_z}[w_{1I}(\mathbf{r}_I,\mathbf{r}_R,t)v_n(\mathbf{r}_R,t) + w_{2I}(\mathbf{r}_I,\mathbf{r}_R,t)p(\mathbf{r}_R,t)]\big|_{t=\tau(\mathbf{r}_S,\mathbf{r}_I,\mathbf{r}_R)}, \quad (2)$$

where $p(\mathbf{r}_R,t)$ and $v_n(\mathbf{r}_R,t)$ are measured pressure and normal component of the particle velocity, respectively, $p_{\mathrm{Im}}(\mathbf{r}_I)$ is the image at point I in terms of reflected pressure, the Green function kernels $w_{1I}(\mathbf{r}_I,\mathbf{r}_R,t)$ and $w_{2I}(\mathbf{r}_I,\mathbf{r}_R,t)$ are given by

$$w_{1I}(\mathbf{r}_I,\mathbf{r}_R,t) = \rho_0 \frac{1}{4\pi r_{IR}}\frac{\partial}{\partial t} \quad (3)$$

$$w_{2I}(\mathbf{r}_I,\mathbf{r}_R,t) = \frac{\cos\phi}{4\pi r_{IR}}\left(\frac{1}{r_{IR}} - \frac{1}{c}\frac{\partial}{\partial t}\right), \quad (4)$$

and

$$\tau(\mathbf{r}_S,\mathbf{r}_I,\mathbf{r}_R) = \frac{r_{SI}+r_{IR}}{c} \quad (5)$$

is the sum of travel times from the source S to the receivers R via the image point I. All other symbols have their usual meanings. The position vector $\mathbf{r}_R$ has the three Cartesian components $(r_{R_x}, r_{R_y}, r_{R_z})$. See Fig. 3 for a definition of the geometry.

Note that S is a constant, I is a parameter and R (more precisely r_{R_x} and r_{R_z}) is the integration variable in Eq. (2). The integration in Eq. (2) with the travel time condition of Eq. (5) is performed over the so-called Huygens surface.[3]

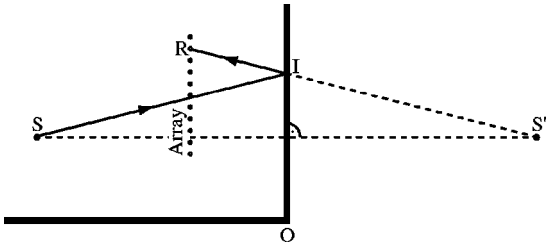


FIG. 4. Illustration of the imaging for one reflecting object.

In the case that only pressure information is available, imaging can be performed based on the Rayleigh II integral. Using the same notation as in Eq. (2), the formulation reads

$$\langle p_{\mathrm{Im}}(\mathbf{r}_I)\rangle = 2\int\int \mathrm{d}r_{R_x}\,\mathrm{d}r_{R_z}\, w_{2I}(\mathbf{r}_I,\mathbf{r}_R,t) \times p(\mathbf{r}_R,t)\big|_{t=\tau(\mathbf{r}_S,\mathbf{r}_I,\mathbf{r}_R)}. \quad (6)$$

The advantage of Eq. (2) over Eq. (6) is that the former allows us to distinguish between front and back by using both pressure and particle velocity. Equation (2) can be expressed in words as follows:

Step 1: Sample the wave field on the array, cf. $v_n(\mathbf{r}_R,t)$ and $p(\mathbf{r}_R,t)$ in Eq. (2).

Step 2: Perform an inverse wave field extrapolation to a point I, cf. $w_{1I}(\mathbf{r}_R,\mathbf{r}_I,t)$ and $w_{2I}(\mathbf{r}_R,\mathbf{r}_I,t)$ in Eq. (2) and $\tau = r_{IR}/c$ in Eq. (5).

Step 3: Calculate the travel time r_{SI}/c for the direct sound to arrive at I and use the extrapolated pressure amplitude at this time as image information, cf. p_{Im} in Eq. (2).

Step 4: Repeat steps 2 and 3 for all I in the region to be imaged.

The scheme has a potential problem when performing imaging on the side of the array containing the primary source. According to step 3 of the imaging process the pressure amplitude at the arrival time of the direct sound is used as image information. For image points in the area between the array and the primary source the inverse extrapolated field will not only contain the reflections but also the direct sound and the image will thus always show the direct sound. The conclusion is that imaging is not possible in the triangle defined by the array edges and the primary source position, corresponding to the gray-shaded area in Fig. 3.

Application of Eq. (2) requires a planar array setup. The formulation can still be employed with a linear array setup. However, since the linear array is rotationally invariant about its axis, the image of the three-dimensional enclosed space will be projected onto a plane (2-D Imaging).

B. Constructing the image of a reflecting object

The understanding of the imaging process is facilitated by the aid of geometrical room acoustics.

Consider the situation in Fig. 4 where a sound ray from the source S is travelling towards the vertical boundary and is reflected at the boundary in point I. For simplicity it is assumed that the reflection is specular and the ray will then be registered at the array in point R. The position of the boundary, however, is unknown in the imaging process. Instead, an inverse wave field extrapolation will be performed for many points between the array and the reflecting object and beyond. If the inverse wave field extrapolation is based on the Kirchhoff integral, both pressure and normal component of the particle velocity are available and consequently the extrapolated amplitude will only be nonzero (or above the noise floor) for points in the direction of the vector $\overrightarrow{RI}$. The second condition to be satisfied is the total travel time corresponding to the distance $\overline{SIR}$. The image point I can then only be placed at the unique position satisfying both conditions. Naturally, the distance $\overline{SI}$ is then equal to the distance $\overline{S'I}$ because that is how the image source was constructed in the first place.

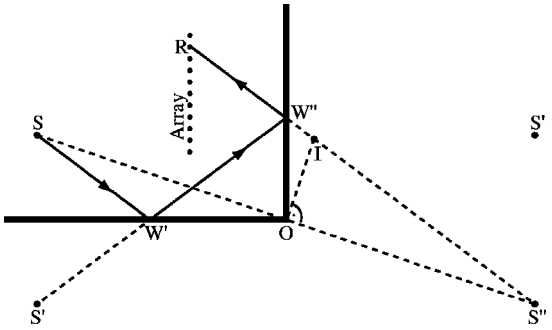


FIG. 5. Illustration of the imaging due to second-order reflections.

C. Constructing the second-order image

In practice wave propagation in enclosed spaces is not constrained to single reflections but subject to second- and higher-order reflections. Referring to Fig. 5, a sound ray travels from S towards the horizontal boundary where it is reflected at W' and then travels towards the vertical boundary. After reflection at point W'' the ray is registered at the array in point R. The imaging process, however, will assume that the ray was only reflected once. The construction of the image then follows along the same lines as in the case of a single reflection. This implies that the only point satisfying both ray direction and total travel time will be point I. The entire vertical boundary will then be imaged on the extension of the vector $\overrightarrow{OI}$. For this situation the distance $\overline{SI}$ is equal to the distance $\overline{S''I}$ for all image points. In the case that a sound ray is reflected first on the vertical and then the horizontal boundary, the image will be appear on the extension of the vector $\overrightarrow{OI}$ below point O.

The idea can also be extended to images from reflections at three objects and more. The image will always appear on the perpendicular bisection of the line from the primary source to the generating mirror image source. In the context of the work presented here, the nth-order image is referred to as the image caused by n reflections or equivalently by a mirror image source of order n.

One potential solution to the problem of higher-order images is to specify source position S' as the actual sound source. The second-order image would then appear at the actual position of the boundary with the drawback that the first-order image would be imaged at a wrong position inside

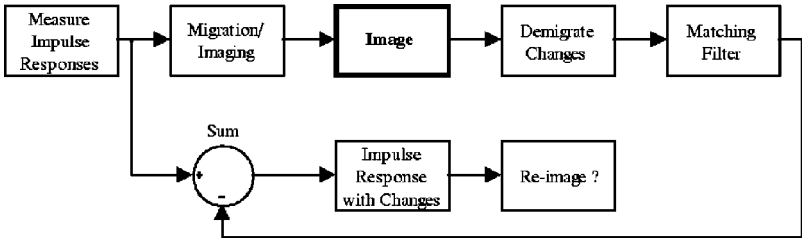


FIG. 6. Block diagram of the process to alter objects.

the enclosure. The effect of specifying an image source as actual sound source can be understood as a translation of the image from one reference frame to another. Rather than mirroring the source, it is also possible to mirror the array, an approach employed in time-reversal focusing.[12] Moreover, there are methods known in seismics to separate or at least attenuate multiple reflections from the RIR;[13] their implementation is the subject of further research.

As will become evident in later sections of this paper, the correct dealing of second- and higher-order images is important because (a) they allow one to add information where the first-order image is not visible and (b) in order to correctly remove the influence of an object from the RIR, not only the first-order reflection but all second- and higher-order reflections involving that object have to be removed.

D. Demigration principle

Demigration is understood as the inversion of the imaging (or migration) process. While imaging can be understood as a transformation of the measured RIRs into an image (from time to distance domain), the demigration transforms the image back into the RIRs (from distance to time domain).

The principle is to interpret each image point as a notional point source firing at the time corresponding to its distance from the primary source and with its strength given by the reflected pressure amplitude. The scheme can be interpreted as the inversion of Eq. (6). The integration is also performed over points in space and looks as follows:

$$\langle p(\mathbf{r}_R,t)\rangle = \int\int \mathrm{d}r_{I_x}\,\mathrm{d}r_{I_z}\, w_{2F}(\mathbf{r}_R,\mathbf{r}_I,t) \times p_{\mathrm{Im}}(\mathbf{r}_I)\big|_{r_{I_y}=\psi(\mathbf{r}_S,r_{I_x},r_{I_z},\mathbf{r}_R,t)}, \tag{7}$$

where

$$w_{2F}(\mathbf{r}_R,\mathbf{r}_I,t) = \frac{\cos\phi}{4\pi r_{IR}}\left(\frac{1}{r_{IR}}+\frac{1}{c}\frac{\partial}{\partial t}\right) \tag{8}$$

and the integration surface $r_{I_y}=\psi(\mathbf{r}_S,r_{I_x},r_{I_z},\mathbf{r}_R,t)$, the surface of equal reflection time or isochrone, is given implicitly by

$$t(\mathbf{r}_S,r_{I_x},r_{I_y}=\psi(\mathbf{r}_S,r_{I_x},r_{I_z},\mathbf{r}_R,t),r_{I_z},\mathbf{r}_R) = \frac{r_{SI}+r_{IR}}{c}, \tag{9}$$

in other words, by the same sum of travel times as in Eq. (5) except that in this case I rather than R varies. Equation (7) essentially describes forward wave field extrapolation with the Rayleigh II integral where the pressure at the image point I is only known at one instant in time (that is the travel time of the direct sound from S to I).

The fact that the notional sources have no directivity is maybe not too intuitive for the specular reflections but many notional sources of equal strength aligned on a plane will lead to a specular reflection.

When applied over all possible image points, application of Eq. (7) will potentially recreate the original RIR. In practice, where imaging is performed with band-limited signals and a finite aperture, it was found that there is a difference between the original and demigrated RIR in phase and amplitude and a matching filter was employed to minimize the difference in a least-square sense. Another point is that the image was sampled according to the minimum sampling criterion ($\Delta y=\lambda_{\min}/4$). Without oversampling in image space, a low-pass filter, analogous to the reconstruction filter in a digital-to-analogue conversion, has to be applied to the reconstructed impulse response in order to suppress the aliased high frequencies.

1. Alteration of objects

Using demigration it is possible to alter objects and investigate the changes in the acoustics by means of the impulse response. The scheme is shown in Fig. 6. The image of the object to be altered is demigrated. The demigration has to be applied only over those image points containing any changes. After applying the matching filter, the obtained impulse responses are subtracted from the original measured ones and the outcome will contain the impulse responses with the changes. The reason for this sequence is that the reflections due to the object to be altered first have to be removed. In order to take multiple reflections into account, the process has to be repeated with the images obtained from specifying the mirror image sources as primary source.

Examples of object alterations are as follows.

(i) Remove the object (this will result in a hole).
(ii) Replace the object with another object.
(iii) Alter the frequency-dependent reflection properties of the object by frequency filtering the demigrated impulse responses before subtraction.
(iv) Render the surface of a planar object more diffusive by displacing individual image points by a small random or periodic offset analogous to diffusors based on reflection gratings.

III. SIMULATIONS

The pressure and particle velocity impulse responses were generated using a mirror image source model of analytical monopoles with the second derivative of a Gaussian

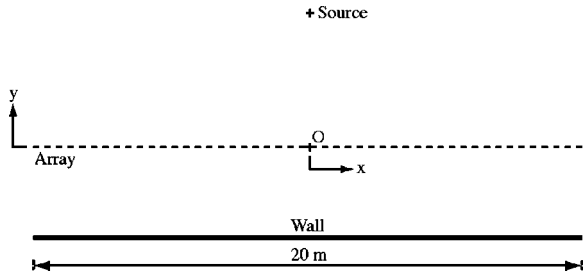


FIG. 7. Geometry of the single reflector imaging.

as wavelet. The array consisted of ideal pressure microphones spaced at 0.05 m. The sampling frequency was 16 kHz.

A. Illustration of the imaging principle

The workings of the imaging are illustrated for the case of one reflector as shown in Fig. 7. The reflector extends from $x=-10\,\mathrm{m}$ to $+10$ m and is at $y=-2.71\,\mathrm{m}$. A point source is placed at $y=4.10\,\mathrm{m}$. The linear array is at $y=0$ and has the same dimension as the reflector. Both, source and array, are placed symmetrically in front of it. In Fig. 8 the inverse extrapolated field is shown for four y positions. This corresponds to step 2 in the process of Sec. II A. The dashed hyperbola represents the travel time for the direct sound to arrive at the particular position and is the line along which the pressure will be used as image content for the given y position.

In Fig. 8(a) the y position is between the array and the reflector and the image content will be zero (or noise). In Fig. 8(b) the field shown is exactly at the reflector position and it can be seen that the hyperbola follows the wave front

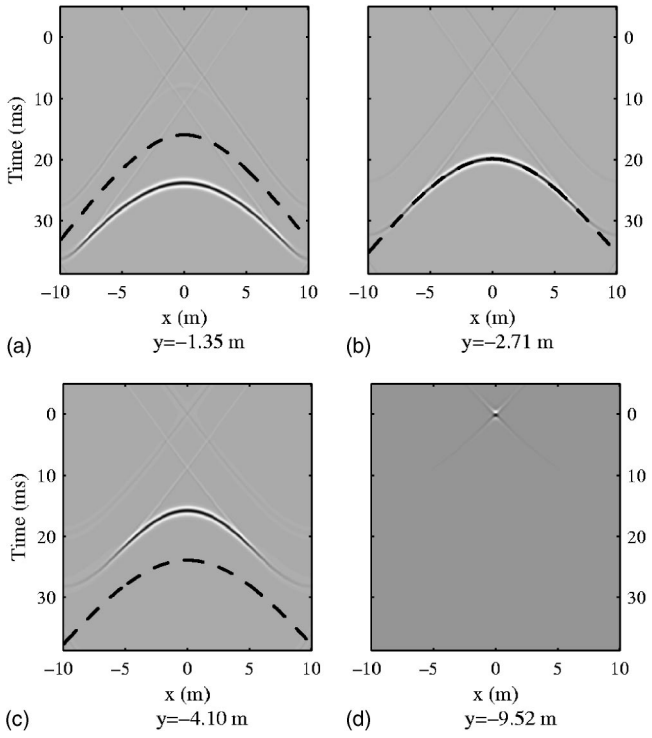


FIG. 8. Inverse extrapolated field due to one reflector at four y positions. The dashed hyperbola is the line along which the pressure will be used as image content.

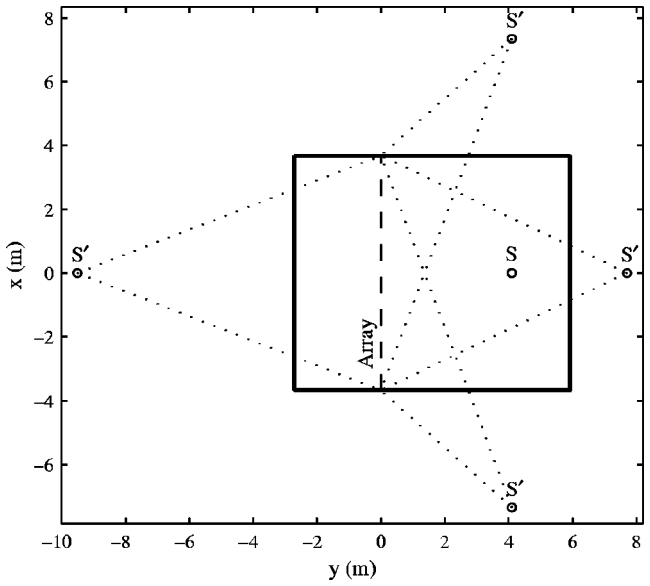


FIG. 9. Geometry of the shoe box enclosure with array and sound source S. First-order mirror image sources are indicated S′.

which will then constitute the image of the reflector. In Fig. 8(c) the y position is beyond the reflector and as in Fig. 8(a) the image will not contain significant amplitudes. The main characteristic when moving away from the array is that the extrapolated wavefront is shifted towards decreasing time and the hyperbola along which the pressure is used as image content is shifted towards increasing time. They only coincide at the position of the wall.

Since the wave front is not a delta function but spread out in time, the image amplitude just in front and behind the reflector is not zero but contains the amplitudes of the wavelet at either side of the main lobe. Imaging with band-limited wavelets will consequently always lead to images of reflectors with finite width.

In Fig. 8(d) the field is shown for the y position of the mirror image source caused by the reflector. The whole wave front has collapsed into a point at $t=0$ and $x=0$ and the mirror image source is in focus. However, the aim is to image the reflector and its secondary sources are in focus in Fig. 8(b).

B. 2-D imaging: Shoe box enclosure

A floor plan of the enclosure together with source and receiver geometry is shown in Fig. 9. The vertical distance from both the source and the array to the floor was 1.20 m and that to the ceiling 1.72 m. The reflection coefficient of all boundaries was set to 1. Since the simulations involved a linear array only 2-D imaging could be performed.

1. Enclosure image without floor and ceiling

In order to remove the floor and ceiling, the reflection coefficient of those two boundaries was set to zero. The obtained image of the enclosure is shown in Fig. 10 where for clarity the negative image amplitudes have been omitted. From the figure, it is immediately evident that only parts of the wall appear in the image. This is basically a visibility issue caused by the finite aperture of the array. In Fig. 9 the

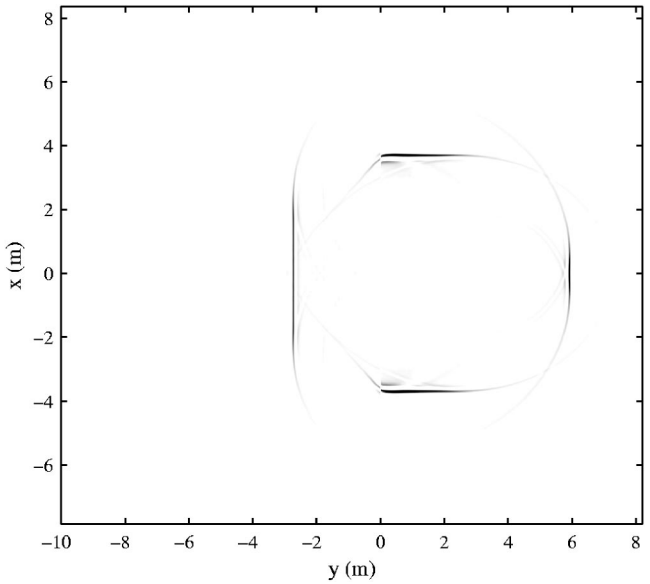


FIG. 10. Acoustic image of the shoe box enclosure from simulated data taking only the first-order image sources of the walls into account.

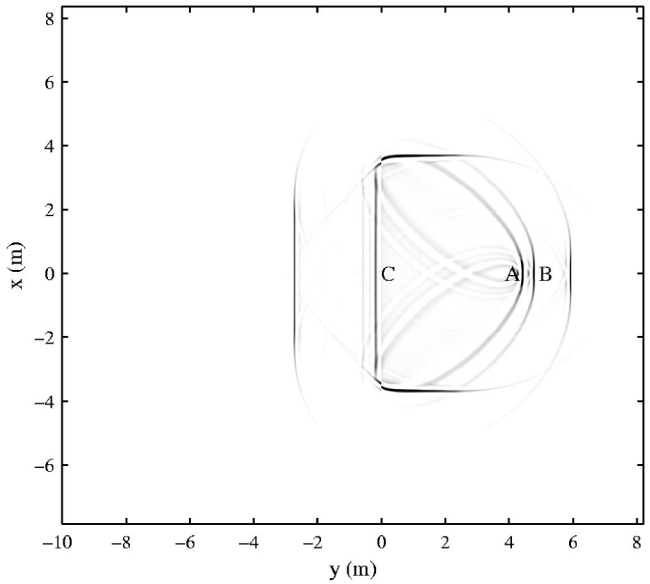


FIG. 12. Acoustic image of the shoe box enclosure from simulated data taking all first-order image sources into account.

two dotted lines from each image source to the array edges indicate which part of the wall will be seen on the image. This example illustrates nicely the helpfulness of the image source model to understand the imaging result.

2. Enclosure image without floor and ceiling from second-order reflections

Figure 11 shows the image of the enclosure when only second-order reflections are used in the simulation. The imaging has been performed specifying the first-order image sources as sound sources. For each one only those second-order reflections involving the particular image source have been included in the simulations. Figure 11 is thus a summation of the four images obtained from all four first-order image sources. From the figure it can be seen that the second-order images allow one to add information about the walls where there was no information in the first-order image, in particular at the enclosure corners.

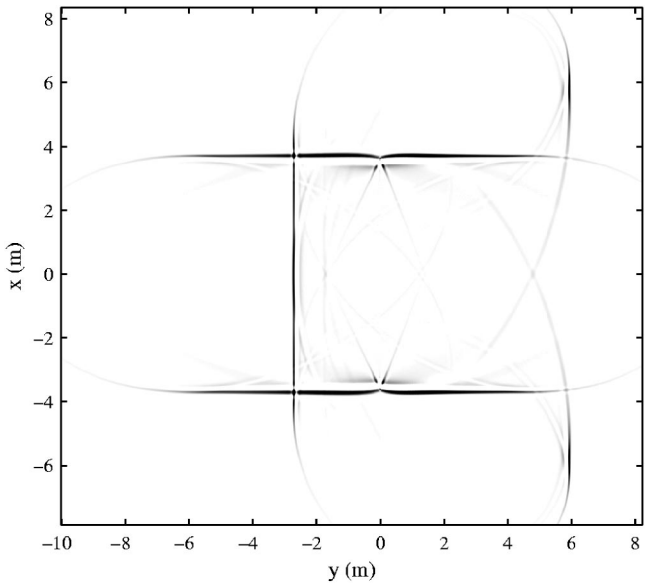


FIG. 11. Acoustic image of the shoe box enclosure from simulated data performing the imaging with respect to the first-order image source and considering only second-order reflections.

3. Enclosure image with floor and ceiling

The situation where floor and ceiling are present is now considered. The image obtained is shown in Fig. 12. Compared to Fig. 10, there are two more events marked ''A'' and ''B'' visible. ''A'' is the image of the floor and ''B'' is the image of the ceiling. The occurrence of these two images is a consequence of the projection resulting from the linear array setup as addressed in Sec. II A. Another issue is that the linear array relies on the polarity of pressure and particle velocity to distinguish between front and back. However, this works only correctly for mirror image sources in the $x-y$ plane and consequently the images of the floor and ceiling will appear twice. In Fig. 12 event ''C'' is the image of the floor on the incorrect side of the array while the one due to the ceiling is barely visible just to the left of it. Events ''A,'' ''B,'' and ''C'' are undesired artifacts purely due to 2-D imaging and will disappear in 3-D imaging.

4. Enclosure image incorporating higher-order image sources

Figures 10 and 12 were obtained by only including first-order mirror image source in the model. For a comparison with images obtained from measured data, the simulated data should incorporate all mirror image sources up to at least a specified distance. This leads to the image of the enclosure in Fig. 13 for the case that image sources within 20 m are included. From the figure, a whole series of additional images is visible. As discussed in Sec. II C, each mirror image source creates its own image. In the same section the position of those images was established. In the figure the three events above the letter ''D'' are the first- and second-order images of the side wall with negative x coordinate while the

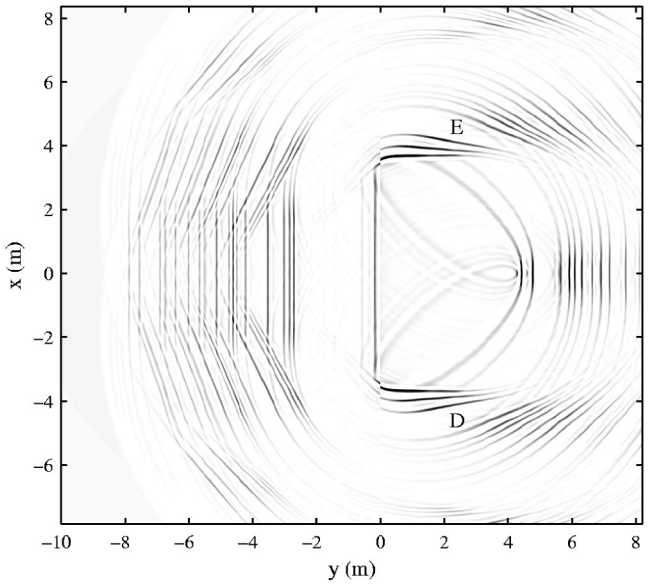


FIG. 13. Acoustic image of the shoe box enclosure from simulated data taking all image sources within a distance of 20 m from the array center into account.

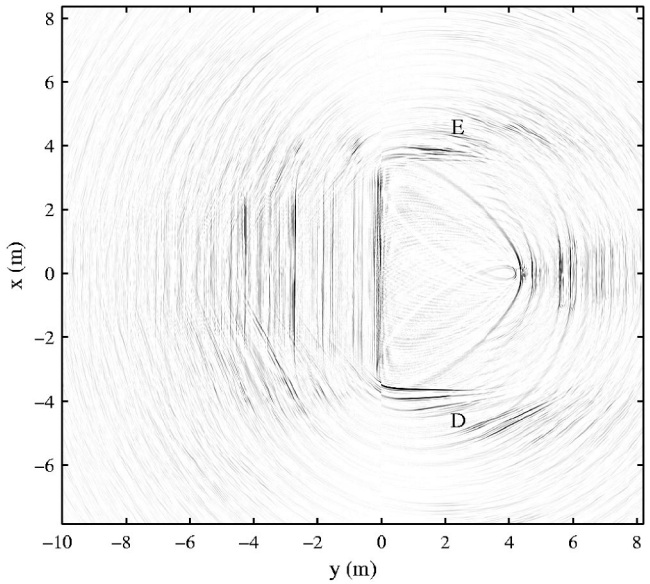


FIG. 14. Acoustic image of the shoe box enclosure from measured data.

events to the right are the higher-order images due to reflections between the side wall and the wall behind the source. The same structure can be seen around ''E'' for the side wall with positive *x* coordinate. The visibility of the higher-order images is defined in the same way as for the first-order images.

In comparison with Fig. 10, Fig. 13 is much more difficult to interpret. It is important to understand at this stage the difference between events due to 2-D imaging of a 3-D space and those due second- and higher-order reflections. The former's position in the image is a function of the room and measurement geometry and they can, as is the case here, appear *within* the physical boundaries of the enclosure. Images of second- and higher-order reflections, on the other hand, will always appear *outside* the physical boundaries of the enclosure. As Fig. 5 illustrates, enclosure corners are the only points where images of different order coincide. Neglecting the artifacts due to 2-D imaging, Fig. 10 can be retrieved from Fig. 13 by constructing radial traces from the source position in all directions and observing the position of the maximum pressure amplitude in those traces. The shortcoming of such an approach is that in the enclosure corners where the first-order image is not visible the algorithm would include the second-order image information. This problem can only be solved by interpreting the image, but in practice the enclosure geometry is more or less known as *a priori* information.

IV. MEASUREMENTS

Measurements have been performed using a source that was reasonably omni-directional within the area of interest. Pressure and particle velocity were measured using a Sound-Field MKV microphone, again with a spacing of 0.05 m. The sampling frequency was 16 kHz except for the shoe box enclosure where it was 14 980 Hz.

A. 2-D imaging: Shoe box enclosure

Figure 14 shows the acoustic image obtained from measurements in a lecture hall with the same dimensions as the enclosure of Sec. III B. When comparing to Fig. 13, it can be seen that all events discussed so far are reproduced in the image. Most of the discrepancies can be explained by the differences in reflection coefficients and the not modelled irregularity of some surfaces. The image of the side wall with positive *x* coordinate (below letter ''E''), for example, looks much more diffuse than the one at negative *x* coordinate (above letter ''D''). This is explained by the fact that the particular wall consisted of a window front covered by curtains rather than a hard wall.

B. 2-D imaging: Concert hall De Doelen

As a more practical case the imaging process is tested on measurements performed in the concert hall De Doelen, Rotterdam. A photograph of the hall is given in Fig. 15 and the floor plan is shown in Fig. 16. The source was situated on the stage approximately at the conductor's position. The linear array was positioned between the tiers of seating and spans the whole width of the inner seating area.

Figure 17 shows the acoustic image of the area within the dash-dotted rectangle of Fig. 16. In the figure the struc-

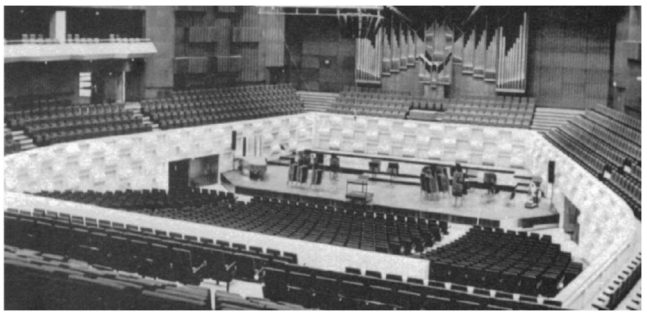


FIG. 15. De Doelen, Rotterdam (from M Barron. *Auditorium acoustics and architectural design*).

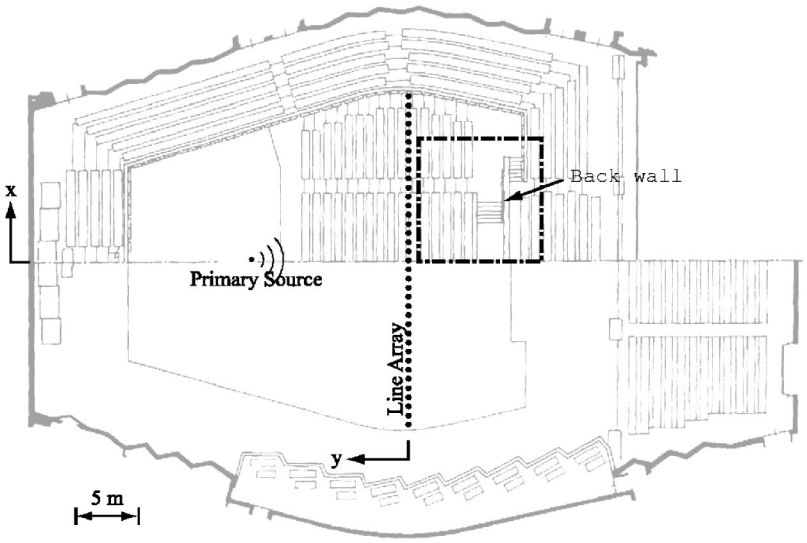


FIG. 16. Floorplan of De Doelen, Rotterdam with source and array position indicated. The dash-dotted rectangle is referred to in the text (from M Barron. *Auditorium acoustics and architectural design*).

ture enclosed in the black rectangle represents the image of the wall behind the last tier. The structure of this diffusive wall can be seen in Fig. 15 as the white checkerboard squares of convex curved shape surrounding the inner seating area. This sequence is very nicely reproduced in the image by segments above and below a median line. The vertical bands on the left side of the figure are the images of the tiers of seating and the dashed line indicates the position of the corridor between them. It is possible to recognize the top of the backrest of each individual seat in the row. Further, one can also observe that below the corridor seats from subsequent rows are offset by half the seat's width such as to improve visibility towards the stage whereas above the corridor this measure has been omitted since the audience already needs to look slightly sidewards to see towards the center of the stage. Also notable is that due to the mapping of the three-dimensional space onto a plane, the rows are not spaced equidistantly.

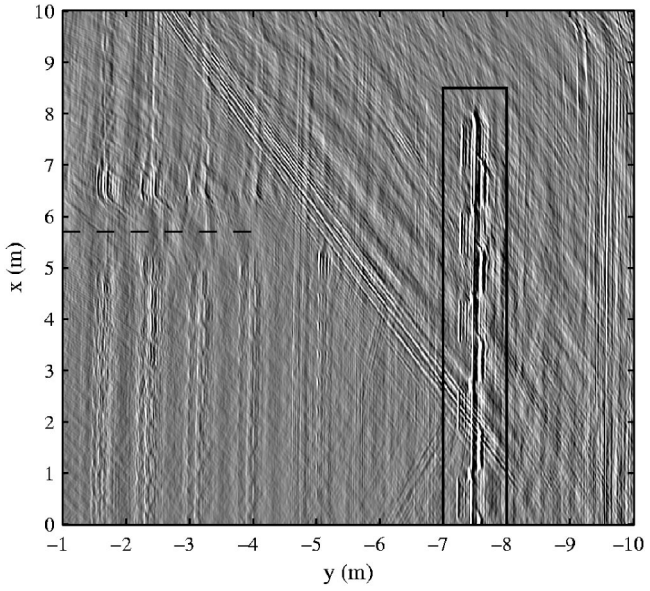


FIG. 17. Acoustic image of the area within the dash-dotted rectangle of Fig. 16.

C. 3-D imaging: Corridor wall

Planar array measurements have been carried out in a simple building corridor. The array runs parallel to the corridor wall at a distance of 0.88 m and over a length of 7.00 m. The height of the array was 2.50 m. The shortest distance between the source and the array was 1.68 m. A photograph of the corridor wall is shown in Fig. 18 while Fig. 19 shows the floor plan and elevation. The chair at the right of the left column is not present in the photograph. The origin of the coordinate system is at the center of the lower edge of the array aperture which is at a height of 0.35 m from the floor. The rather unusual axis orientation is due to the measurement setup.

The problem of displaying the three-dimensional image was approached by showing horizontal slices of the image at different heights and an attempt to visualize the wall surface in three dimensions.

1. Slices at different heights

In Fig. 20 horizontal slices are shown at four representative heights. With reference to Fig. 19 and the photograph in Fig. 18 most objects can be identified.

In all four images the vertical columns are seen at $x=-1$ m and $x=2$ m and also the wall in-between is imaged well. In Figs. 20(a)–(c) the front of the closet is seen between approximately $x=-3$ m and $x=-2$ m while in Fig.

FIG. 18. Photograph of the corridor wall.

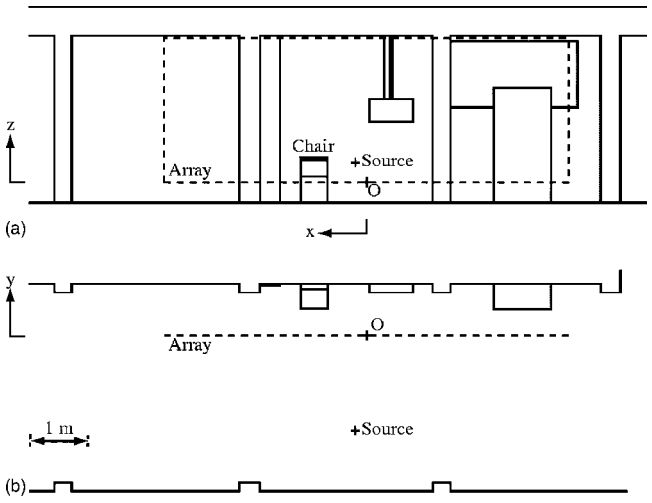


FIG. 19. Elevation (a) and floor plan (b) of the corridor.

20(d) it has disappeared since the slice is at a height above the upper edge of the closet. In Fig. 20(a) the backrest of the chair in front of the wall can be seen at $x=1$ m. In Fig. 20(c) the image of the wall between the columns is discontinued between $x=-0.5$ m and $x=0.4$ m by the electrical distribution box. Also seen, particularly well in Fig. 20(b), is the small protrusion, 0.33 m wide and constructed out of the same material as the columns, to the right of the left column.

The four slices shown were selected rather arbitrarily but showed the main objects present. Naturally, one could also display slices in the vertical plane. However, the visual perception of an object is limited in either case.

2. Surface representation

The second approach used to display the information is based on the fact that the wall and the objects in front form a surface. For each (x,z) pair of the volume image, the y position of the maximum pressure amplitude in the image is recorded. The motivation is that a ray towards the wall will be reflected at only one point (unless there are acoustically transparent materials). The surface formed is then passing through all those points. The result is shown in Fig. 21 for the part of the wall visible from the array. In comparison to the slices the interpretation is much more intuitive. The closet and the two vertical columns can be recognized easily. Also the structure of the distribution box can be identified. The lump at the bottom 1 m to the right of the left column is the surface due to the chair while the spikes to the left are due to the washbasin. The fine vertical line just left of the column is the image of the plastic drain pipe. The smallest object that can be positively identified is the light switch mounted at half-height on the left column.

V. DEMIGRATION

The images shown in Sec. IV were interesting to look at but compared to a photograph their creation is much more laborious and, moreover, due to the longer wavelength, they do not show the same amount of detail. The main difference, however, is that the image is obtained in terms of acoustic rather than electromagnetic reflection properties. The second

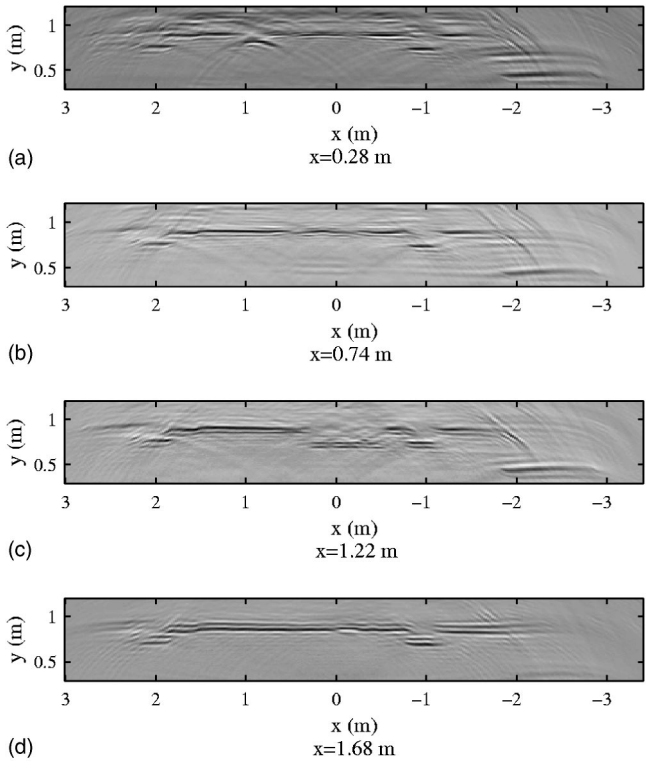


FIG. 20. Image slices at different heights.

important issue is that the room impulse responses used to obtain the image represent the transfer function between the source and the receiver and thus define the acoustic quality in an objective and subjective sense. The interesting part, introduced in this chapter, is the possibility to remove or alter objects in the image and investigate the influence on a particular impulse response either by objective parameters or perceptually.

A. Removal of objects

The process of removing objects was investigated for the distribution box and the closet in front of the corridor wall from Sec. IV C. In Fig. 22 a horizontal slice of the image is shown where the cross section of the volume over which the demigration was performed is indicated by the dashed rectangle. The volume extended over the whole height of the distribution box. The same volume was used in the demigration of the second-order images caused by ground and ceiling reflection, respectively.

For the closet the whole area of x values smaller than $x=-1.62$ m over its height was removed. The motivation was to include the second-order image caused by reflections from the side of the closet. For both objects the consequence of removal is that there is a hole in the wall. With the closet, the wall was effectively shortened.

1. Influence on the impulse responses

Another way to investigate the influence of the object on the impulse response is to calculate the energy contained in the reflections from the object after demigration. Figure 23 displays a view at the planar array aperture where the grayness is proportional to the reflected energy at that measure-

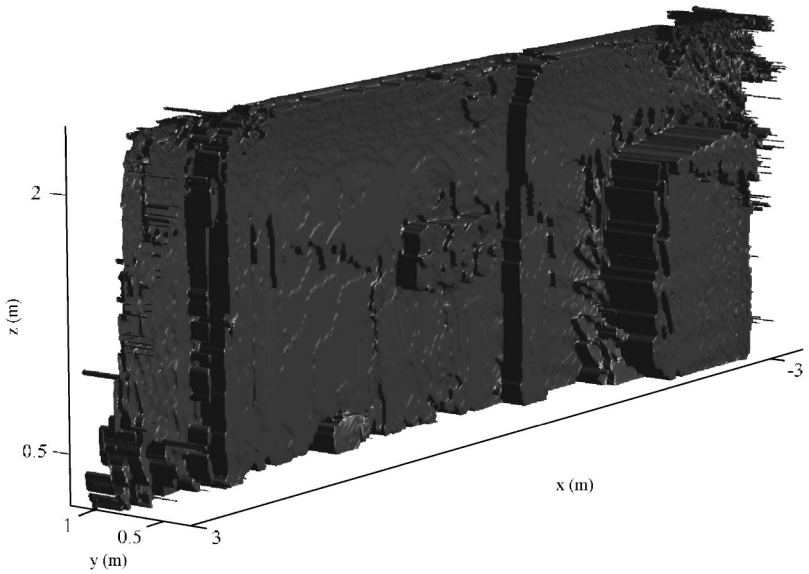


FIG. 21. Surface representation of the corridor wall.

ment position. The dotted rectangles mark the projection of the distribution box and the closet onto that plane.

The influence of the closet is generally confined within an area right in front of it though more concentrated on the right side because of the direction of the reflected waves from the source. There is some spread above the upper edge due to the second-order reflection via the floor. The gray area between $x=-1$ m and $x=-2$ m is caused by second-order reflections from the side of the closet and the wall. The influence of the distribution box is largest just above its position, again due to the direction of the reflected waves. The larger spread compared to the closet is partly due to second-order reflections via floor and ceiling and partly due to the fact that the distribution box is of irregular shape causing scattering. For better visibility the energy due to the distribution box has been amplified by a factor of 3.

B. Reimaging

The ultimate test to check whether the removal of the object was performed in a physically correct way is to reimage from the impulse responses obtained after demigration. This was performed for the distribution box. The result is shown in Fig. 24 for the same slice as in Fig. 22. As the figure shows, the object has really been removed almost completely while all other objects are still present. Presumably due to the matching filter, there are residues of the removed object and some minor differences in the fine structure of the image. When using the same surface representation as in Sec. IV C 2 in combination with a threshold to block small image amplitudes, Fig. 25 is obtained. From the figure it can be seen that, in fact, the object has been removed with the exceptions of some residues partly due to imperfections in the matching filter process and partly due to fact that the control volume did not extend deep enough in the z direction.

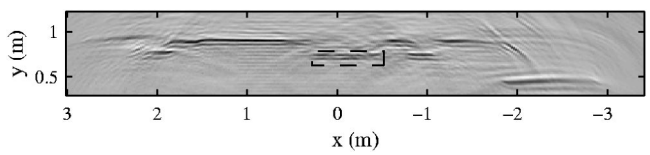


FIG. 22. Image slice at distribution box height ($z=1.16$ m).

VI. LISTENING TESTS

A complete psychoacoustical evaluation of the audibility of reflections from an indivual object is beyond the scope of this paper, Instead, a short initial study has been conducted with the aim to obtain a basic understanding to what extent the reflections from an individual object are perceptually relevant.

An extensive discussion about the subjective effects of reflections is discussed by Kuttruff.[14] In short, the main issues are that with a transient sound such as speech, the reflection will, depending on the delay time, either add to the loudness of the direct sound or be perceived as a distinctive echo. With a continuous sound such as noise the reflection will cause spectral colouration.

A. Experimental setup

The listening positions chosen were close to or right in front of the removed objects (closet and distribution box) and selected such that the energy ratios within the first 64 ms between altered and original RIR were distributed approximately even between 0.60 and 0.98. The selected positions

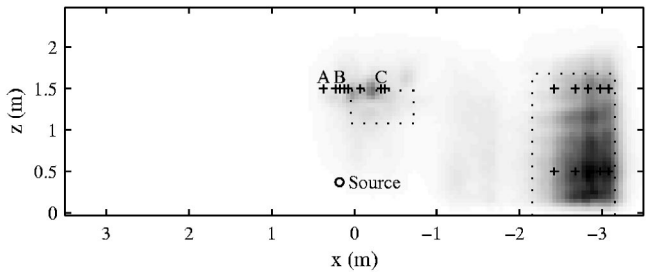


FIG. 23. View at array aperture. Grayness proportional to energy contained in the reflections from the closet and the distribution box. The dotted rectangles mark the projection of those onto the array plane. Positions included in the listening test are marked with crosses. Letters A-C are referred to in the text.

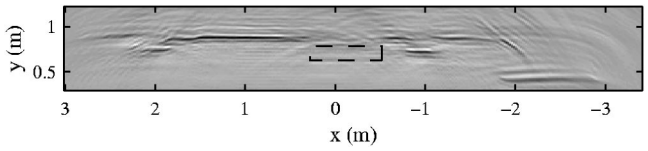


FIG. 24. Reimaged slice after removal of distribution box ($z = 1.16$ m).

used in the listening tests are indicated with crosses in Fig. 23. The shortest distance to the reflecting object was 0.45 m in the case of the closet and 0.75 m in the case of the distribution box.

The tests were performed binaurally with a closed headphone and selecting two RIRs at the approximate ear separation though no attempt has been made to correct for the head-related transfer function. The signals were fed to the headphone as if the subject was facing the reflecting wall. The RIRs were convolved with two monophonic signals, one being 4-s of noise shaped by the spectrum of male speech,[15] the other being a female speech sample of 2-s length.[16] The speech noise signals were normalized for equal energy of original and altered sample. There were a total of 13 listening positions with and 13 listening positions without direct sound. In the case without direct sound, the reflection from the object was the first wave front to arrive at the listening position.

A triple stimulus *AXB* listening test was employed where either of *A* and *B* were the samples with the original and altered RIR and *X* was the reference. The subjects were forced to choose whether *A* or *B* was equivalent to *X*. *X* was always the sample with the original RIR but this was not known by the test subjects.

The subjects were allowed and encouraged to take any perceived information such as loudness, spaciousness, coloration, and intelligibility into account. Because of the wide range of energy ratios, the difficulty of the task at hand varied by a great amount. The subjects were allowed to listen to a few randomly selected sets before the actual test for training purpose and also permitted to repeat any set during the test as many times as they wished to.

Fourteen subjects participated in the test, one of whom was later excluded due to inconsistent answers. All subjects stated to have normal hearing but their experience with listening tests varied by a great amount.

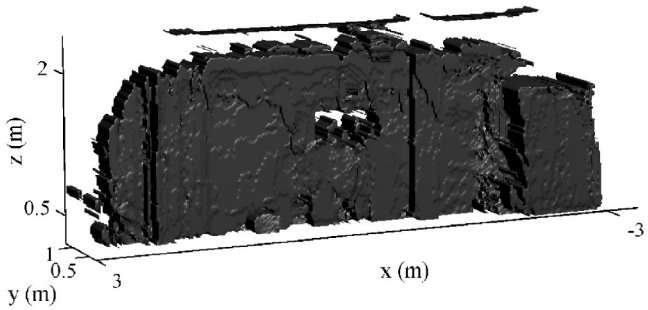


FIG. 25. Surface representation of the corridor wall without the distribution box.

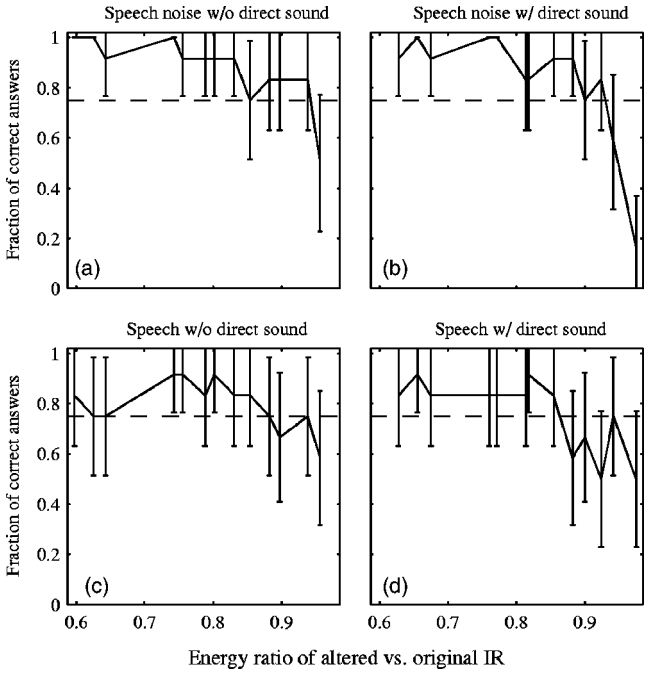


FIG. 26. Outcome of the listening tests expressed as mean between the subjects together with the 95% confidence bounds.

B. Results

The outcome of the listening tests is shown in Fig. 26 expressed as fraction of correct answers among the subjects versus energy ratio. The threshold, indicated by the dashed line, lies at 0.75 because statistically there is a 50% chance to guess the correct answer. The general trend observed in all graphs of Fig. 26 is a decrease from a fraction value close to unity towards a value of 0.5 for an increasing energy ratio. Values below 0.5 are a result of the small population size. Even though the confidence bounds are rather large, it appears safe to conclude that an energy ratio smaller than approximately 0.85 is required for a consistent detection of the objects' presence. Naturally, the energy in the RIR is a rather crude descriptor since the perception depends on the delay time and strength of individual reflections. Another observation is that with the direct sound it appears to be only slightly more difficult to detect the difference. It is possible that the comb filter effect occurring due to the short time delay between direct sound and reflection from the object facilitates the detection.

Looking at three positions from Fig. 23 in more detail, it was found that at position ''A'' none of the subjects were able to detect the difference. At position ''B'' the difference was detected reliably only in the case of speech noise without direct sound. At position ''C'' samples without direct sound allowed for the detection with both signals while with direct sound and speech noise the response was on the threshold. For the closet the differences were detected regardless whether with or without direct sound and speech or speech noise.

VII. CONCLUSIONS

A method has been proposed to map reflections in a room impulse response to the reflecting object resulting in an image of the enclosed space in terms of reflected pressure. The imaging process is based on inverse wave field extrapolation from the receiver to the object's position.

Simulations have been carried with the model of a simple shoe box enclosure. It was found that multiple reflections from more than one object lead to images of higher order. The position of those is generally outside the physical boundaries of the enclosed space. A comparison with measurements in a real enclosure of identical dimensions showed good agreement. Measurements performed in a corridor and a concert hall showed that considerable detail in the image can be obtained even at large distances.

By performing the inverse step of the imaging process it is possible to recreate the original impulse responses. This allows altering of objects in the image and their influence on the acoustics can (a) be calculated in a physically correct way and (b) be evaluated perceptually without actually altering the object. The process has been validated with two objects on one of the corridor's walls. Preliminary results from a listening test indicate that the influence of a reflecting object can only be perceived in its close proximity.

The applicability of the proposed method is limited because of the elaborate measurements associated with a planar array. Consequently, a way has to be found to deal with the inherent mapping of the three-dimensional space onto a plane due to linear array imaging. A possible approach is to exploit the fact that in a Cartesian system there are two normal particle velocity components on the array of which only one has currently been used. Further, the possibility of identifying and removing reflections involving a particular surface such as, e.g., the floor, would be highly beneficial.

ACKNOWLEDGMENTS

The authors would like to thank Eric Verschuur for his enthusiasm in helping them to apply the concepts of seismic imaging to room acoustics.

[1] J. Baan and D. de Vries, ''Auralization of sound fields by wave field synthesis,'' in *106th Convention* (Audio Engineering Society, Copenhagen, 1999).

[2] A. J. Berkhout, *Seismic Migration: Imaging of Acoustic Energy by Wave Field Extrapolation* (Elsevier, Amsterdam, 1984), Vol. 12.

[3] P. Hubral, J. Schleicher, and M. Tygel, ''A unified approach to 3-D seismic reflection imaging, Part I: Basic concepts,'' Geophysics **61**, 742–758 (1996).

[4] A. J. Berkhout, D. de Vries, and J. J. Sonke, ''Array technology for acoustic wave field analysis in enclosures,'' J. Acoust. Soc. Am. **102**, 2757–2770 (1997).

[5] J. S. Bradley, ''Comparison of concert hall measurements of spatial impression,'' J. Acoust. Soc. Am. **96**, 3525–3535 (1994).

[6] M. Barron and L. J. Lee, ''Energy relations in concert auditoriums. I,'' J. Acoust. Soc. Am. **84**, 618–628 (1988).

[7] J. D. Maynard, E. G. Williams, and Y. Lee, ''Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH,'' J. Acoust. Soc. Am. **78**, 1395–1412 (1985).

[8] A. J. Berkhout, D. de Vries, and P. Vogel, ''Acoustic control by wave field synthesis,'' J. Acoust. Soc. Am. **93**, 2764–2778 (1993).

[9] A. D. Pierce, in *Acoustics, an Introduction to its Physical Principles and Applications* (McGraw–Hill, New York, 1981), Chap. 4.

[10] A. J. Berkhout, in *Applied Seismic Wave Theory* (Elsevier, Amsterdam, 1987), Chap. 10.

[11] M. Tygel, J. Schleicher, and P. Hubral, ''A unified approach to 3-d seismic reflection imaging, part ii: Theory,'' Geophysics **61**, 759 (1996).

[12] S. Yon, M. Tanter, and M. Fink, ''Sound focusing in rooms: The time-reversal approach,'' J. Acoust. Soc. Am. **113**, 1533–1543 (2003).

[13] A. B. Weglein, ''Multiple attenuation: an overview of recent advances and the road ahead (1999),'' The Leading Edge **18**, 40–44 (1999).

[14] H. Kuttruff, in *Room Acoustics*, 4th ed. (Spon, London, 2000), Chap. 7.

[15] Instituut voor Zintuigfysiologie TNO, Spraakmateriaal behorende bij de test voor het meten van de spraakverstaanbaarheidsdrempel voor korte zinnen in stilte en in stationaire of fluctuerende ruis (English translation: Speech material accompanying a test for measuring the speech intelligibility threshold for short sentences in anechoic and stationary or fluctuating noise conditions), Audio CD, 1988.

[16] European Broadcasting Union, Sound Quality Assessment Material Recording for Subjective Tests, Audio CD, 1988, track 53.

Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array

Bradford N. Gover[a)]
Department of Physics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada, and Institute for Microstructural Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada

James G. Ryan[b)] and Michael R. Stinson
Institute for Microstructural Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada

(Received 13 February 2003; revised 13 May 2004; accepted 6 July 2004)

The directional variation of sound at a point has been studied in three rooms, using a measurement system described previously [J. Acoust. Soc. Am. **112**, 1980–1991 (2002)]. The system uses a pair of 32-element spherical microphone arrays to obtain directional impulse responses in each of 60 steering directions, with an angular resolution of 28°, covering all directions in the whole solid angle. Together, the array measurements span the frequency range from 300 to 3300 Hz. The angular distribution of incident sound energy is visualized on a three-dimensional plot, and quantified by computing the directional diffusion and the directional peak-to-average level difference (''anisotropy index'') of the sound field. The small-to-medium-sized rooms had reverberation times of 360, 400, and 600 ms. Measurements were made for several source and receiver locations in each, and were analyzed over several time ranges (full decay time of room, late time decay, 2-ms windows throughout the decay). All measured sound fields were found to be highly directional, the distribution of arriving energy at a point greatly influenced by the early specular reflections. The directions and arrival times of these reflections were identified from the measurements, giving excellent agreement with those expected from knowledge of the room geometry. It was observed that as time progressed, the sound fields initially exhibited increasing isotropy, followed by increasing *anisotropy*, due to nonuniform absorption in the rooms. The measurement system is capable of yielding detailed information about the reverberant sound field in a room, and is easily modified to be able to analyze ambient or time-varying fields. [DOI: 10.1121/1.1787525]

PACS numbers: 43.55.Mc, 43.55.Gx, 43.55.Br [MK] Pages: 2138–2148

I. INTRODUCTION

When studying the acoustics of reverberant spaces, the directional nature of the sound field is frequently of interest. Knowledge of how the sound field varies with direction at a point, and how this varies with time, can potentially be of great value in areas such as auditorium assessment or correction,[1] determination of psychoacoustic indicators,[2] validation of diffuse field assumptions,[3] or in optimal microphone array design.[4] The direction and time of arrival of individual reflections or an assessment of the overall diffuseness or isotropy of the field are commonly sought. The present study reports the results from the use of a new system to measure this information in sound fields in actual rooms.

Frequently, to study the sound field in a room, a single omnidirectional pressure microphone is used. Such measurements cannot yield spatial information as the directions of arrival of the component sound waves are lost. However, through inspection of a measured pressure-time signal, or more usually a room impulse response, the times of arrival of a few reflections are sometimes identifiable. Together with a detailed knowledge of the room geometry, it is occasionally possible to also identify reflecting surfaces and directions of arrival. This is not always possible, and, even when it is, can be a significant amount of work.

The correlation of pressure measured at different points in a room can be used to infer confidence in the diffuseness of the field.[5,6] However, this technique is not without its short-comings since the correlation coefficient deviates only weakly from its ideal diffuse values even in the presence of strong anisotropy.[7,8]

Measuring sound fields in rooms with an intensity probe, as is also common, can be useful in finding the directions of net energy transport.[9] Even measuring the intensity instantaneously does not necessarily aid in resolving reflections, however, since this direction of net energy flow is not always in the direction of a wavefront arrival. Analysis of intensity can, nevertheless, shed light on the overall diffuseness of the field.[10] There have been several studies reported in the past 10 years employing three-dimensional instantaneous intensity probes, including those by Oguro *et al.*,[11] Guy and Abdou,[12] Hori,[13] Peltonen *et al.*,[14] and Merimaa *et al.*[15]

It is possible to employ multiple pressure microphones in a variety of ways to derive directionally influenced information in a sound field. Examples of such studies include those by Yamasaki and Itow,[16] Sekiguchi *et al.*,[17] Täger and

[a)]Present address: Institute for Research in Construction, National Research Council, Ottawa, Ontario K1A 0R6, Canada. Electronic mail: brad.gover@nrc-cnrc.gc.ca
[b)]Present address: Gennum Corporation, 232 Herzberg Rd., Kanata, Ontario K2K 2A1, Canada.

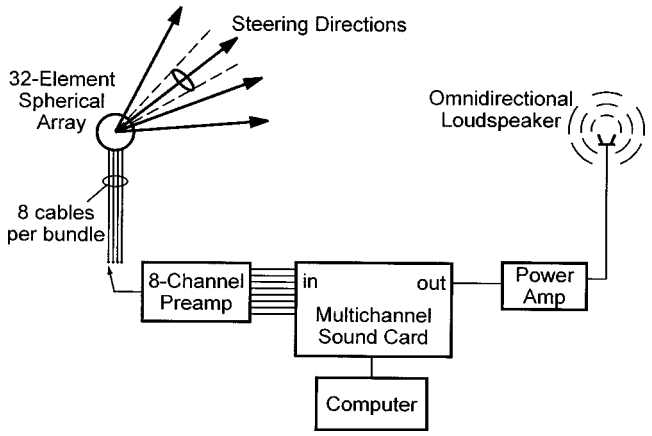


FIG. 1. Block diagram of measurement system. The microphone array steers a 28°-wide directional beam in 60 steering directions, detecting the sound arriving from each.

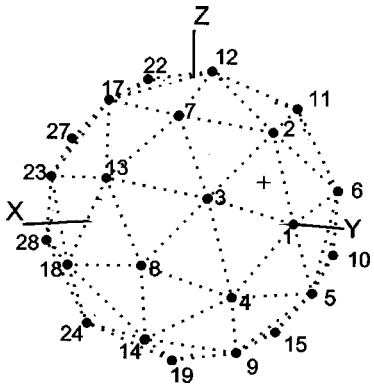


FIG. 2. Array geometry depicting microphone locations (black dots) and the primary steering direction (+). The 32 microphone positions (those on back half of sphere not shown) lie at the vertices of a geodesic sphere (shown with dotted lines). The other steering directions are through the centers of the triangular faces of the polyhedron.

Mahieux,[18] Abel and Begault,[19] Essert,[20] and Becker *et al.*[21]

There is another group of existing studies that address the directional measurement problem by using a directional receiver or beamforming microphone array to look in one direction at a time. These include those by Thiele,[22] Broadhurst,[23] Nishi,[24] and Okubo *et al.*[25]

The present measurement system, described by the authors in Ref. 26, is also of this latter type, and is capable of detailed directional and temporal analysis. This system makes use of 32-microphone spherical arrays to achieve a highly directional beam pattern. Directional gains of over 14 dB are realized in 60 steering directions, with a 28° beamwidth. Responses can be analyzed instantaneously at each sample, or over arbitrary time windows. The measurements presented in this paper describe its application in three real rooms, and demonstrate how directional information is determined.

II. MEASUREMENT SYSTEM

The measurement system used in this work has been described in detail in Refs. 8 and 26. A block diagram summarizing the system is shown in Fig. 1. The system works as follows: A sound field is established in a room by playing a test signal over an omnidirectional loudspeaker. Simultaneously, a recording is made of the signals arriving at the microphones comprising the detecting array. The array has 32 elements, but in the present state of the system only 8 can be measured at a time, necessitating the repetition of the measurement four times. It is assumed that the system (room) being measured is time invariant so that the identical sound field is established each of the four times. By using a maximum-length-sequence (MLS) to drive the loudspeaker, cross-correlation of the test signal with each microphone signal results in the traditional (omnidirectional) room impulse responses at the array microphones.[27] Other stimuli may of course be used to measure the omnidirectional impulse responses.[28] Through subsequent beamforming in software, the array output is obtained in each of 60 steering directions. These outputs are the directional impulse responses—the pressure response at the array center due solely to sound arriving from within the directional array aperture.

At the heart of the system is the spherical microphone array. The geometrical arrangement of the 32 array microphones is that of a geodesic sphere, shown in Fig. 2. This arrangement lends itself naturally to the 60 symmetric steering directions, through the triangular faces of the polyhedron defined by the microphone positions (as vertices). Due to the symmetry, only one set of beamformer weights need be designed, which can be used to obtain the array output for all 60 steering directions from one set of measurements. The beamformer weights were designed from a constrained optimization procedure, maximizing the array gain subject to a constraint on the white noise gain.[26,29] The constraint was derived from measured magnitude and phase variances among the microphones, and serves to incorporate robustness to microphone noise and mismatch into the design. As a result, the array achieves a very high directionality over a broad range of frequency.

Over a range of about 1.7 octaves, the directional gain of the beam is above 14 dB, and the 3-dB beamwidth is 28°. As a comparison, the familiar first-order hypercardioid pattern has a directional gain of 6 dB and a 3-dB beamwidth of 105°.[4] The position in frequency of the usable range of the array scales with the radius of the sphere (the breadth remains 1.7 octaves). To cover the range of frequencies from 300 to 3300 Hz, two arrays of different sizes were constructed. The construction of both employed 6-mm electret microphones (Panasonic WM-61A102B) mounted onto thin (1.6 mm) stainless steel rod frames. The arrays were both fitted with the same size support base to allow precise placement and repositioning. In replacing one array with the other to repeat a measurement over a different frequency range, the array centers were easily positioned at the same location. The larger of the arrays has a diameter of 48 cm and covers the range from 300 to 1000 Hz; the smaller has a diameter of 16 cm and operates over 1000–3300 Hz. Plots of the beam pattern (averaged over the usable frequency range) and of the directional gain versus frequency for both arrays are shown in Fig. 3. These were calculated from the beamformer filters, using simulated anechoic data. The performance of the ar-

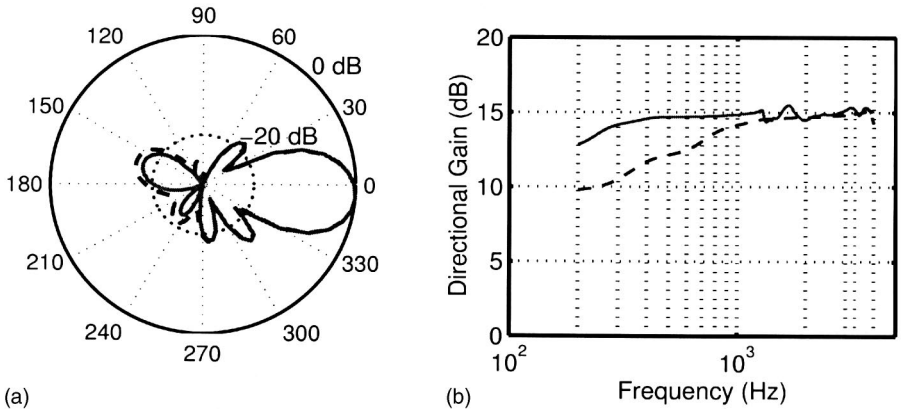


FIG. 3. Array beam properties, computed from the beamformer coefficients: (a) wide-band beam patterns for 16-cm array (dashed line) and 48-cm array (solid line), in the plane which passes through the primary steering direction, microphone 1, and the center of the array (see Fig. 2). The steering direction is at 0°, microphone 1 is at 21°. The asymmetry of the patterns are due to the asymmetry of the array with respect to this plane. (b) Directional gain versus frequency for 16-cm array (dashed line) and 48-cm array (solid line).

rays was previously demonstrated in anechoic and reverberant chambers, and was found to be consistent with theoretical predictions.[26]

The control computer was equipped with a multichannel sound card (Echo Audio Layla24), which was used to play the MLS and simultaneously record from eight of the array microphones. The omnidirectional loudspeaker (B&K Omnisource Type 4295) was driven by a power amplifier (Bryston 2B), and all playback/recording was controlled from software (Syntrillium Software CoolEdit SE). It is true that the loudspeaker becomes directional at higher frequencies, but up to 3300 Hz (the highest frequency of interest in this paper), it satisfies ISO standard requirements for omnidirectionality.[30] The microphone signals were sampled at a rate of 44.1 kHz but were downsampled prior to beamforming. The low-frequency array operates at a sampling rate of 2756.25 Hz, and the high-frequency array at a sampling rate of 11 025 Hz.

III. SYSTEM OUTPUT AND ANALYSIS

As mentioned above, the time of arrival of discrete reflections is frequently identifiable from a room impulse response. With traditional omnidirectional responses, however, no information regarding from which direction the reflection arrives is available. In contrast, when inspecting a directional impulse response as measured with the present system, the direction of arrival of the reflection is clear. Figure 4 illustrates these concepts for a measurement made in an anechoic chamber, in the presence of two parallel planar reflectors. On the omnidirectional response [panel (a)], the arrival time of the reflections is identifiable (7.5 ms), but it is not evident that two reflections arrive simultaneously. The directional measurements [panels (b)–(d)] clearly resolve the directions of arrival of both reflections, even though they arrive simultaneously.

The impulse response can be integrated to deduce the energy arriving at the receiving position over a given portion of the room decay. The energy E in a pressure impulse response $h(n)$ from time sample n_1 to time sample n_2 is proportional to the integral (sum) of the squared response:

$$E \propto \sum_{n=n_1}^{n_2} h^2(nT_s), \tag{1}$$

where $T_s = 1/F_s$ is the sampling period of a sequence sampled at rate F_s. This is the ''incident'' or ''arriving'' energy, assumed to be carried towards the receiving point by a plane wave. This quantity is proportional to the potential, and therefore total, energy of this plane wave. It is not the same as the total acoustic field energy at the receiving position. By computing the incident energy in each of a set of measured directional responses, the arriving energy versus direction can be graphed and/or analyzed. A normalized set of incident energies can be obtained from

$$e_i = \frac{\Sigma_{n=n_1}^{n_2} h_i^2(nT_s)}{\max_i\{\Sigma_{n=n_1}^{n_2} h_i^2(nT_s)\}}, \tag{2}$$

where $h_i(n)$ is the directional impulse response for direction i.

Three-dimensional plots of e_i versus direction yield clear pictures of the direction of arrival of the incident sound energy (the directions between the 60 measured directions can be interpolated). This can be of use in identifying directions of significant or insignificant sound arrival. See Fig. 5, which depicts the incident energy distribution for the same parallel reflector measurement as discussed in Fig. 4. Panel (a) of the figure depicts the geometry of the measurement situation. Panel (b) shows the arriving energy versus all directions at once. Azimuth is indicated horizontally; elevation vertically along the curves. Panels (c)–(e) show three different views of the same surface, which plots the arriving energy as a function as direction. The radius and gray level of the surface in a particular direction indicate the energy incident from that direction. Notice that the direct sound and reflection pair are easily resolved.

The overall isotropy of the sound field can be assessed qualitatively from the plot, but also numerically from directional quantifiers. In particular, the directional diffusion and the peak-to-average level difference of the sound field can be easily computed.

The directional diffusion d is defined as[22,31]

$$d = \left(1 - \frac{\mu}{\mu_0}\right) \times 100\%, \tag{3}$$

with μ computed from

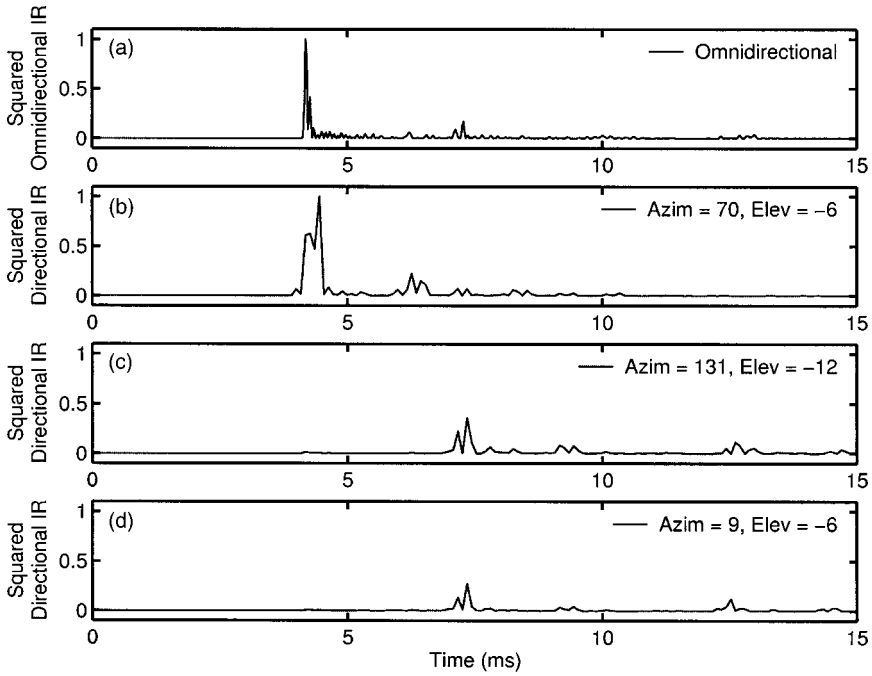


FIG. 4. Squared impulse responses measured for parallel reflectors in anechoic chamber (all self-normalized to have a maximum of unity): (a) Omnidirectional impulse response at one array microphone. Note the two symmetric reflections arriving at 7.5 ms are not resolvable in direction. (b) Directional impulse response in direction of source. The direct sound is picked up, but the pair of reflections is not. The peak at 6 ms is attributed to scattering from the loudspeaker support stand. (c) Directional impulse response in direction 61° to one side of the source direction. One lateral reflection is seen arriving at 7.5 ms. (d) Directional impulse response in direction 61° to the other side of the source direction. The other simultaneously arriving lateral reflection is seen at 7.5 ms.

$$\mu = \frac{1}{\langle e \rangle} \sum_{i=1}^{60} |e_i - \langle e \rangle|, \tag{4}$$

where e_i is computed from the measured directional impulse responses via Eq. (2), and $\langle e \rangle$ is the mean incident energy over all directions, given by

$$\langle e \rangle = \frac{1}{60} \sum_{i=1}^{60} e_i . \tag{5}$$

The quantity μ is similar to the variance of the set of directional incident energies; μ_0 is the value for single plane-wave (anechoic) incidence and is determined by the beam pattern of the detector. d is merely a renormalization to change the range of values from 0% (anechoic) to 100% (isotropic). The system has been previously used to find a value of 21% in an anechoic chamber, and of 91% in a reverberation chamber.[26]

Since the incident energy in each directional response depends on the width of the receiver pattern, so too does d, and hence the same sound field measured with different receivers could yield differing values. With the present system, the same receiver is used in all cases, so comparisons among different measurements can be safely made.

The directional peak-to-average level difference L_{PA} of the sound field is obtained from

$$L_{PA} = 10 \log \left[\frac{\max(e_i)}{\langle e \rangle} \right], \tag{6}$$

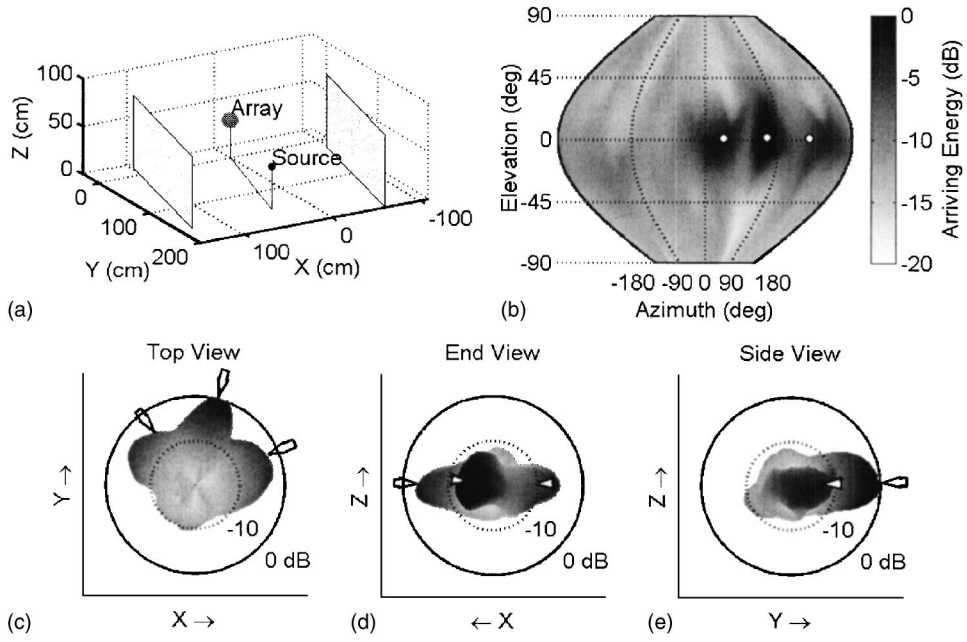


FIG. 5. Results for parallel reflectors in anechoic chamber, measured with high-frequency array (1000–3300 Hz). (a) Geometry of the measurement setup. The line joining source and array is parallel to and equidistant from each reflecting panel. (b) Arriving sound energy as a gray scale map (azimuth horizontal, elevation vertical, measured from the array center). Directions of expected sound incidence (direct sound, a pair of first-order reflections) are indicated by white circles. (c)–(e) Three different views of the same measured data, plotted three-dimensionally versus azimuth and elevation. The radius and gray level of the surface are proportional to the arriving energy (in dB) in each direction. The directions of expected sound arrival are indicated by white arrows pointing inward.

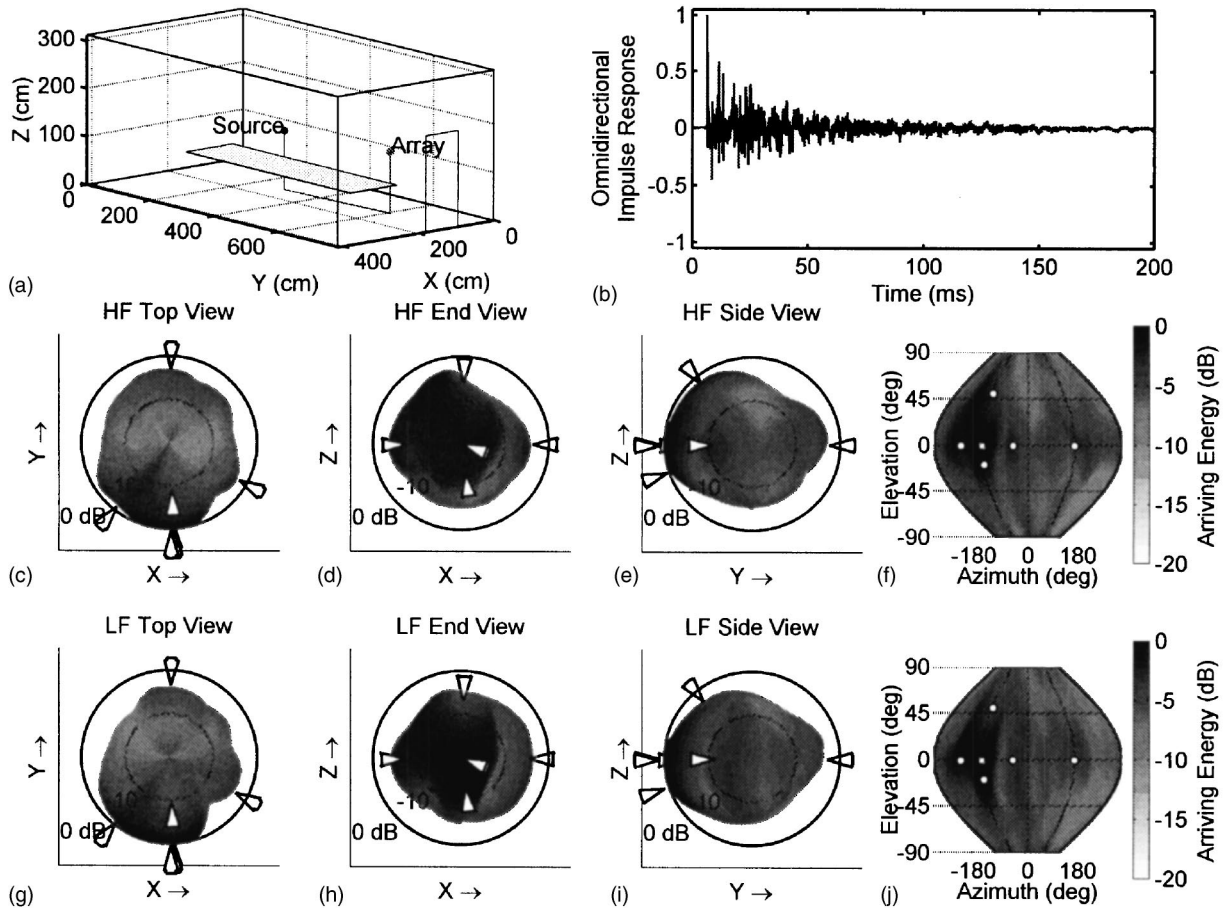


FIG. 6. Full decay time measurement in a small meeting room: (a) Geometry of room and source and array positions. (b) Omnidirectional room impulse response measured with one of the array microphones. (c)–(e) Three views of incident energy versus direction measured with the high-frequency array (1000–3300 Hz). (f) Arriving energy versus all directions measured with the high-frequency array. (g)–(i) Three views of incident energy versus direction measured with the low-frequency array (300–1000 Hz). (j) Arriving energy versus all directions measured with the low-frequency array.

the result being expressed in decibels. This is a quantity that is directly analogous to the directivity index (of a beam pattern), commonly used to describe the directionality of transducers and transducer arrays. It is logical therefore to refer to L_{PA} as the "anisotropy index" at the measurement point; it increases with increasing anisotropy of the sound field.

The anisotropy index takes on a minimum value of 0 dB when the sound field is perfectly isotropic. The microphone arrays used in the system have a beam pattern with a directivity index of about 14 dB, which is therefore the upper theoretical limit we can expect to measure. The system has been previously used to measure values of 12.8 and 1.5 dB in an anechoic and a reverberation chamber, respectively.[26]

It should be noted that by considering different values for n_1 and n_2 in Eq. (2) it is possible to plot the incident energy distribution and compute directional diffusion and anisotropy index at each time instant or for any given time range.

IV. FULL DECAY TIME

The length of the impulse responses measured with the system should be at least as long as the reverberation time of the room. This ensures detection of all arriving energy at the receiving point, from the direct arrival until the sound field has decayed by 60 dB. By integrating the squared impulse response over this entire time range, the energy in the full decay of the room is considered.

A. Small meeting room

Full decay time results for a measurement in a typical meeting room are shown in Fig. 6. The geometry of the room and source and array locations are shown in panel (a). The walls were wood-paneled, the floor had carpet over concrete, and the ceiling was acoustical tile. The room was fitted with a large table and 12 padded chairs (chairs not shown). The room dimensions were 4.14 m wide, 7.91 m long, and 3.13 m high. The broadband (300–3300 Hz) reverberation time was 360 ms. The distance from source to array was 3.14 m. Panel (b) shows the omnidirectional room impulse response measured with one of the array microphones. Panels (c)–(f) are the directional incident energy results measured with the high-frequency array (1000–3300 Hz): three views of the incident energy versus direction surface, and the gray map depicting all directions at once. Panels (g)–(j) are corresponding plots for the low-frequency array measurement (300–1000 Hz).

As a qualitative observation, notice that the distribution was not isotropic. The directions of greater arriving energy are more or less the same as the indicated early-arriving (and

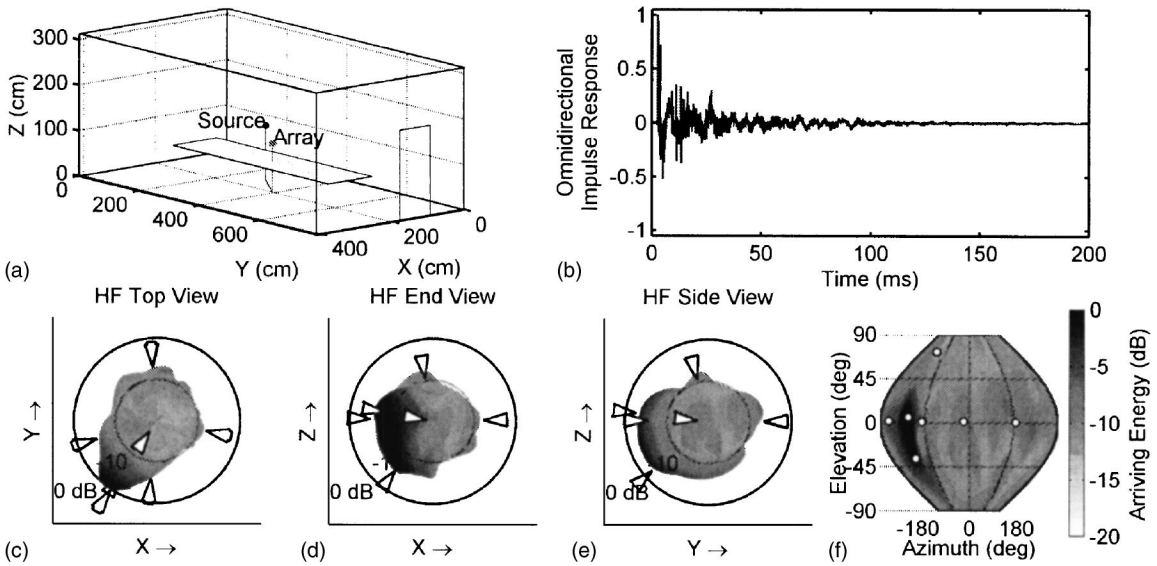


FIG. 7. Full decay time measurement with the high-frequency (1000–3300 Hz) array in a small meeting room, source in the same position as for the measurement shown in Fig. 6. Layout is the same as Fig. 6, panels (a)–(f). Note the increased anisotropy in this measurement as compared to that presented in Fig. 6.

strong) discrete reflections. Quantitatively, the value of the directional diffusion was 70% for the high-frequency measurement, and 73% for the low-frequency measurement. The value of the anisotropy index was 4.0 dB for the high-frequency measurement, and 4.2 dB for the low-frequency measurement.

1. Same source position, different receiver position

For the same source position in the same room, the array was moved to a new position and the measurement repeated. Care was taken to ensure that the source remained outside the near-field of the array, where the beamforming filters designed for plane-wave incidence would be inappropriate.[32] The full decay time results for the high-frequency measurement only are shown in Fig. 7, the layout being the same as panels (a)–(f) of Fig. 6. The low-frequency measurement is similar, and is omitted for brevity. Compared to the distribution in Fig. 6, the sound arriving at the array position for this case is clearly more anisotropic. The contribution from the direction of the source is larger, which is expected since the source-to-array distance was only 1.25 m for this measurement (as compared to 3.14 m previously). This observation is reinforced by the directional diffusion value of 59% and the anisotropy index value of 7.8 dB, as compared to 70% and 4.0 dB previously.

2. Same source-to-receiver separation

Full decay time results for a third measurement in the same meeting room are shown in Fig. 8, the layout being the same as Fig. 6, panels (a)–(f). In the interest of conciseness, only the high-frequency results are discussed. For this measurement, the source and array are in positions different than those used for the measurement presented in Fig. 6, but the distance between them is the same. This means that the direct sound strength is the same in both cases, and therefore any differences in the isotropy are due to the room (reflections, scattering, etc.). Observe that there are differences in the ''roundness'' of the plot—evident in the comparison of panel (c) in particular. This observation is borne out by the value of anisotropy index (6.5 dB, compared to 4.0 dB previously), but interestingly not by the value of directional diffusion (71%, compared to 70% previously). The directional diffusion fails to indicate the difference between these sound fields.

B. Videoconferencing room

Full decay time results for a measurement in a small videoconferencing room are shown in Fig. 9. The layout of the figure is the same as Fig. 6. The room was 7.23 m wide, 8.33 m long, 3.01 m high, had plaster walls, carpet over concrete floor, and acoustical tile ceiling. The room was fitted with a large table, several padded chairs (not shown), a cabinet housing videoconferencing equipment, and had a large heavy curtain covering the back wall. The broadband (300–3300 Hz) reverberation time was 400 ms. The source and array were positioned 2.03 m apart.

Notice again that the distribution of sound was not isotropic, and that the directions of the early-arriving discrete reflections shape the overall surface. A large proportion of the arriving energy is carried in these reflections. The value of the directional diffusion was 58% for the high-frequency measurement, and 57% for the low-frequency measurement. The value of the anisotropy index was 7.4 dB for the high-frequency measurement, and 8.6 dB for the low-frequency measurement.

C. Lecture theater

Full decay time results for a measurement in a small lecture theater are shown in Fig. 10. The layout of the figure is the same as Fig. 6. The hall had plaster walls, a sloping, carpeted floor with 12 rows of upholstered, padded seats, and a sculpted ceiling. The room was 12.62 m wide and 16.45 m

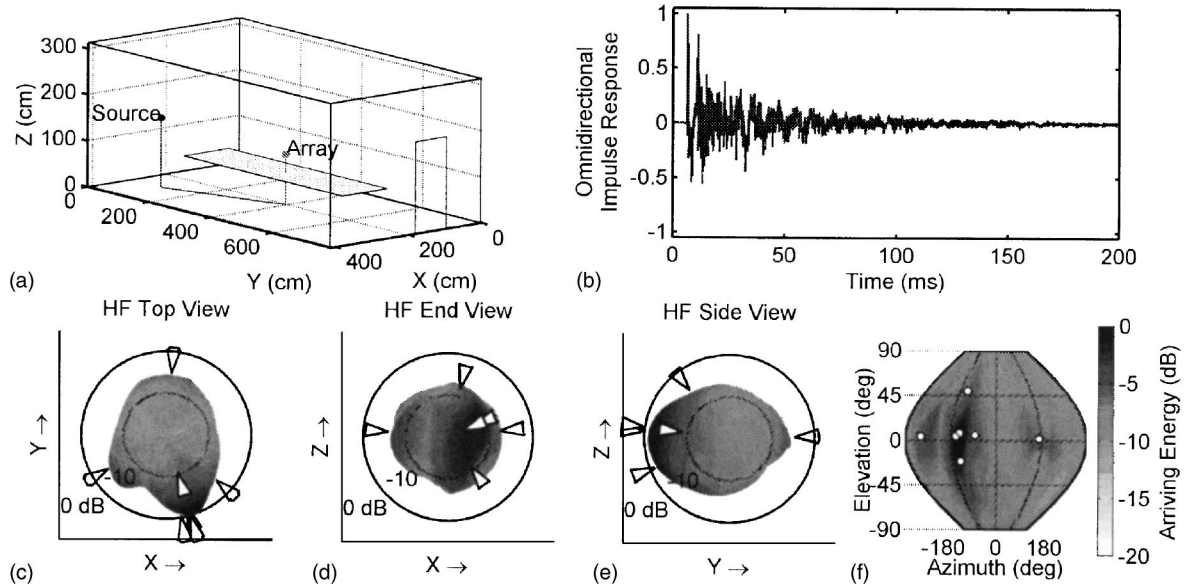


FIG. 8. Full decay time measurement with the high-frequency (1000–3300 Hz) array in a small meeting room, source-to-array distance the same as for the measurement shown in Fig. 6. Layout is the same as Fig. 6, panels (a)–(f). Note the increased anisotropy in this measurement as compared to that presented in Fig. 6.

front-to-back. The height was 5.69 m at the front and 2.75 m at the back. The side walls were parallel. The broadband (300–3300 Hz) reverberation time was 600 ms, and the source and array were situated 4.01 m apart.

This room was larger than the other two, and the sound field was less diffuse, as evidenced by the directionality indicated in the figure. This is reinforced by the numerical values of directional diffusion and anisotropy index: respec-

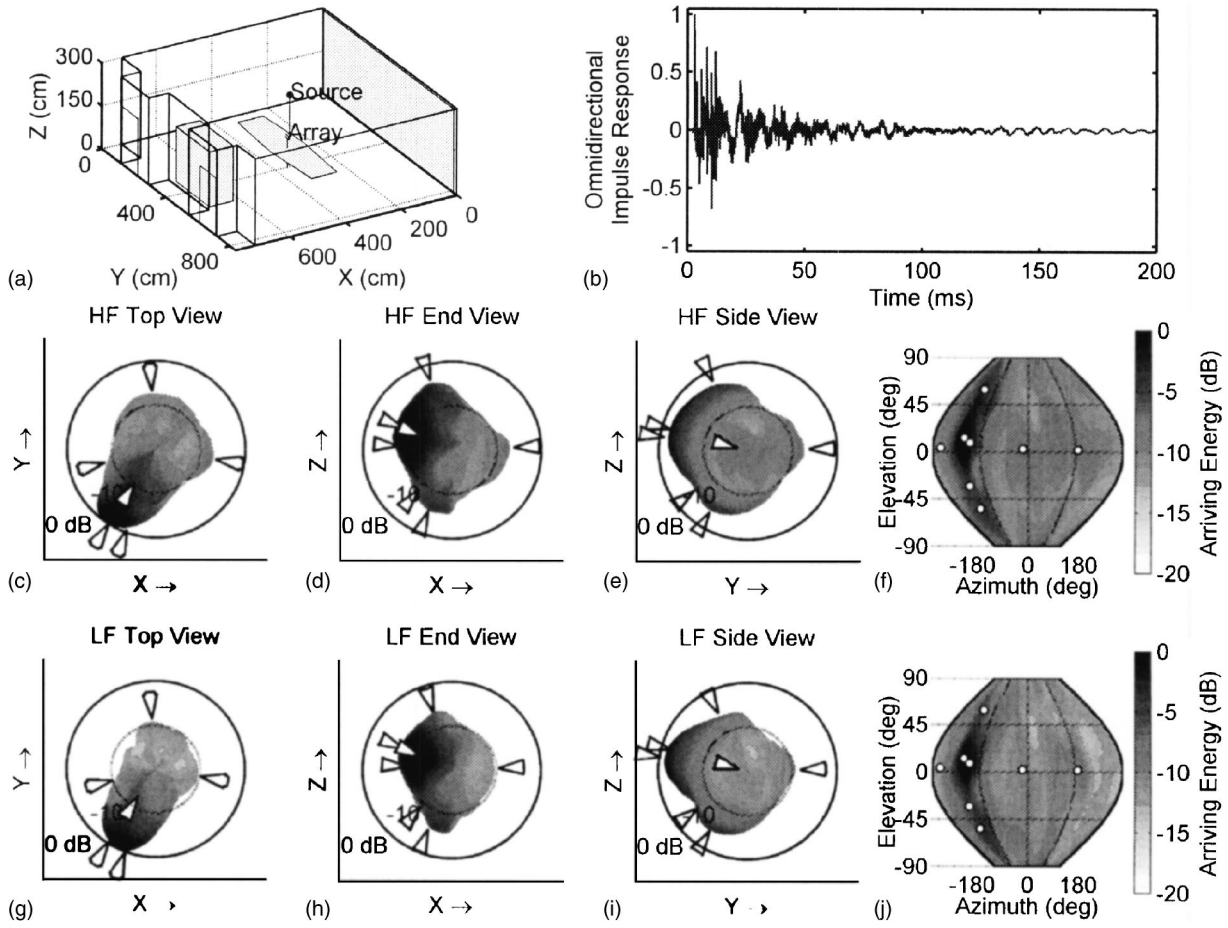


FIG. 9. Full decay time measurement in a videoconferencing room. Layout is the same as Fig. 6.

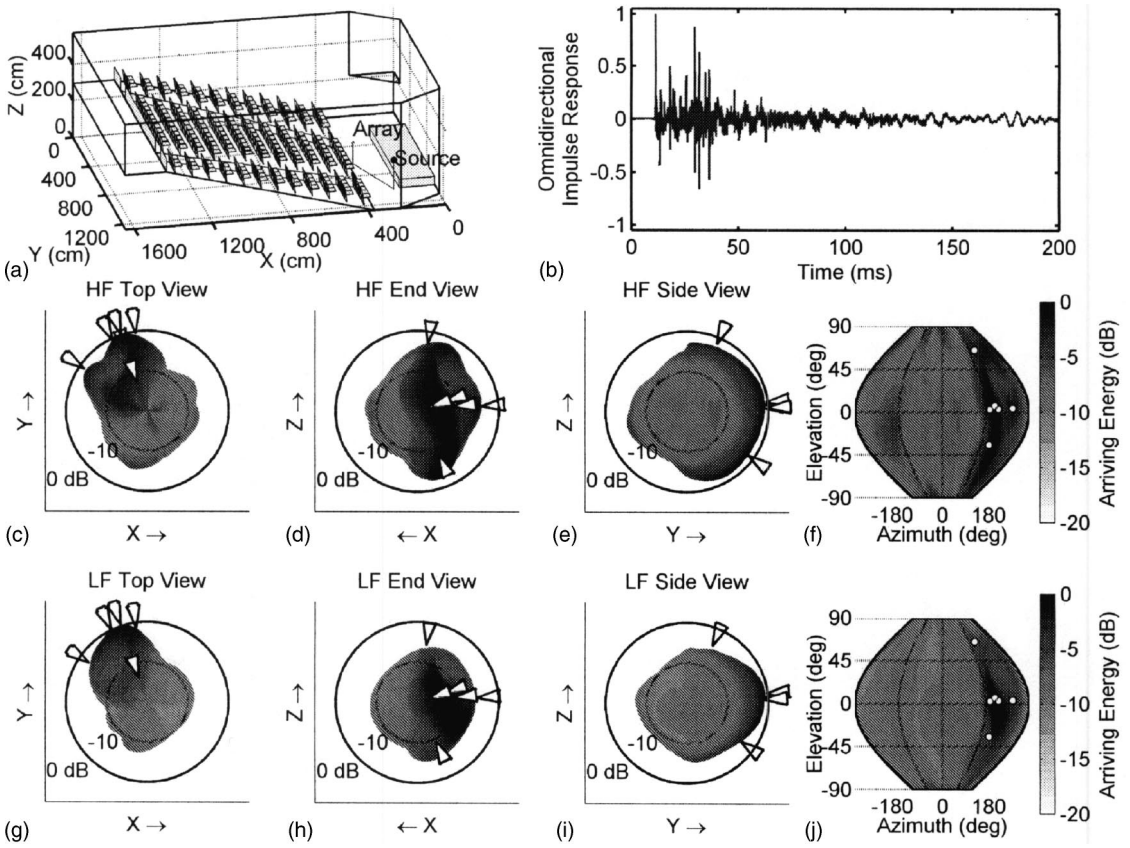


FIG. 10. Full decay time measurement in a small lecture theater. Layout is the same as Fig. 6.

tively 61% and 5.5 dB for the high-frequency measurement, 62% and 6.8 dB for the low-frequency measurement.

V. TIME WINDOWING

Knowledge of the directional impulse responses is particularly useful for examining the isotropy of the sound field over restricted time ranges. For example, it can be of interest to gate out the discrete early arrivals, and examine only the reverberant tail. This was done for measurements in the three rooms. The high-frequency measurements only shall be discussed; the low-frequency measurements are similar.

A. Small meeting room

Figure 11 shows several restricted time range results for a measurement with the high-frequency array in the small meeting room described in Sec. IV A. Panel (a) of the figure shows the room geometry and the source and array locations. Panel (b) shows the omnidirectional room impulse response measured at one of the array microphones. Panel (c) shows

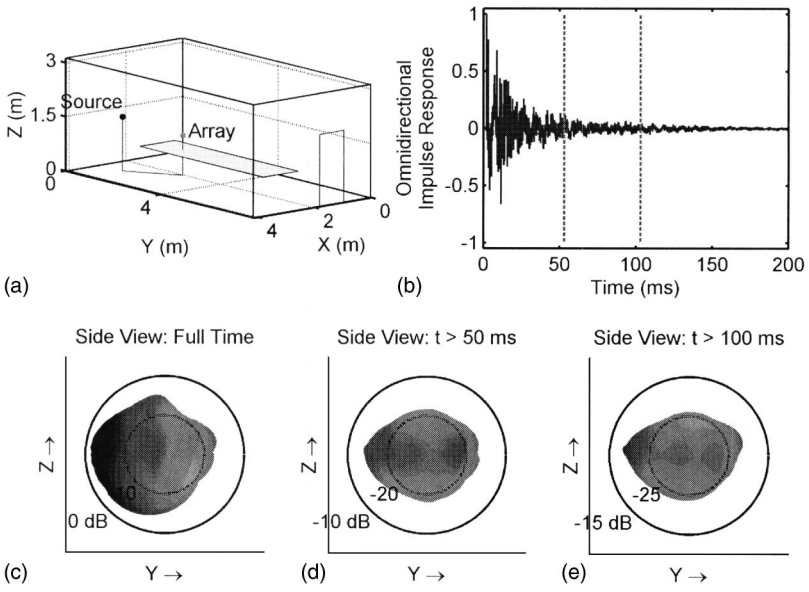


FIG. 11. Restricted time range results for a measurement in the small meeting room, made with the high-frequency array (1000–3300 Hz). (a) Geometry of the room depicting source and array locations. (b) Omnidirectional impulse response measured at one of the array microphones. The vertical dotted lines indicate time cutoff points, 50 and 100 ms after the direct sound arrival. (c) One view of the arriving energy versus direction surface, including energy arriving over the full time of decay of the room. (d) Same as panel (c), but the energy arriving in the initial 50 ms is excluded. (e) Same as panel (c), but the energy arriving in the initial 100 ms is excluded.

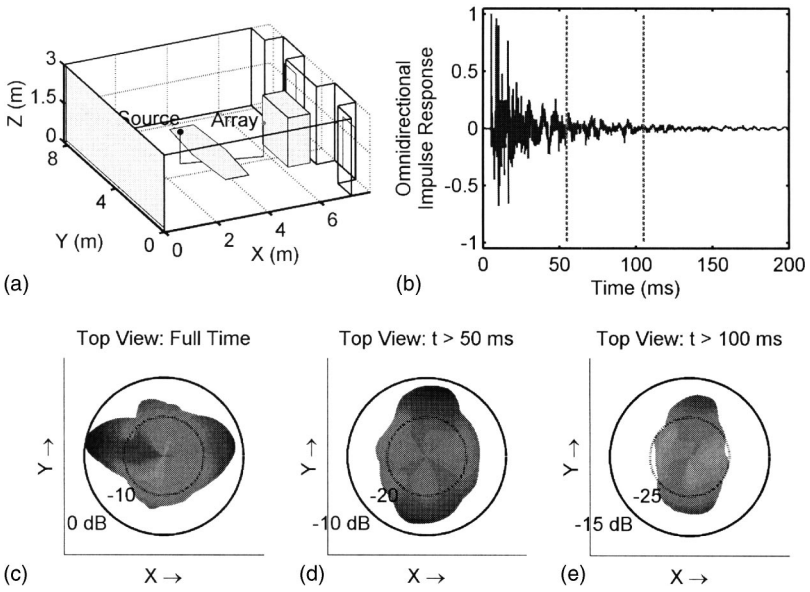


FIG. 12. Restricted time range results for a measurement in the videoconferencing room, made with the high-frequency array (1000–3300 Hz). Layout is the same as Fig. 11.

one view of the plot of full-decay-time arriving energy. Panel (d) shows a similar plot, but the energy arriving in the first 50 ms after the direct arrival is not included. This is the so-called ''late'' response only. Panel (e) is similar, but omits the first 100 ms of the response after the direct arrival.

Notice from panel (c) that in considering the full decay time of the room, there are significant sound contributions from the source direction (from the left) and from the reflections from above and below. The value of directional diffusion in this case is 64%, and that of anisotropy index is 6.2 dB. The shape of the surface in panel (d) indicates that having gated the early arrivals out, the late sound is distributed more uniformly, but evidently is elongated along the long dimension of the room. The numerical indicators do indicate increased isotropy, having values of 77% for directional diffusion and 3.4 dB for anisotropy index. This elongated shape is even more pronounced in the plot of the later time range ($t>100$ ms), seen in panel (e), which appears less-evenly distributed in direction. For this case, directional diffusion is 73% and anisotropy index is 4.4 dB, indeed indicating a more anisotropic distribution. There is longer-persisting sound propagating back-and-forth the length of the room, after having been absorbed in other directions.

B. Videoconferencing room

Figure 12 shows the restricted time range results for a measurement in the small videoconferencing room described in Sec. IV B, using the high-frequency array. The layout of the figure is the same as Fig. 11.

By comparing panels (c) and (d), notice that gating the early response results in a more uniform distribution of sound with direction, but not entirely so. The sound arriving in the late time period is in some ways more evenly distributed, but there is a strong back-and-forth component arriving laterally. These observations are reflected in the quantifiers' directional diffusion and anisotropy index, which respectively were 67% and 6.7 dB for the full response, and 74% and 4.3 dB for the late response. It is interesting to note that for the later response ($t>100$ ms), the quantifiers again indicate an increasing anisotropy (taking on values of 69% and 5.1 dB). As time progresses, sound is absorbed more quickly by the front and back walls than by the painted plaster side walls.

C. Lecture theater

Figure 13 shows the restricted time range results for a measurement in the small lecture theater described in Sec. IV C, using the high-frequency array. Panels (a)–(e) of the figure are as in Figs. 11 and 12, plus an additional panel is added, (f), which shows results excluding the first 150 ms of arriving energy.

This extra time range was included to reveal the increasing anisotropy in this sound field as time passed. This behavior was not overly evident from inspecting the plots in panels (c)–(e), which do indicate increasing isotropy initially, as before. However, it can be seen that as the sound field decays, there are persistent arrivals back-and-forth across the room (between the parallel side walls). The directional diffusion and anisotropy index indicate this behavior as well, respectively having values of 61% and 5.5 dB for the full decay time, 84% and 2.3 dB excluding the first 50 ms, 86% and 2.7 dB excluding the first 100 ms, and 81% and 3.8 dB excluding the first 150 ms. To observe this behavior in this larger room, more time is required to allow for sufficient reflections (i.e., sound absorptions) than in the smaller rooms.

VI. EVOLUTION OF THE SOUND FIELD

By considering the measured impulse responses over different ranges, it is possible to investigate the temporal evolution of the sound field, and to isolate discrete reflections in time and direction. For instance, computing the energy incident from each steering direction at each sample enables visualization of evolution of the sound field. Alternatively, by considering a short sliding time window, individual reflections can be located.

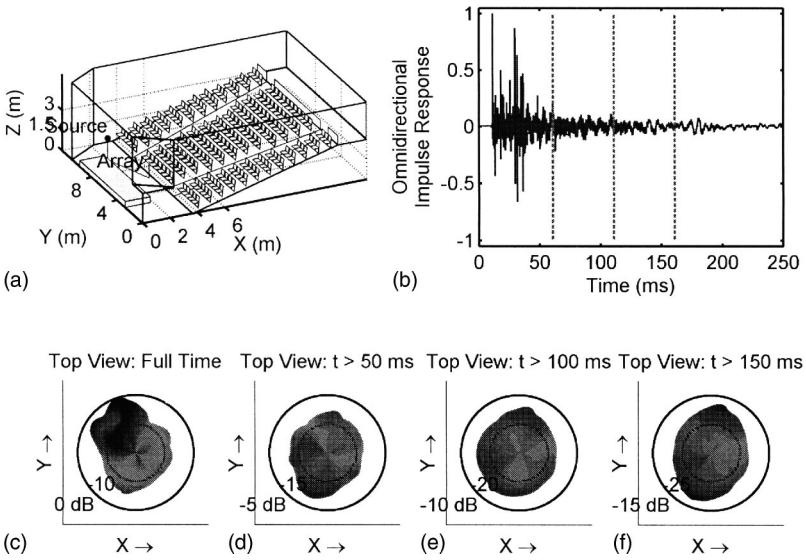


FIG. 13. Restricted time range results for a measurement in the lecture theater made with the high-frequency array (1000–3300 Hz). Layout of panels (a)–(e) is the same as Fig. 11. Panel (f) shows the arriving energy versus direction, excluding the energy arriving in the first 150 ms after the direct sound arrival.

Figure 14 shows results for a measurement made in the small lecture theater with the high-frequency array. Panel (a) shows the room geometry and the source and array locations. Panel (b) shows the omnidirectional room impulse response measured at one of the array microphones. Panel (c) shows the incident energy versus direction surface for the full decay time of the room. The plot has been superimposed on a three-dimensional drawing of the room. On a computer screen, this enables rotation and zooming to quickly identify room surfaces that cause reflections, for instance. Panel (d) shows likewise for the direct sound arrival, obtained by integrating the impulse responses from $t=11$ ms to $t=13$ ms only. Panel (e) shows the reflection arriving at time $t=30$ ms, which is the first-order reflection off the ceiling. The incident directions of the direct sound and this reflection were not known *a priori*—they followed from the measurements as output.

VII. CONCLUSIONS

The spherical array measurement system enables detailed analysis of reverberant sound fields. The measurements made in the three rooms discussed herein reveal not only the anisotropy of the sound fields, but details of that anisotropy in terms of time, direction, and overall quantification. The directions of the strong, early-arriving reflections are confirmed to be important, and are easily determined without the need for tedious measurement of the room geometry and inspection of omnidirectional impulse responses. It has been observed that source-to-array separation is important. However, for the same source-to-array separation, differences in diffusion related to position in the room were observed. It has further been noticed that the sound field in the larger room was less diffuse than those in the smaller rooms.

With this system, the time evolution and decay of the reverberant sound field can be investigated. Initially, the room reflections serve to ''build up'' the sound field at the receiving point in a manner which increases the isotropy. It was subsequently observed that the absorption of sound at different rates in different directions can lead to increasing anisotropy in the sound field as it decays.

All that is required to extend the present system to have the capability of analyzing nonreverberant (i.e., time vary-

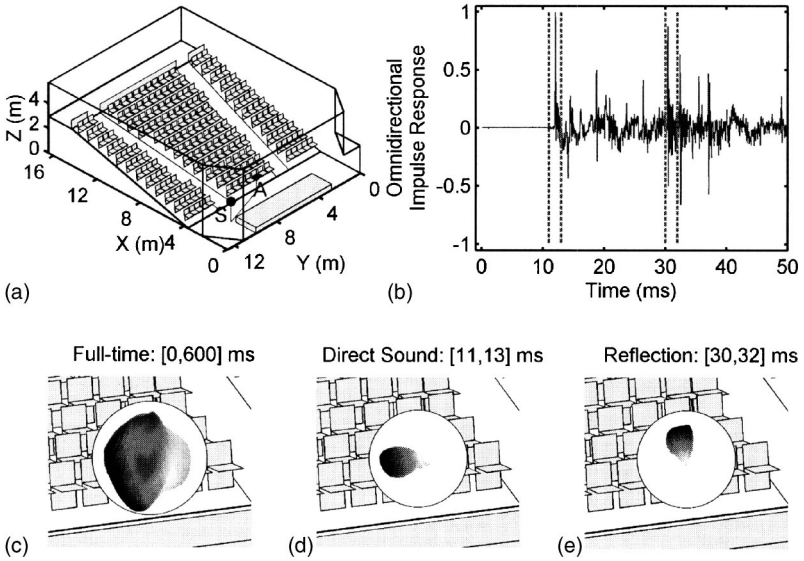


FIG. 14. Time-windowed results for a measurement in the lecture theater made with the high-frequency array (1000–3300 Hz). (a) Geometry of the room, depicting the source (S) and array (A) positions. (b) Initial part of omnidirectional room impulse response measured with one of the array microphones. The vertical dotted lines indicate the short-time windows used to resolve the direct sound and one reflection. (c) Arriving energy versus direction plot superimposed on a drawing of the room, zoomed in on the array position. Energy integrated over full decay of room. (d) Same as panel (c), but incident energy integrated from $t=11$ ms to $t=13$ ms only. This shows the arrival of the direct sound. (e) Same as panel (c), but incident energy integrated from $t=30$ ms to $t=32$ ms only, revealing the arrival direction of a ceiling reflection.

ing) sound fields is the addition of 24 more simultaneous data acquisition channels. Such hardware is easily accessible nowadays, and such a modification would enable study of ambient sound fields, including human talkers or machinery noise, for example. It is hoped that the directional information measurable with the system can be of value in a range of application areas.

ACKNOWLEDGMENTS

The authors would like to thank to René St. Denis and John Quaroni at NRC for technical assistance. One author (BG) would additionally like to thank Dr. John Vanderkooy at the University of Waterloo for financial support and for many helpful discussions and suggestions.

[1] L. L. Beranek, *Music, Acoustics & Architecture* (Wiley, New York, 1962).
[2] L. Cremer and H. A. Müller, *Principles and Applications of Room Acoustics*, Vol. 1, translated by T. J. Schultz (Applied Science, Essex, UK, 1982).
[3] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1989).
[4] G. W. Elko, ''Superdirectional Microphone Arrays,'' Chap. 10 of *Acoustic Signal Processing for Telecommunication*, edited by S. L. Gay and J. Benesty (Kluwer Academic, Norwell, MA, 2000).
[5] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, ''Measurement of correlation coefficients in reverberant sound fields,'' J. Acoust. Soc. Am. **27**, 1072–1077 (1955).
[6] C. T. Morrow, ''Point-to-point correlation of sound pressures in reverberation chambers,'' J. Sound Vib. **16**, 29–42 (1971).
[7] T. J. Schultz, ''Diffusion in reverberation rooms,'' J. Sound Vib. **16**, 17–28 (1971).
[8] B. N. Gover, ''Development and use of a spherical microphone array for measurement of spatial properties of reverberant sound fields,'' Ph.D. thesis, Department of Physics, University of Waterloo, Waterloo, Canada, 2001.
[9] F. J. Fahy, *Sound Intensity*, 2nd ed. (Spon, London, 1995).
[10] A. Abdou and R. W. Guy, ''Spatial information of sound fields for room-acoustics evaluation and diagnosis,'' J. Acoust. Soc. Am. **100**, 3215–3226 (1996).
[11] S. Oguro, M. Anzai, H. Suzuki, and T. Ono, ''A three-dimensional sound intensity probe,'' J. Acoust. Soc. Am. **91**, 2370(A) (1992).
[12] R. W. Guy and A. Abdou, ''A measurement system and method to investigate the directional characteristics of sound fields in enclosures,'' Noise Control Eng. J. **42**, 8–18 (1994).
[13] K. Hori, ''4 microphones power advanced, 3-dimensional sound intensity measuring system,'' J. Electron. Eng. February, 47–49 (1994).
[14] T. Peltonen, T. Lokki, B. Gouatarbès, J. Merimaa, and M. Karjalainen, ''A system for multi-channel and binaural room response measurements,'' Convention Paper 5289, presented at the 110th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, May 2001.
[15] J. Merimaa, T. Lokki, T. Peltonen, and M. Karjalainen, ''Measurement, analysis, and visualization of directional room responses,'' Convention Paper 5449, presented at the 111th Convention of the Audio Engineering Society, New York, September 2001.
[16] Y. Yamasaki and T. Itow, ''Measurement of spatial information in sound fields by closely located four point microphone method,'' J. Acoust. Soc. Jpn. (E) **10**, 101–110 (1989).
[17] K. Sekiguchi, S. Kimura, and T. Hanyuu, ''Analysis of sound field on spatial information using a four-channel microphone system based on regular tetrahedron peak point method,'' Appl. Acoust. **37**, 305–323 (1992).
[18] W. Täger and Y. Mahieux, ''Reverberant sound field analysis using a microphone array,'' in Proceedings ICASSP-97, IEEE International Conference of Acoustics, Speech, and Signal Processing, Munich, Germany, April 1997.
[19] J. S. Abel and D. R. Begault, ''Methods for room acoustic analysis using a monopole-dipole microphone array,'' in Proceedings InterNoise98, Paper 123. International Congress on Noise Control Engineering, Christchurch, New Zealand, November 1998.
[20] R. Essert, ''Progress in concert hall design—developing an awareness of spatial sound and learning how to control it,'' EBU Technical Review, Winter Issue (1997), pp. 31–39.
[21] J. Becker, M. Sapp, and O. Schmitz, ''Four-microphone-array measurement combined with geometrical room acoustic technique,'' Collected papers from the 2nd Forum Acusticum/137th Meeting of the Acoustical Society of America, Oldenburg, Germany, 1999.
[22] R. Thiele, ''Richtungsverteilung und Zeitfolge der Schallrückwurfe in Räumen,'' Acustica **3**, 291–302 (1953).
[23] A. D. Broadhurst, ''An acoustic telescope for architectural acoustic measurements,'' Acustica **46**, 299–310 (1980).
[24] T. Nishi, ''Relation between objective criteria and subjective factors in a sound field, determined by multivariate analyses,'' Acustica **76**, 153–162 (1992).
[25] H. Okubo, M. Otani, R. Ikezawa, S. Komiyama, and K. Nakabayashi, ''A system for measuring the directional room acoustical parameters,'' Appl. Acoust. **62**, 203–215 (2001).
[26] B. N. Gover, J. G. Ryan, and M. R. Stinson, ''Microphone array measurement system for analysis of directional and spatial variations of sound fields,'' J. Acoust. Soc. Am. **112**, 1980–1991 (2002).
[27] D. D. Rife and J. Vanderkooy, ''Transfer-function measurement with maximum-length sequences,'' J. Audio Eng. Soc. **37**, 419–444 (1989).
[28] S. Müller and P. Massarani, ''Transfer-function measurement with sweeps,'' J. Audio Eng. Soc. **49**, 443–471 (2001).
[29] H. Cox, R. M. Zeskind, and M. M. Owen, ''Robust adaptive beamforming,'' IEEE Trans. Acoust., Speech, Signal Process. **35**, 1365–1376 (1987).
[30] Product Data Sheet for OmniSource Sound Source Type 4295, Brüel & Kjær, available from www.bksv.com
[31] H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier Applied Science, London, 1991).
[32] J. G. Ryan, ''Criterion for the minimum source distance at which plane-wave beamforming can be applied,'' J. Acoust. Soc. Am. **104**, 595–598 (1998).

Plane-wave decomposition of the sound field on a sphere by spherical convolution

Boaz Rafaely[a)]
Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

(Received 16 May 2003; revised 14 July 2004; accepted 19 July 2004)

Spherical microphone arrays have been recently studied for sound analysis and sound recordings, which have the advantage of spherical symmetry facilitating three-dimensional analysis. This paper complements the recent microphone array design studies by presenting a theoretical analysis of plane-wave decomposition given the sound pressure on a sphere. The analysis uses the spherical Fourier transform and the spherical convolution, where it is shown that the amplitudes of the incident plane waves can be calculated as a spherical convolution between the pressure on the sphere and another function which depends on frequency and the sphere radius. The spatial resolution of plane-wave decomposition given limited bandwidth in the spherical Fourier domain is formulated, and ways to improve the computation efficiency of plane-wave decomposition are introduced. The paper concludes with a simulation example of plane-wave decomposition. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1792643]

PACS numbers: 43.60.Ac, 43.60.Fg, 43.55.Br [EJS] Pages: 2149–2157

I. INTRODUCTION

Reverberant sound fields have been widely studied, as they have a significant influence on the acoustic performance of enclosures in a variety of applications. For example, the intelligibility of speech in lecture rooms; the quality of music in auditoria; and the noise level in offices, are all affected by the enclosed sound field. Systems that measure reverberant sound fields or operate in such fields often require microphone arrays to achieve the high spatial detail imposed by the complexity of the sound field. Linear, circular, and planar microphone arrays have a well developed theoretical analysis,[1,2] and have been applied to speech enhancement and speaker tracking in conference rooms;[3,4] auralization of sound fields measured in concert halls;[2] and the study and modifications of sound fields in auditoria,[5,6] as a few examples.

Recently, spherical microphone arrays have been suggested for spatial beam forming,[7–9] sound recordings with high spatial detail,[10,11] and sound-field measurement and analysis.[12–15] The advantage of the spherical microphone array is the spherical symmetry which facilitates three-dimensional analysis. Although useful microphone array designs and applications have been presented in previous papers, they do not provide a theoretical analysis similar to that available for linear, circular, and planar arrays. In particular, we would like to have detailed spatial information on the sound field given the pressure on a sphere as measured by spherical microphone arrays. Such information could be beneficial in many of the applications described above, and in this work is achieved through plane-wave decomposition. The paper therefore introduces the theory of plane-wave decomposition, and in addition presents topics of practical importance such as the spatial resolution and the computation complexity of plane-wave decomposition. Two simplifying assumptions are made.

(i) The sound pressure on the entire sphere is known. Although this is usually not true in practice, this assumption simplifies the development of the theoretical results. The reader is referred to Rafaely[16] and references within for discussion of spatial sampling with application to spherical microphone arrays.
(ii) The sound field is composed of plane waves. This will be approximately true in reverberant sound fields for waves which have traveled sufficient distance from their source.

Given these assumptions, we show in Sec. III, following a review of the spherical Fourier transform and spherical convolution, that the relation between the pressure on a sphere and the complex amplitude of the plane waves composing the sound field around the sphere is that of spherical convolution. In other words, plane-wave decomposition is achieved by convolving the pressure on a sphere with another function which depends only on frequency and the sphere radius. This is in contrast to linear and planar array configurations, where the directivity of the sound field is given by the Fourier transform of the pressure on the line or plane.[1] Since perfect plane-wave decomposition requires an infinite spherical harmonic order, an expression for the spatial resolution in the more practical case of finite harmonic order is derived in Sec. IV and compared to the spatial resolution of a finite aperture planar array. The computation complexity of plane-wave decomposition and ways to reduce it are introduced in Sec. V, and the paper concludes with a simulation example of plane-wave decomposition in Sec. VI.

II. SPHERICAL FOURIER TRANSFORM

The spherical Fourier transform,[17] or spherical harmonics decomposition,[1] is reviewed in this section. In particular,

a)Electronic mail: br@ee.bgu.ac.il

the Fourier transform and convolution on the sphere, and some properties of spherical harmonics are presented which are then used in the derivations that follow. The standard Cartesian (x,y,z) and spherical (r,θ,ϕ) coordinate systems are used throughout this paper.[18] Consider a function $f(\theta,\phi)$ which is square integrable on the unit sphere, then the spherical Fourier transform of f, denoted by f_{nm} is given by[17]

$$f_{nm}=\int_{\Omega\in S^2} f(\theta,\phi)Y_n^{m*}(\theta,\phi)d\Omega=\mathcal{S}\{f(\theta,\phi)\}, \tag{1}$$

$$f(\theta,\phi)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} f_{nm}Y_n^m(\theta,\phi)=\mathcal{S}^{-1}\{f_{nm}\}, \tag{2}$$

where $\mathcal{S}$ and $\mathcal{S}^{-1}$ represent the forward and inverse spherical Fourier transforms respectively, "*" represents complex conjugate, $\Omega=(\theta,\phi)$, and the integral $\int_{\Omega\in S^2} d\Omega = \int_0^{2\pi}\int_0^{\pi}\sin\theta\, d\theta\, d\phi$ covers the entire surface area of the unit sphere,[1] denoted by S^2. The spherical harmonics Y_n^m are defined by

$$Y_n^m(\theta,\phi)\equiv\sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\theta)e^{im\phi}, \tag{3}$$

where n is the order of the spherical harmonics and $i=\sqrt{-1}$. The spherical harmonics are the solution to the wave equation, or the Helmholtz equation in spherical coordinates.[1,18] The associated Legendre function P_n^m represents standing spherical waves in θ while the factor $e^{im\phi}$ represents traveling spherical waves in ϕ. The spherical harmonics are orthonormal, and so the following relations hold:[18]

$$\int_{\Omega\in S^2} Y_{n'}^{m'}(\theta,\phi)Y_n^{m*}(\theta,\phi)d\Omega=\delta_{nn'}\delta_{mm'}, \tag{4}$$

$$\sum_{n=0}^{\infty}\sum_{m=-n}^{n} Y_n^{m*}(\theta',\phi')Y_n^m(\theta,\phi) = \delta(\phi-\phi')\delta(\cos\theta-\cos\theta'), \tag{5}$$

where $\delta_{nn'}$ and $\delta_{mm'}$ are the Kronecker delta functions and $\delta(\phi-\phi')$ and $\delta(\cos\theta-\cos\theta')$ are the Dirac delta functions.[1] The next properties relate the Fourier coefficients of a function to its symmetry on the sphere with respect to the azimuth ϕ,

$$f_{nm}=f_n\delta_m\Leftrightarrow f(\theta,\phi)=f(\theta), \tag{6}$$

$$f_{nm}=f_{n(-m)}\Leftrightarrow f(\theta,\phi)=f(\theta,\pi-\phi). \tag{7}$$

Equation (6) which shows that when the coefficients are zero for all $m\neq 0$ then the function is constant along ϕ, can be derived from Eqs. (1) and (2). Equation (7) which shows that symmetry over $m=0$ results in symmetry over $\phi=90°$, 270°, can be derived from Eq. (2) and the relation $Y_n^m(\theta,\pi-\phi)=Y_n^{-m}(\theta,\phi)$.[19] When the Fourier coefficients are equal for all m, i.e., $f_{nm}=f_n$, then the values for $m=0$ correspond to the part which is constant with ϕ as in Eq. (6), while the function is symmetric along $\phi=90°$, 270° as Eq. (7) is satisfied. This result will be used later in this paper.

The final property relates multiplication of spherical Fourier transform coefficients to spherical convolution.[17] Consider two functions $f(\gamma)$ and $h(\gamma)$ square integrable on the unit sphere, with the vector of Cartesian coordinates $\gamma(\theta,\phi)=(\sin\theta\cos\phi,\sin\theta\sin\phi,\cos\theta)^T$ representing a point on the unit sphere. The spherical convolution of f with h denoted by $f*h$ is defined as the integral of the product of f and a rotated h, for all possible rotations,[17]

$$f(\gamma)*h(\gamma)=\int_{g\in \mathrm{SO}(3)} f(g\,\eta)h(g^{-1}\gamma)dg, \tag{8}$$

where the special orthogonal group of 3×3 matrices, SO(3), represents the three-dimensional rotation group, with $g\in \mathrm{SO}(3)$ spanning all rotations, and $\eta=(0,0,1)^T$ is the north pole. Euler matrix representation of rotation[17] with $g=R_z(\phi)R_y(\theta)R_z(\psi)$ can be used, where the 3×3 matrices R_z and R_y, defined in Cartesian coordinates, represent counterclockwise rotations about the z and y axes, respectively, and the rotation integral written as[17] $\int_{g\in \mathrm{SO}(3)}dg = \int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi}\sin\theta\, d\theta\, d\phi\, d\psi$.

Driscoll and Healy[17] show that spherical convolution converts to multiplication through the spherical Fourier transform, i.e.,

$$\mathcal{S}\{f(\gamma)*h(\gamma)\}=2\pi\sqrt{\frac{4\pi}{2n+1}}f_{nm}\cdot h_{n0}. \tag{9}$$

Note that h_{nm} is evaluated only at $m=0$, which means that $f*g\neq g*h$ and so the spherical convolution is noncommutative, or non-Abelian. The convolution $f*h$ is denoted the left convolution[17] of h by f. Equation (9) suggests that only the coefficients h_{n0} which correspond to the azimuthally symmetric part of h participate in the convolution, and that the function with Fourier coefficients h_{n0} is symmetric along ϕ in a manner described above. Also note that an alternative definition of spherical convolution exists,[20] although the one presented here was found to be more useful in this paper.

III. PLANE-WAVE DECOMPOSITION

In this section the sound pressure measured on a sphere for a plane-wave sound field is first presented, leading to derivation of plane-wave decomposition, i.e., finding the waves composing the sound field given the sound pressure on the sphere. Consider a unit amplitude incident plane wave arriving from (θ_l,ϕ_l). The pressure p_l on a sphere due to the incident plane wave at position (r,θ,ϕ) and wave number k, can be written using spherical harmonics as[1]

$$p_l(kr,\theta,\phi)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} b_n(kr)Y_n^{m*}(\theta_l,\phi_l)Y_n^m(\theta,\phi), \tag{10}$$

where b_n is generalized for open spheres (virtual spheres) and rigid spheres as follows:

$$b_n(kr)=\begin{cases} 4\pi i^n\left(j_n(kr)-\dfrac{j_n'(ka)}{h_n'(ka)}h_n(kr)\right) & \text{rigid sphere,}\\ 4\pi i^n j_n(kr) & \text{open sphere,}\end{cases} \tag{11}$$

where j_n and h_n are the spherical Bessel and Hankel functions, j_n' and h_n' are their derivatives, and $a \leqslant r$ is the radius of the rigid sphere. Note that Williams[1] uses a notation with wave propagation directions (rather then arrival directions), but as the wave number is in opposite sign compared to the notation used here, the result for b_n is the same. Taking the spherical Fourier transform [as in Eq. (1)] of Eq. (10), the coefficients $p_{l_{nm}}$ are given by

$$p_{l_{nm}}(kr) = b_n(kr) Y_n^{m*}(\theta_l, \phi_l). \tag{12}$$

We now assume that an infinite number of plane waves arrive at the sphere from all directions $\Omega_l = (\theta_l, \phi_l)$, with amplitudes $w(k, \theta_l, \phi_l)$ (strictly, w is the amplitude density). The spherical Fourier transform of the total pressure on the sphere due to all waves can be calculated by integrating Eq. (12) over these directions,

$$\begin{aligned} p_{nm}(kr) &= \int_{\Omega_l \in S^2} w(k, \theta_l, \phi_l) b_n(kr) Y_n^{m*}(\theta_l, \phi_l) d\Omega_l \\ &= w_{nm}(k) b_n(kr), \end{aligned} \tag{13}$$

where w_{nm} is the spherical Fourier transform of sound field directivity function w. Equation (13) can be rewritten as

$$w_{nm} = p_{nm} \cdot \frac{1}{b_n}, \tag{14}$$

where the dependence on k has been omitted for simplicity. Equation (14) shows that plane-wave decomposition can be performed in the spherical Fourier domain simply by dividing the pressure coefficients with b_n. Plane-wave decomposition can also be formulated as spherical convolution by taking the inverse spherical Fourier transform of both sides of Eq. (14) and using the result in Eq. (9),

$$w = p * \beta, \tag{15}$$

where

$$\beta = \mathcal{S}^{-1}\left\{ \frac{1}{2\pi} \sqrt{\frac{2n+1}{4\pi}} \frac{1}{b_n} \right\}. \tag{16}$$

Equation (15) shows that plane-wave decomposition can be performed in the spatial domain by spherically convolving p with β.

Taking the simple case of a single plane wave with a unit amplitude arriving from direction (θ_l, ϕ_l), we can substitute Eq. (12) in Eq. (14), and use Eqs. (2) and (5) to get

$$\begin{aligned} w(\theta, \phi) &= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^{m*}(\theta_l, \phi_l) Y_n^m(\theta, \phi) \\ &= \delta(\phi - \phi_l)\, \delta(\cos\theta - \cos\theta_l). \end{aligned} \tag{17}$$

This confirms that the directivity of a single plane wave is a delta function pointing to the arrival direction of the plane wave.

Figures 1(a) and 1(b) show $b_n(kr)$ for a rigid and an open sphere, respectively, as a function of kr and the order n. For both rigid and open spheres, terms with orders higher than about kr contribute significantly less than the lower order terms. Also, for an open sphere, some coefficients b_n will be zero for some values of kr, which means that Eqs. (15) and (14), which include a division by b_n, cannot be employed for these kr values. The rigid sphere therefore provides a better conditioned problem compared to the open sphere. Figure 2 further illustrates b_n for a rigid sphere and several values of ka, showing clearly how b_n starts to decay in magnitude beyond about $n = ka$.

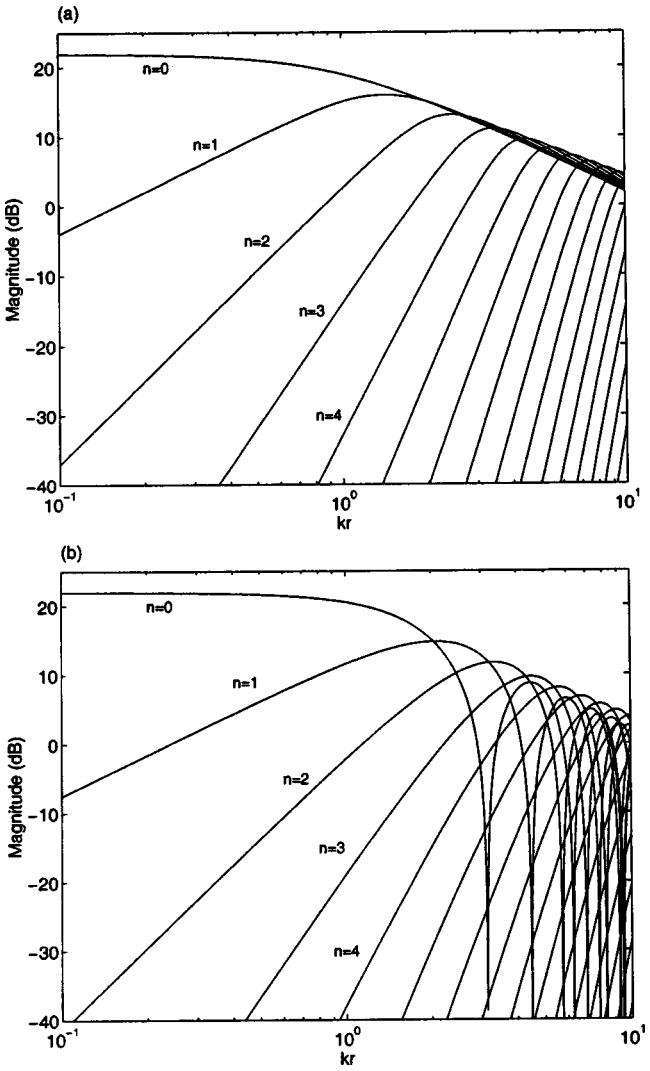


FIG. 1. Magnitude of $b_n(kr)$ for (a) rigid sphere with $kr = ka$, (b) open sphere, with orders of $n = 0$ to 4 denoted on the figures.

An example of the function β for a rigid sphere, evaluated for $kr = ka = 5$ and $n \leqslant 10$ using Eq. (16) over a finite grid on the unit sphere, is presented next. Figures 3(a) and 3(b) show $\beta(\theta, \phi)$ using a three-dimensional (3D) plot on a unit sphere, and on a two-dimensional (2D) plot as a function of θ and ϕ. Note the symmetry around $\phi = 90°$ and 270°, as expected from the symmetry property presented in Eq. (7) satisfied by $1/b_n$.

For band-limited p, i.e., $p_{nm} = 0$ for $n > N$, the corresponding w is also band limited as evident from Eq. (14). This means that to find w, a band-limited version of β is required, i.e., only $1/b_n$ for $n \leqslant N$ is required. A band-limited version of Eq. (15) can therefore be written as

$$w_N = p_N * \beta_N, \tag{18}$$

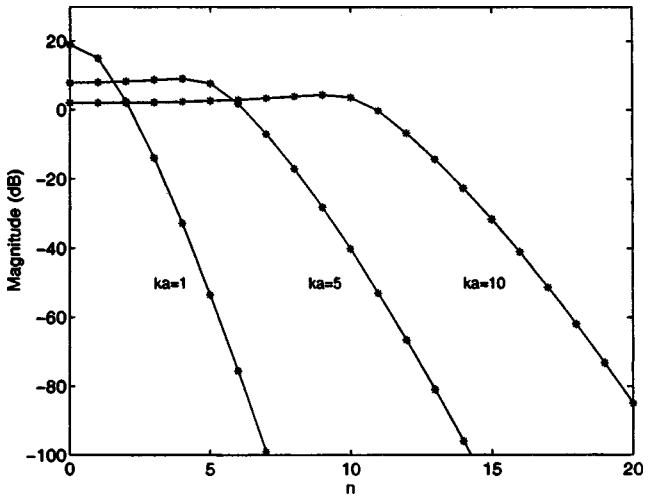


FIG. 2. Magnitude of $b_n(kr)$ for a rigid sphere as a function of n, for $ka=1$, 5, and 10 as denoted on the figure, and $kr=ka$.

where the functions w_N, p_N, and β_N are calculated using the inverse spherical Fourier transform as in Eq. (2) only with the summation over n taken from 0 to N and not infinity. This is a useful result since the calculation of β using Eq. (16) involves an inverse of b_n, and as illustrated in Fig. 2, $b_n \to 0$ as $n \to \infty$. To ensure stability of β, b_n must be truncated, and so if p is originally measured over a finite band, b_n is truncated to the same order with no further loss of accuracy in plane-wave decomposition. Note from Eq. (12) that b_n also appears in the magnitude of the pressure on the sphere due to a plane wave, and so given the frequency and the sphere dimensions, kr, the magnitude of the Fourier coefficients of orders larger then kr is expected to diminish, and so the assumption that p is band limited will be approximately correct. Nevertheless, as shown in the next section, the spatial resolution of plane-wave decomposition depends on the order N, and so truncating the bandwidth of p could limit plane-wave decomposition.

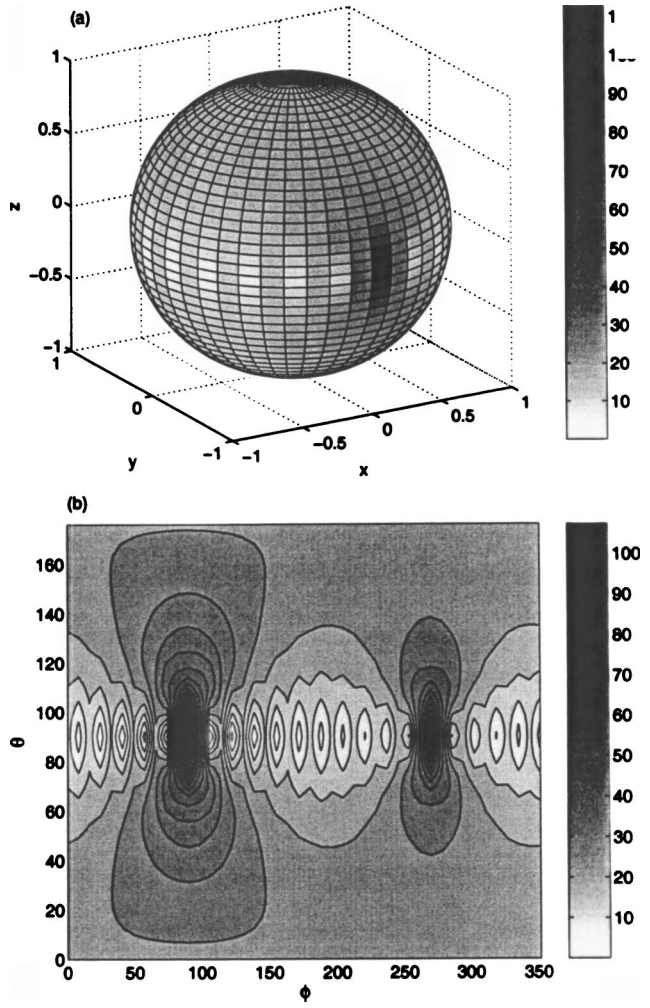


FIG. 3. Magnitude of $\beta(\theta,\phi)$ for a rigid sphere, $kr=ka=5$, $n \leqslant 10$, (a) three-dimensional plot on the unit sphere and (b) two-dimensional plot as a function of (θ,ϕ).

IV. SPATIAL RESOLUTION

In practice it might not be possible to measure all spherical Fourier coefficients of the pressure, since coefficients of orders higher than about kr have diminishing amplitudes. The directivity of the sound field in this case is calculated by taking the inverse spherical Fourier transform of Eq. (14), only with a finite order N,

$$w_N(\theta,\phi)=\sum_{n=0}^{N}\sum_{m=-n}^{n} p_{nm}\frac{1}{b_n}Y_n^m(\theta,\phi). \tag{19}$$

When the sound field is composed of a unit amplitude plane wave arriving from (θ_l,ϕ_l), w_N is evaluated by substituting Eq. (12) in Eq. (19),

$$\begin{aligned} w_N(\theta,\phi) &= \sum_{n=0}^{N}\sum_{m=-n}^{n} Y_n^{m*}(\theta_l,\phi_l)Y_n^m(\theta,\phi) \\ &= \sum_{n=0}^{N}\frac{2n+1}{4\pi}P_n(\cos\Theta) \\ &= \frac{N+1}{4\pi(\cos\Theta-1)}[P_{N+1}(\cos\Theta)-P_N(\cos\Theta)] \\ &= w_N(\Theta), \end{aligned} \tag{20}$$

where Θ is the angle between the vectors with directions (θ,ϕ) and (θ_l,ϕ_l), which can be written as[18]

$$\cos\Theta=\cos\theta\cos\theta_l+\cos(\phi-\phi_l)\sin\theta\sin\theta_l. \tag{21}$$

The spherical harmonics addition theorem (Arfken and Weber,[18] p. 796 and onwards) was used in the first two lines of Eq. (20), while the third line used a result from Gradshteyn and Ryzhik[21] [p. 1026, Sec. 8.915 (1)]. Figure 4 shows the directivity $w_N(\Theta)$ for various values of N, evaluated from Eq. (20), and normalized to have a unit gain at $\Theta=0$. The figure shows that w_N widens as N decreases, and so it is clear that with a finite order spatial resolution and plane-wave decomposition are limited. Notice that w_N is now only a function of a single angle Θ around an axis given by the vector (θ_l,ϕ_l).

We now formulate a relation between N and the resolution in plane-wave decomposition, by considering the first (smallest) zero Θ_0 of $w_N(\Theta)$, and taking twice that value, or the width of the main lobe, as the resolution. This implies

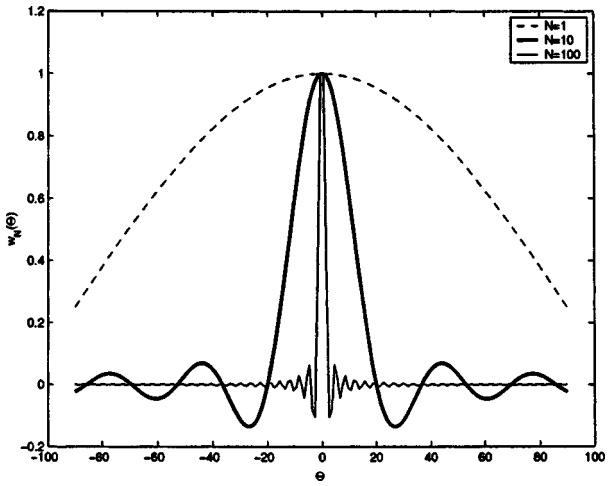


FIG. 4. Normalized directivity $w_N(\Theta)$ for a plane wave as a function of order N.

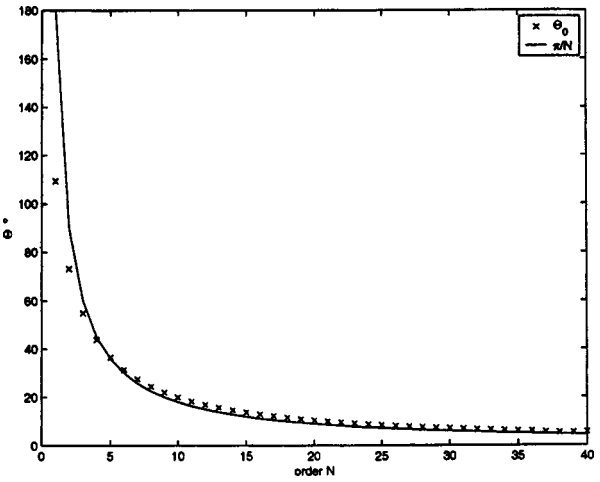


FIG. 5. The first zero $w_N(\Theta_0)=0$ as a function of N, and the approximation $\Theta_0 \approx \pi/N$.

that plane waves separated by $2\Theta_0$ should be decomposed reasonably well. Equation (20) shows that w_N is a sum of two Legendre polynomials in $\cos\Theta$. We can therefore compute the coefficients of these polynomials for every N, find their zeros, and locate the smallest zero Θ_0. As Legendre polynomials $P_n(z)$ are defined over $z\in[-1,1]$, we find the zero which is nearest to 1 (such that $\cos\Theta$ is nearest zero). It turns out that the polynomial defined by $P_{N+1}(z)-P_N(z)$ has the first zero at $z=1$, which is canceled by a pole at $z=1$, i.e., the term $\cos\Theta-1$ in Eq. (20). Therefore the next zero nearest 1 is chosen. Table I shows w_N expressed as a polynomial for $N\leqslant 5$. This provides an explicit form for the directivity function, although the result in Eq. (20) can be used more generally.

Figure 5 shows Θ_0 as a function of N, and as approximated by

$$\Theta_0 \approx \frac{\pi}{N}. \tag{22}$$

The error of this approximation was found to be less then 2° for $N\in[4,40]$. Assuming that for a given frequency we choose $N=kr$, i.e., neglecting the higher order harmonics which are less significant and more difficult to compute accurately, as discussed above, then a new relation for the width or resolution $2\Theta_0$ can be written using Eq. (22), in terms of the sphere radius $r=a$ and the wave number k, as

$$2\Theta_0 \approx \frac{2\pi}{ka}. \tag{23}$$

We now compare the spatial resolution derived here with the already established results of plane-wave decomposition given the pressure on a plane. Given $\mathbf{k}=(k_x,k_y,k_z)$ the wave vector, then it can be shown[1] that the pressure on the $x-y$ plane (at $z=0$) is given as the two-dimensional Fourier transform of the wave amplitudes, i.e., $p(x,y)$ and $w(k_x,k_y)$ are Fourier pairs. Plane-wave decomposition can now be readily calculated, and is nonambiguous if we assume that the waves arrive only from one side of the plane, i.e.,

$$w(k_x,k_y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)e^{-i(k_x x+k_y y)}\,dx\,dy. \tag{24}$$

Pressure and directivity are therefore related by the Fourier transform in the planar case in contrast to spherical convolution in the spherical case. When p is a plane wave, w in Eq. (24) becomes a delta function.[22] But as spatial resolution for the spherical case was limited by finite order, the planar case is limited by the dimensions of the measurement plane. As a simple example consider a plane on the z axis defined by $x\in[-L/2,L/2]$ and $y\in[-L/2,L/2]$ and a unit amplitude plane wave propagating perpendicular to the plane, defined by $k_x=k_y=0$. The directivity for a plane wave sound field in this case is found by solving the integral in Eq. (24) with finite limits for $p=1$ and is given by

$$w_L(k_x,k_y)=L^2\,\mathrm{sinc}(k_xL/2)\,\mathrm{sinc}(k_yL/2). \tag{25}$$

With a directivity of a sinc function, plane-wave decomposition cannot be achieved perfectly. The resolution or the width of the main lobe of the sinc function depends on frequency and the aperture L. The first zeros of the sinc function in this case form a square confined by $k_x=k_y=\pm 2\pi/L$, while its peak points to $\theta=0$ ($k_x=k_y=0$). We further simplify the analysis by approximating this square by an inner circle of radius $\sqrt{k_x^2+k_y^2}=2\pi/L$. Writing $k_x=k\sin\theta\cos\phi$

TABLE I. Directivity function w_N for several orders N.

Order N	Directivity $w_N(z)$, $z=\cos\Theta$
0	$\frac{1}{4\pi}$
1	$\frac{1}{4\pi}(3z+1)$
2	$\frac{1}{4\pi}\frac{3}{2}(5z^2+2z-1)$
3	$\frac{1}{4\pi}\frac{1}{2}(35z^3+15z^2-15z-3)$
4	$\frac{1}{4\pi}\frac{5}{8}(63z^4+28z^3-42z^2-12z+3)$
5	$\frac{1}{4\pi}\frac{6}{16}(231z^5+105z^4-210z^3-70z^2+35z+5)$

TABLE II. Resolution, or width of main lobe for the spherical configuration ($2\Theta_0$) and planar configuration ($2\theta_0$) as a function of harmonic order N, diameter of measurement sphere ($2a$) or plane aperture (L).

Sphere $2\Theta_0$ Plane $2\theta_0$	N	$2a(\lambda)$ $L(\lambda)$
360.0°	1	0.3
180.0°	2	0.6
72.0°	5	1.6
36.0°	10	3.2
12.0°	30	9.5
7.2°	50	15.9
3.6°	100	31.8
1.2°	300	95.5

and $k_y = k \sin\theta \sin\phi$, we can approximate the position of the zeros as $\sin\theta_0 \approx 2\pi/kL$. For apertures which are much larger then a wavelength, $kL \gg 2\pi$, we can further approximate $\sin\theta_0$ by θ_0, and the width of the main lobe, or resolution, is approximated by

$$2\theta_0 \approx \frac{4\pi}{kL}. \tag{26}$$

The width of the main lobe, or resolution, approximated by Eq. (26) for the plane configuration, can be compared to Eq. (23) for the spherical configuration. These are similar when L, the width of the plane, is comparable to $2a$, the diameter of the sphere. This shows that when imposing finite size constraints on the plane, and finite order constraint (which relate to the sphere size) on the sphere, both provide similar spatial resolution for plane-wave decomposition. Table II illustrates the spatial resolution in plane-wave decomposition by showing $2\Theta_0$ for the sphere and $2\theta_0$ for the plane as a function of sphere diameter and plane aperture. The table shows that high harmonic order, or spheres and planes of large dimensions are required to achieve high resolution in plane-wave decomposition.

V. COMPUTATION COMPLEXITY

Computation of plane-wave decomposition in practice is based on data measured by microphones that spatially sample the sound field. Various spherical sampling schemes are available, which are aliasing-free for band-limited functions.[16] In particular, equiangular sampling use equally spaced samples on θ and ϕ, with a total of $4(N+1)^2$ samples.[17] Gaussian sampling is similar but use nearly equally spaced samples on θ, with a total of only $2(N+1)^2$ samples.[23] Nearly uniform sampling schemes require the least number of spatial samples, or microphones, around $\approx 1.5(N+1)^2$, which are distributed approximately uniformly on the sphere.[24] In the discussion that follows we use integrals over the various angles as a general way to represent the required computations, but we note that since the measured pressure is only available at the sample points, these integrals will be approximated in practice by summations, with order $O(N)$ samples for each integral or angle.

Computation of w for plane-wave decomposition can be performed in the spherical Fourier transform domain by first estimating p_{nm}, then calculating w_{nm} using Eq. (14), and then transforming to w. Assuming a microphone array of order $n \leqslant N$ is used such that the number of samples is of order $O(N^2)$, the computation of w will be of order $O(N^4)$ which results from the spherical Fourier transform calculations. Improvement in computation efficiency can be achieved for equiangle and Gaussian sampling schemes by separating the computations over θ and ϕ,

$$\begin{aligned} f_{nm} &= \int_0^{2\pi}\int_0^{\pi} f(\theta,\phi) Y_n^{m*}(\theta,\phi) \sin\theta \, d\theta \, d\phi \\ &= q_n^m \int_0^{\pi} \left[\int_0^{2\pi} f(\theta,\phi) e^{-im\phi} d\phi \right] P_n^m(\cos\theta) \sin\theta \, d\theta, \end{aligned} \tag{27}$$

where q_n^m denotes the square-root factor in Eq. (3). The term in the square brackets is evaluated by order $O(N)$ for each θ and m, i.e., by a total order of $O(N^3)$. Then, the integral over θ is evaluated by order $O(N)$ for each m and n, i.e., by a total order of $O(N^3)$. The entire computation is therefore achieved with the reduced order of $O(N^3)$. Further reductions can be achieved if the computation in the square brackets is performed using the fast Fourier transform (FFT),[25] and the integral over θ is evaluated using fast Legendre transforms.[17,25] A similar approach of separation over n and m can be used to reduce the computation of the inverse transform. First, the summation over m is zero padded to $m \in [-N, N]$, and then the same approach as for the forward transform follows:[17]

$$\begin{aligned} f(\theta,\phi) &= \sum_{n=0}^{N} \sum_{m=-N}^{N} f_{nm} Y_n^m(\theta,\phi) \\ &= \sum_{m=-N}^{N} \left[\sum_{n=0}^{N} f_{nm} q_n^m P_n^m(\cos\theta) \right] e^{im\phi}. \end{aligned} \tag{28}$$

Plane-wave decomposition can also be calculated by direct evaluation of spherical convolution without the use of the spherical Fourier transform. Evaluating the triple integral (or when approximated by summations) in Eq. (8) at $O(N^2)$ sample directions of w will require order $O(N^5)$ computations. However, the convolution integral can be reformulated as follows:

TABLE III. Direction of arrival (θ_l, ϕ_l), magnitude and phase of each of the five plane waves.

	θ_l	ϕ_l	$\lvert w(\theta_l,\phi_l)\rvert$	$\angle w(\theta_l,\phi_l)$
(i)	20°	200°	1	0°
(ii)	45°	270°	1	20°
(iii)	60°	270°	1	80°
(iv)	120°	210°	1	180°
(v)	120°	250°	1	270°

$$w(\theta,\phi)=\int_0^{2\pi}\int_0^{2\pi}\int_0^{\pi} p(g(\theta',\phi',\psi')\eta)\beta(g^{-1}(\theta',\phi',\psi')\gamma(\theta,\phi))\sin\theta'\,d\theta'\,d\phi'\,d\psi'$$

$$=\int_0^{2\pi}\int_0^{\pi} p(\theta',\phi')\left[\int_0^{2\pi}\beta(g^{-1}(\theta',\phi',\psi')\gamma(\theta,\phi))d\psi'\right]\sin\theta'\,d\theta'\,d\phi', \tag{29}$$

where $g\eta \equiv g(\theta',\phi',\psi')(0,0,1)^T$ reduces to (θ',ϕ') using Euler matrix representation of rotation. The term in the square brackets is known and can therefore be precomputed for each θ' and ϕ'. The order of computation is now reduced to $O(N^4)$. If the analysis of the sound field is performed in a limited range of directions only, e.g., when several waves of interest arrive from approximately the same direction, then the computation of w for these selected directions is of order $O(N^2)$, which is a significant reduction compared to computation by spherical Fourier transform which remains the same under these conditions. This case resembles beam forming with the term in the square brackets acting as the spatial weighting function.[9,16]

VI. SIMULATION EXAMPLE

An example of plane-wave decomposition given the sound pressure on a sphere is presented in this section to illustrate the theoretical results of this paper. A single frequency sound field around a rigid sphere of radius $r=a$, with $ka=5$, is considered. The sound field is composed of five plane waves, with their directions of arrival and amplitudes detailed in Table III. The spherical Fourier coefficients of the pressure on the sphere due to each of the waves were calculated using Eq. (12), and the spherical Fourier coefficients of the total sound pressure on the sphere calculated as the sum of the five sets of coefficients. The Fourier coefficients were truncated at $N=10\,(n\leqslant 10)$. As shown in Fig. 1(a), the contribution of the truncated coefficients (b_n for n

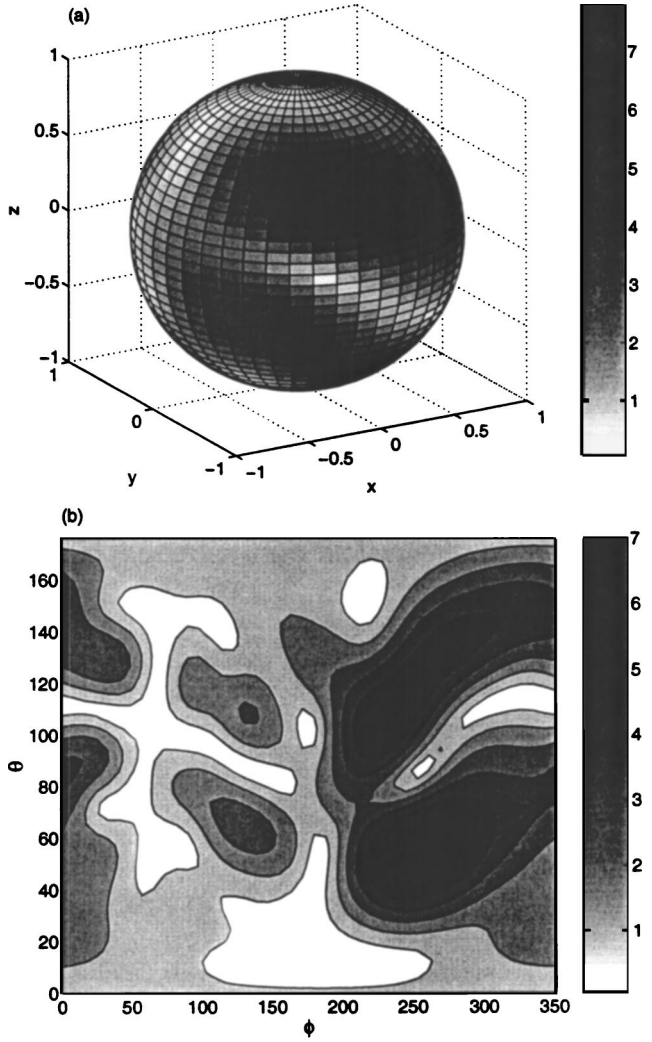


FIG. 6. Magnitude of the sound pressure level $p_N(\theta,\phi)$ on the sphere due to the plane waves described in Table III, (a) three-dimensional plot on the unit sphere and (b) two-dimensional plot as a function of (θ,ϕ).

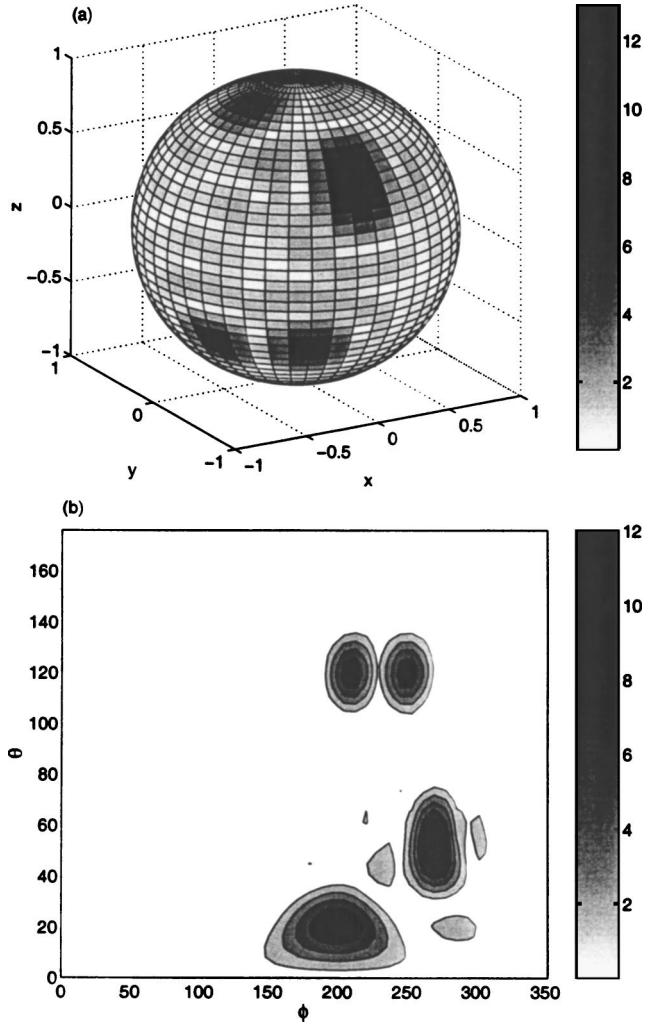


FIG. 7. Directivity w_N of the sound field due to the plane waves described in Table III, calculated for $n\leqslant 10$, illustrated using (a) three-dimensional plot on the unit sphere and (b) two-dimensional plot as a function of (θ,ϕ).

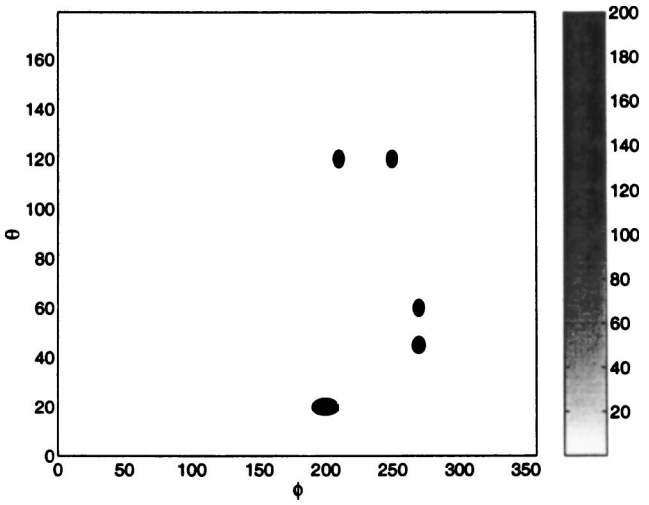


FIG. 8. Directivity w_N of the sound field due to the plane waves described in Table III, calculated for $n \leqslant 50$, illustrated using two-dimensional plot as a function of (θ,ϕ).

>10) at $ka=5$ is small. The directivity w_N, or plane-wave decomposition, was then calculated using Eq. (19).

Figure 6 illustrates the total pressure on the sphere due to the five plane waves. As can be seen it is not clear where the waves are arriving from given the total pressure. Figure 7 shows similar plots for w_N, the plane wave amplitudes after plane-wave decomposition was performed, where the directions of arrival are illustrated more clearly. Table II shows that for $N=10$ the spatial resolution, or width of the main lobe, is 36°. The waves denoted by (iv) and (v) in Table III are well separated in Fig. 7, as their spatial separation is 40°. However, the two waves denoted by (ii) and (iii), separated by only 15°, cannot be resolved in Fig. 7, as their separation is smaller than the resolution. Figure 8 shows a plot similar to that of Fig. 7(b), only here a higher order of $N=50$ was used for plane-wave decomposition, corresponding to a spatial resolution of 7.2° (see Table II). All five waves can now be resolved with this improved resolution.

This example illustrated how plane-wave decomposition can be achieved using spherical convolution and how finite harmonic order limits the spatial resolution of plane-wave decomposition.

VII. CONCLUSIONS

This paper presented the theory and an example of plane-wave decomposition given the pressure on a sphere. Formulation of plane-wave decomposition was developed, represented as spherical convolution, and an expression for the spatial resolution was derived, relating it explicitly to the limit on the harmonic order. As plane-wave decomposition can be computationally complex, especially when a high harmonic order is employed, ways to reduce it were proposed, for computations both in the spherical Fourier domain and the spatial domain. The results developed in this paper can be useful in the analysis and design of spherical microphone arrays for sound field analysis in various applications. Furthermore, the spherical convolution results provide a signal-system type framework for the analysis of sound on a sphere, which could be further explored in future work.

ACKNOWLEDGMENTS

Valuable discussions which initiated the research presented in this paper took place at the Research Lab of Electronics, MIT, with Julie E. Greenberg and other members of the Sensory Communication Group, while the author was on a Sabbatical leave at MIT, which was supported by EPSRC Grant No. GR/R90048/01 and the Royal Academy of Engineering grant reference IJB/AH/ITG 01-658. Part of the work presented in this paper was performed while the author was at the Institute of Sound and Vibration Research, University of Southampton, UK.

[1] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, New York, 1999).
[2] E. Hulsebos, D. De Vries, and E. Bourdillat, "Improved microphone array configurations for auralizations of sound fields by wave-field synthesis," J. Audio Eng. Soc. **50**, 779–790 (2002).
[3] H. F. Silverman, W. R. Patterson III, and J. L. Flanagan, "The huge microphone array," IEEE Concurrency **6**, 36–46 (1998).
[4] H. F. Silverman, W. R. Patterson III, and J. Sachar, "Factors affecting the performance of large-aperture microphone arrays," J. Acoust. Soc. Am. **111**, 2140–2157 (2002).
[5] A. J. Berkhout, D. De Vries, and P. Vogel, "Acoustic control by wave field synthesis," J. Acoust. Soc. Am. **93**, 2764–2778 (1993).
[6] A. J. Berkhout, D. De Vries, and J. J. Sonke, "Array technology for acoustic wave field analysis in enclosures," J. Acoust. Soc. Am. **102**, 2757–2770 (1997).
[7] H. Nomura and H. Miyata, "Microphone arrays for improving speech intelligibility in a reverberant or noisy space," J. Audio Eng. Soc. **41**, 771–781 (1993).
[8] J. Meyers, "Beamforming for a circular microphone array mounted on spherical shaped objects," J. Acoust. Soc. Am. **15**, 636–651 (2001).
[9] J. Meyers and G. W. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," Proceedings of ICASSP 2002, Vol. II, pp. 1781–1784.
[10] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," Proceedings of ICASSP 2002, Vol. II, pp. 1949–1952.
[11] J. Meyers and G. W. Elko, "A spherical microphone array for spatial sound recordings," J. Acoust. Soc. Am. **111**, 2346 (2002).
[12] B. N. Gover, J. G. Ryan, and M. R. Stinson, "Microphone array measurement system for analysis of directional and spatial variations of sound fields," J. Acoust. Soc. Am. **112**, 1980–1991 (2002).
[13] G. Weinreich and E. Arnold, "Method for acoustic radiation fields," J. Acoust. Soc. Am. **68**, 404–411 (1980).
[14] B. Rafaely, "Decomposition of reverberant sound fields into plane waves using microphone arrays," Presented at the 5th International Workshop on microphone array systems, theory and practice, Erlangen-Nuremberg, Germany, 15–16 May, 2003.
[15] B. Rafaely, "Plane-wave decomposition by spherical-convolution microphone array," J. Acoust. Soc. Am. **115**, 2578 (2004).
[16] B. Rafaely, Design and analysis of spherical microphone arrays (unpublished).
[17] J. R. Driscoll and D. M. Healy, Jr., "Computing Fourier transforms and convolutions on the 2-sphere," Adv. Appl. Math. **15**, 202–250 (1994).
[18] G. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, 5th ed. (Academic, San Diego, 2001).
[19] Wolfram Research on-line publication of mathematical functions, http://www.function.wolfram.com (2003).
[20] B. D. Wandelt and K. M. Gorski, "Fast convolution on the sphere," Phys. Rev. D **63**, 123002 (2001).
[21] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1980).
[22] A. V. Oppeneim, A. S. Willsky, and S. H. Nawab, *Signals and Systems* (Prentice Hall, New Jersey, 1997).

[23] V. I. Krylov, *Approximate Calculation of Integrals* (McMillan, New York, 1962). Translated from Russian by A. H. Stroud.
[24] R. H. Hardin and N. J. A. Sloane, ''McLaren's improved snub cube and other new spherical designs in three dimensions,'' Discrete Comput. Geom. **15**, 429–441 (1996). See also web site at http://www.research.att-com/njas/sphdesigns/
[25] M. J. Mohlenkamp, ''A fast transform for spherical harmonics,'' J. Fourier Analysis Appl. **5**, 159–184 (1999).

Radial reflection diffraction tomography

Sean K. Lehman[a)]
Lawrence Livermore National Laboratory, L-154, 7000 East Avenue, Livermore, California 94566

Stephen J. Norton
Geophex, Ltd., 605 Mercury Street, Raleigh, North Carolina 27603

(Received 24 November 2003; revised 28 June 2004; accepted 3 July 2004)

A wave-based tomographic imaging algorithm based upon a single rotating radially outward oriented transducer is developed. At successive angular locations at a fixed radius, the transducer launches a primary field and collects the backscattered field in a ''pitch/catch'' operation. The hardware configuration, operating mode, and data collection method are identical to that of most medical intravascular ultrasound (IVUS) systems. IVUS systems form images of the medium surrounding the probe based upon ultrasonic B scans, using a straight-ray model of sound propagation. The goal of this research is to develop a wave-based imaging algorithm using diffraction tomography techniques. Given the hardware configuration and the imaging method, this system is referred to as ''radial reflection diffraction tomography.'' Two hardware configurations are considered: a multimonostatic mode using a single transducer as described above, and a multistatic mode consisting of a single transmitter and an aperture formed by multiple receivers. In this latter case, the entire source/receiver aperture rotates about the fixed radius. Practically, such a probe is mounted at the end of a catheter or snaking tube that can be inserted into a part or medium with the goal of forming images of the plane perpendicular to the axis of rotation. An analytic expression for the multimonostatic inverse is derived, but ultimately the new Hilbert space inverse wave (HSIW) algorithm is used to construct images using both operating modes. Applications include improved IVUS imaging, bore hole tomography, and nondestructive evaluation (NDE) of parts with existing access holes. [DOI: 10.1121/1.1785651]

PACS numbers: 43.60.Pt, 43.60.Ac, 43.35.Zc [EJS] Pages: 2158–2172

I. INTRODUCTION

Consider a wave-based tomographic imaging tool consisting of a single transducer rotating about a fixed center. The transducer is oriented such that it launches fields radially outward. At each angular location, the transducer launches a primary field and collects the backscattered field in a ''pitch/catch'' operation. This configuration, in which a single transducer acts as both source and receiver at multiple spatial locations, is known as *multimonostatic*. The configuration is shown in Fig. 1(a). One may also consider an annular array of fixed transducers. In succession, each transducer launches a primary field and the backscattered field is measured on all the transducers. This configuration, with multiple spatially diverse transmitters and receivers, is a *multistatic* operating mode and is shown in Fig. 1(b). A second multistatic configuration, presented in Fig. 1(c), consists of a rotating subaperture formed by a single transmitter surrounded by multiple receivers. At each angular location, the transmitter launches the primary field and the backscattered field is measured on all receivers.

The goal of this tool is to use inverse wave techniques to reconstruct, that is, to form images, of the medium surrounding the probe, in the plane perpendicular to the axis of rotation. Given the arrangement of transducer(s) at a fixed radius collecting reflected scattered fields, we refer to this as a *radial reflection* configuration. We use the diffraction tomography[1,2] technique, based upon a linearized scattering model, to form images. Thus, given the physical transducer configuration and the mathematical method used to invert the scattering, we call the algorithm *radial reflection diffraction tomography* (RRDT).

When operating in a multimonostatic reflection mode, a spectrally wide band, that is a frequency diverse, incident source must be used because there is insufficient spatial diversity to form images of the surrounding medium. The planar reconstruction has two spatial variables, by the ''golden rule'' of tomography,[3] the measurement system must then have at least two free parameters. Angular location is one and incident source frequency is the other. Conceptually, the forward scattering process maps the two spatial variables of a physical object into the angular location and frequency parameters of the measured field.

Practically, such a probe is mounted at the end of a catheter or snaking tube that can be inserted into a part or medium with the purpose of forming images of the plane perpendicular to the axis of rotation. Applications include intravascular ultrasound (IVUS),[4] bore hole tomography, and nondestructive evaluation (NDE) of parts with existing access holes. We seek a wave-based imaging algorithm rather than a straight-ray B-scan[2] algorithm which is used, for example, in IVUS systems. A wave-based inversion more accurately describes the physics of wave field propagation.

In the next section, we develop the forward linearized multimonostatic scattering model. We initially develop a three-dimensional model. As we wish, however, to form im-

a)Electronic mail: lehman2@llnl.gov

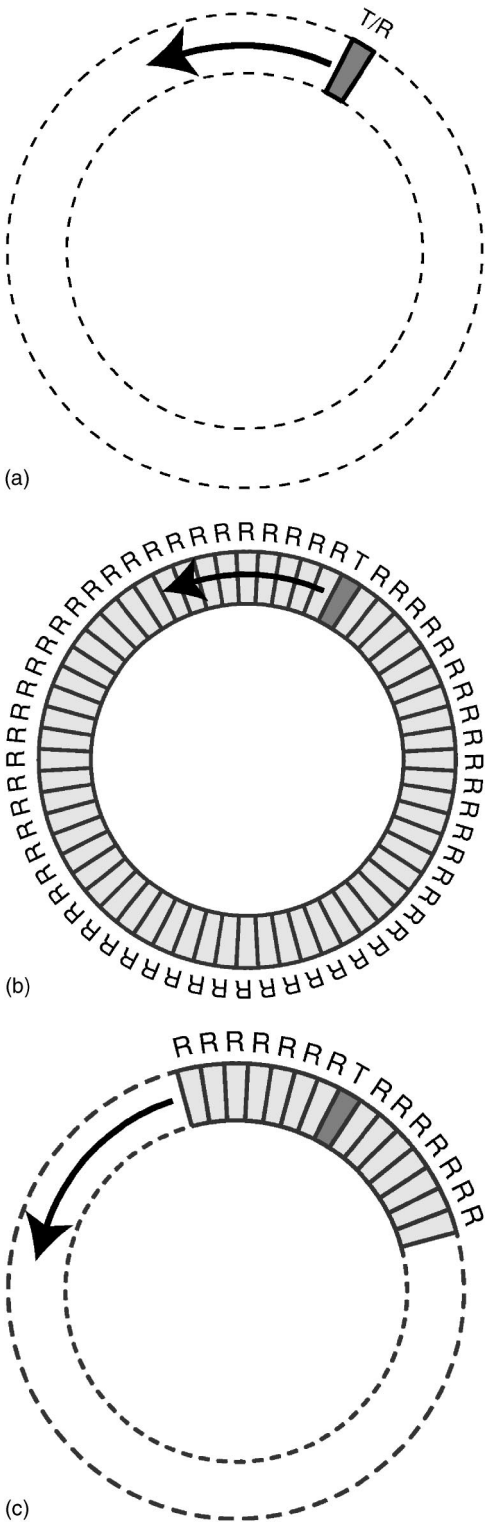


FIG. 1. Three radial reflection transducer configurations. (a) The multimonostatic mode: a single transducer rotates about a fixed center. At each angular location it launches the primary field and collects the reflected scattered field. (b) A multistatic case consisting of a fixed annular array of outward-looking transducers. In succession, each transducer launches the primary field and the reflected scattered field is measured at all the transducers. (c) A multistatic configuration consisting of a rotating aperture.

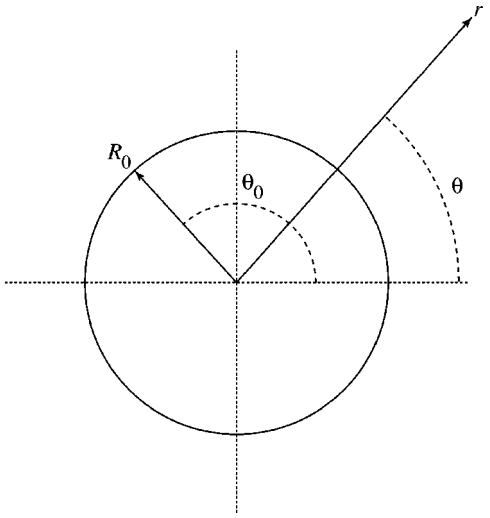


FIG. 2. Radial reflection geometry. R_0 is the probe diameter, $\mathbf{r}_0 \equiv R_0(\cos\theta_0, \sin\theta_0)$ is the planar location of the transducer, $\mathbf{r} \equiv r(\cos\theta, \sin\theta)$ is a point within the surrounding medium.

ages in a plane normal to the probe axis, we specialize the model to 2.5 dimensions assuming no variation in z, that is, along the axis of the catheter. In Sec. III, we develop the RRDT Fourier diffraction theorem,[1,2] (FDT) which governs how the medium's spatial spectrum is sampled by the forward measurement process. The RRDT FDT dictates the resolution of the reconstruction. We develop an analytic expression for multimonostatic inversion in Sec. IV. This inverse, however, proved impractical to implement numerically. In its place, we use the new Hilbert space inverse wave (HSIW) algorithm[5] outlined in Sec. V to achieve imaging algorithms not only for the multimonostatic configuration but also for the multistatic configurations.

Due to the heavy computation nature of the HSIW, a trade-off must be made between the number of transducers and frequencies used in the reconstruction, and the quality of the reconstructed image. Ideally, an aperture consisting of a 360-deg annular array would yield the best image. Practically, it is currently not possible to construct such a fully wired array in, for example, a 0.25-mm IVUS probe. Thus, we study rotating subapertures such as presented in Fig. 1(c), and compare reconstruction results to the single multimonostatic configuration.

As explained above, when operating in a reflection mode, the imaging mathematics dictates that spectrally wideband pulses must be used. The range resolution of the reconstructed image is proportional to the number of frequencies used in the reconstruction. Under the Hilbert space inverse wave algorithm, however, increasing the number of frequencies and transducers increases the complexity of the reconstruction, the size of the intermediary data files, reconstruction time, and computer memory requirement. Thus, the trade-off between computer resources and resolution must also be considered.

We initially study these issues in the proof-of-principle presented in Sec. VI, where we reconstruct scattered fields obtained from a simulation based upon the linearized forward scattering model of Sec. II. The simulated medium consists of a collection of point scatterers. We compare the multistatic and multimonostatic configurations. In Sec. VII, we provide a more rigorous test of the HSIW algorithm by inverting data obtained from Lawrence Livermore National

Laboratory's acoustic/elastic finite-difference time-domain propagation and scattering code, E3D.[6–11] We simulated a medium with two scattering objects whose physical properties varied from 5% below the background to 15% above it. Again, we compare multimonostatic and multistatic configurations. Conclusions are presented in Sec. VIII.

II. MULTIMONOSTATIC FORWARD SCATTERING MODEL

We develop here the linearized forward scattering model derivation for the multimonostatic configuration. We are interested in imaging an area surrounding the probe in the plane normal to the axis of rotation. The geometry is that of Fig. 2. The transducer location is specified by $\mathbf{r}_0 \equiv R_0(\cos\theta_0, \sin\theta_0)$, where R_0 is the probe radius, a constant. At each angular location, θ_0, the transducer launches the primary field radially outward into the medium and measures the reflected scattered field. The wave propagation and scattering is governed by the Helmholtz equation

$$[\nabla^2 + k^2(\mathbf{r})]\psi(\mathbf{r},\omega) = -p(\mathbf{r},\omega), \tag{1}$$

where $\mathbf{r} \equiv r(\cos\theta, \sin\theta)$ is the spatial coordinate of a point in the surrounding medium, ω is the temporal frequency, $k(\mathbf{r})$ is the wave number of the medium, $\psi(\mathbf{r},\omega)$ is the total field, and $p(\mathbf{r},\omega)$ is the incident pulse temporal spectrum.

The derivation is simplified by rearranging Eq. (1) to remove the spatial inhomogeneity in the scattering operator in the left-hand side. We add the background wave number, $k_0(\omega) \equiv \omega / v_0$, to both sides of Eq. (1) and move the inhomogeneous term to the right-hand side to obtain

$$[\nabla^2 + k_0^2]\psi(\mathbf{r},\omega) = -p(\mathbf{r},\omega) - [k^2(\mathbf{r}) - k_0^2(\omega)]\psi(\mathbf{r},\omega). \tag{2}$$

Define the *object function* as

$$o(\mathbf{r}) \equiv \frac{k^2(\mathbf{r})}{k_0^2} - 1, \tag{3}$$

and express Eq. (2) as

$$[\nabla^2 + k_0^2]\psi(\mathbf{r},\omega) = -p(\mathbf{r},\omega) - k_0^2(\omega)o(\mathbf{r})\psi(\mathbf{r},\omega). \tag{4}$$

The second term on the right-hand side is known as the *secondary source* which creates the scattered field. We may use Green's theorem[12] to cast the differential equation of Eq. (4) into an integral equation via

$$\psi(\mathbf{R},\omega) = \int d\mathbf{r}\, G(\mathbf{R},\mathbf{r},\omega)p(\mathbf{r},\omega) + k_0^2(\omega)\int d\mathbf{r}\, G(\mathbf{R},\mathbf{r},\omega)o(\mathbf{r})\psi(\mathbf{r}), \tag{5}$$

where the Green function is

$$G(\mathbf{R},\mathbf{r},\omega) = \frac{e^{ik_0(\omega)|\mathbf{R}-\mathbf{r}|}}{4\pi|\mathbf{R}-\mathbf{r}|}. \tag{6}$$

The first integral in Eq. (5) is the *primary field*, $\psi^{\text{inc}}(\mathbf{R},\omega)$. Subtracting it from the total field yields the *scattered field*

$$\psi^{\text{scatt}}(\mathbf{R},\omega) \equiv \psi(\mathbf{R},\omega) - \psi^{\text{inc}}(\mathbf{R},\omega) = k_0^2(\omega)\int d\mathbf{r}\, G(\mathbf{R},\mathbf{r},\omega)o(\mathbf{r})\psi(\mathbf{r},\omega).$$

An expression for the measured scattered field is obtained when $\psi^{\text{scatt}}(\mathbf{R},\omega)$ is evaluated on the measurement surface, $\mathbf{r}_0$

$$\psi^{\text{scatt}}(\mathbf{r}_0,\omega) = k_0^2(\omega)\int d\mathbf{r}\, G(\mathbf{r}_0,\mathbf{r},\omega)o(\mathbf{r})\psi(\mathbf{r},\omega). \tag{7}$$

Given the scattered field, the data on the measurement surface, $\mathbf{r}_0 = (R_0, \theta_0) \forall\, \theta_0$, we wish to invert Eq. (7) to reconstruct an estimate of the scattering medium, $o(\mathbf{r})$, in the plane perpendicular to the axis of transducer rotation. The equation is nonlinear, in its inverse, in that the total field, which is the sum of the incident and scattered fields, appears under the integral. We may simplify this if we assume the medium is weakly scattering and that the first Born approximation holds. We then neglect the scattered field term and express Eq. (7) as

$$\psi_B^{\text{scatt}}(\mathbf{r}_0,\omega) = k_0^2(\omega)\int d\mathbf{r}\, G(\mathbf{r}_0,\mathbf{r},\omega)o(\mathbf{r})\psi^{\text{inc}}(\mathbf{r},\omega), \tag{8}$$

where we use the B subscript to indicate the Born approximation has been used. We further assume the incident field is the result of a point source located at $\mathbf{r}_0$, so that $p(\mathbf{r},\omega) = P(\omega)\delta(\mathbf{r}_0 - \mathbf{r})$, where $P(\omega)$ is the incident pulse spectrum. We model neither antenna characteristics nor beam pattern. With this assumption, the incident field is

$$\psi^{\text{inc}}(\mathbf{r},\omega) = P(\omega)G(\mathbf{r}_0,\mathbf{r},\omega),$$

and Eq. (8) reads

$$\psi_B^{\text{scatt}}(\mathbf{r}_0,\omega) = P(\omega)k_0^2(\omega)\int d\mathbf{r}\, G^2(\mathbf{r}_0,\mathbf{r},\omega)o(\mathbf{r}), \tag{9}$$

where the squared Green function is a result of the transmitter and receiver being colocated. Using Eq. (6), we explicitly express the forward scattering model as

$$\psi_B^{\text{scatt}}(\mathbf{r}_0,\omega) = \frac{P(\omega)k_0^2(\omega)}{(4\pi)^2}\int d\mathbf{r}\, \frac{e^{i2k_0(\omega)|\mathbf{r}_0-\mathbf{r}|}}{|\mathbf{r}_0-\mathbf{r}|^2}o(\mathbf{r}). \tag{10}$$

The squared Green function is problematic in the development of an inverse scattering expression since the diffraction tomography technique requires an expansion of the integral kernel in a series along the measurement surface. We simplify Eq. (10) following a technique used by Norton and Linzer.[13] We first define the *weighted scattered field* as

$$\psi_W^{\text{scatt}}(\mathbf{r}_0,2\omega) \equiv \frac{4\pi}{P(\omega)k_0^2(\omega)}\psi_B^{\text{scatt}}(\mathbf{r}_0,\omega), = \frac{1}{4\pi}\int d\mathbf{r}\, \frac{e^{i2k_0(\omega)|\mathbf{r}_0-\mathbf{r}|}}{|\mathbf{r}_0-\mathbf{r}|^2}o(\mathbf{r}), \tag{11}$$

and then differentiate it with respect to $k_0(\omega)$

$$\frac{d}{dk_0(\omega)}\psi_W^{\text{scatt}}(\mathbf{r}_0,2\omega)=\frac{i2}{4\pi}\int d\mathbf{r}\,\frac{e^{i2k_0(\omega)|\mathbf{r}_0-\mathbf{r}|}}{|\mathbf{r}_0-\mathbf{r}|}o(\mathbf{r}),$$

$$=i2\int d\mathbf{r}\,G(\mathbf{r}_0,\mathbf{r},2\omega)o(\mathbf{r}). \quad (12)$$

For notational convenience, we define

$$\phi(\mathbf{r}_0,2\omega)\equiv\frac{d}{dk_0(\omega)}\psi_W^{\text{scatt}}(\mathbf{r}_0,2\omega)=v_0\frac{d}{d\omega}\psi_W^{\text{scatt}}(\mathbf{r}_0,2\omega), \quad (13)$$

and express Eq. (12) as

$$\phi(\mathbf{r}_0,2\omega)=i2\int d\mathbf{r}\,G(\mathbf{r}_0,\mathbf{r},2\omega)o(\mathbf{r}). \quad (14)$$

Equation (14) is our forward scattering model in three dimensions. Since our goal is to invert this equation to reconstruct the medium in a plane perpendicular to the axis of rotation, we simplify this further, in the next section, by assuming the object function, $o(\mathbf{r})$, is independent of the vertical (z) coordinate. This is known as a "2.5-dimensional" problem.

A. 2.5-dimensional problem

In cylindrical coordinates, $\mathbf{r}\equiv(r,\theta,z)$ and the measurement surface is $\mathbf{r}_0\equiv(R_0,\theta_0,z_0)$ for R_0 fixed and $0\leqslant\theta_0<2\pi$. Thus, the volume integral of Eq. (14) becomes

$$\phi(\mathbf{r}_0,2\omega)=i2\int_0^\infty r\,dr\int_0^{2\pi}d\theta\,o(r,\theta)\int_{-\infty}^{\infty}dz\,G(\mathbf{r}_0,\mathbf{r},2\omega). \quad (15)$$

The distance between the observation point, $\mathbf{r}_0$, and the scattering point is given by

$$|\mathbf{r}_0-\mathbf{r}|^2=R^2+(z_0-z)^2,$$

where the planar component of the distance is

$$R^2\equiv R_0^2+r^2-2R_0r\cos(\theta_0-\theta).$$

With this definition, we note that the z integral of the Green function reduces to

$$\int_{-\infty}^{\infty}dz\,G(\mathbf{r}_0,\mathbf{r},2\omega)=\frac{i}{4}H_0^{(1)}(2k_0(\omega)R),$$

so that Eq. (15) reads

$$\phi(\mathbf{r}_0,2\omega)=-\tfrac{1}{2}\int_0^\infty r\,dr\int_0^{2\pi}d\theta\,o(r,\theta)H_0^{(1)}(2k_0(\omega)R). \quad (16)$$

This is the 2.5-dimensional forward scattering model. It can be interpreted as a mapping from the object's two-dimensional (r,θ) space into the measured data's two-dimensional (θ_0,ω) measurement space. In the next section, we develop the equivalent of the Fourier diffraction theorem[1,2] (FDT) for this measurement system. The FDT determines the spatial resolution of the reconstruction by showing what part of the object's spatial spectrum is sampled by the measurement system.

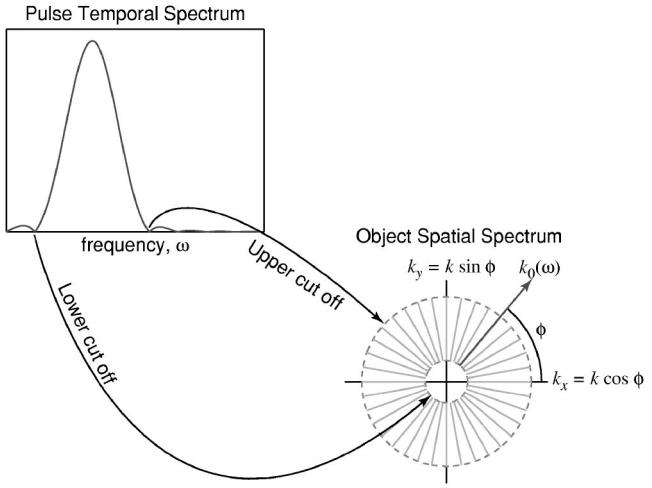


FIG. 3. Radial reflection Fourier diffraction theorem. Each radial line in the object's spatial spectrum (k_x,K_y), represents the locus of points sampled in at single wideband measurement location.

III. FOURIER DIFFRACTION THEOREM

The Fourier diffraction theorem (FDT)[1,14] relates the one-dimensional spectrum of the measured data along the measurement surface to the planar (two-dimensional) spatial spectrum of the object function. We begin the derivation by replacing the Hankel function in Eq. (16) with the expansion

$$H_0^{(1)}(k_0(\omega)R)=\sum_{n=-\infty}^{\infty}J_n(k_0(\omega)R_0)H_n^{(1)}(k_0(\omega)r)\times e^{in(\theta_0-\theta)},$$

where $R_0<r$. Substituting this into Eq. (16) yields

$$\phi(\mathbf{r}_0,2\omega)=-\frac{1}{2}\sum_{n=-\infty}^{\infty}\int_0^\infty r\,dr\int_0^{2\pi}d\theta\,o(r,\theta)\times J_n(2k_0(\omega)R_0)H_n^{(1)}(2k_0(\omega)r)e^{in(\theta_0-\theta)}. \quad (17)$$

We now Fourier expand the object and measured field functions along the measurement surface using the following transform pairs:

$$o_n(r)=\frac{1}{2\pi}\int_0^{2\pi}d\theta\,o(r,\theta)e^{-in\theta}, \quad (18)$$

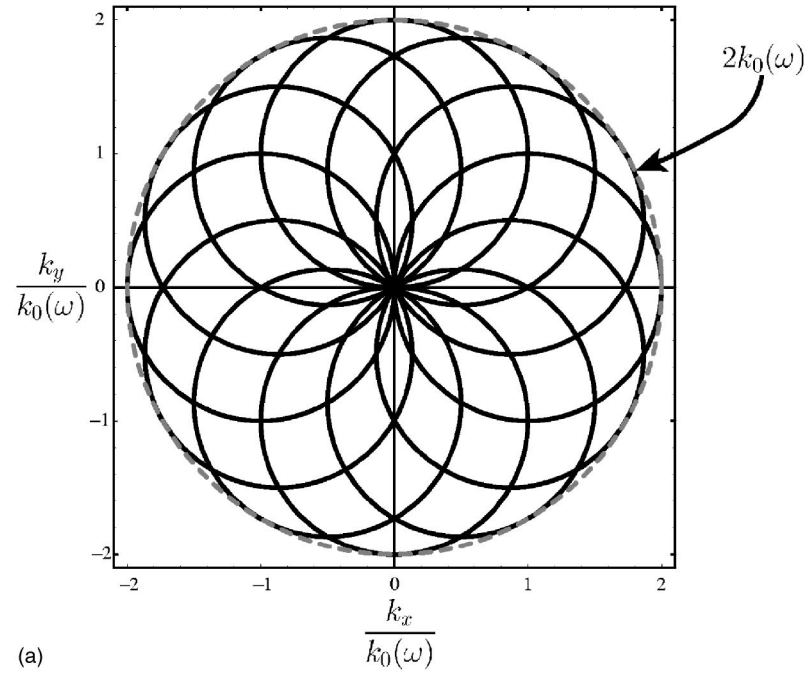
$$o(r,\theta)=\sum_{n=-\infty}^{\infty}o_n(r)e^{in\theta}, \quad (19)$$

and

$$\phi_n(2\omega)=\frac{1}{2\pi}\int_0^{2\pi}d\theta_0\,\phi(\mathbf{r}_0,2\omega)e^{-in\theta_0}, \quad (20)$$

$$\phi(\mathbf{r}_0,2\omega)=\sum_{n=-\infty}^{\infty}\phi_n(2\omega)e^{in\theta_0}. \quad (21)$$

Using Eq. (20) to transform Eq. (17) yields



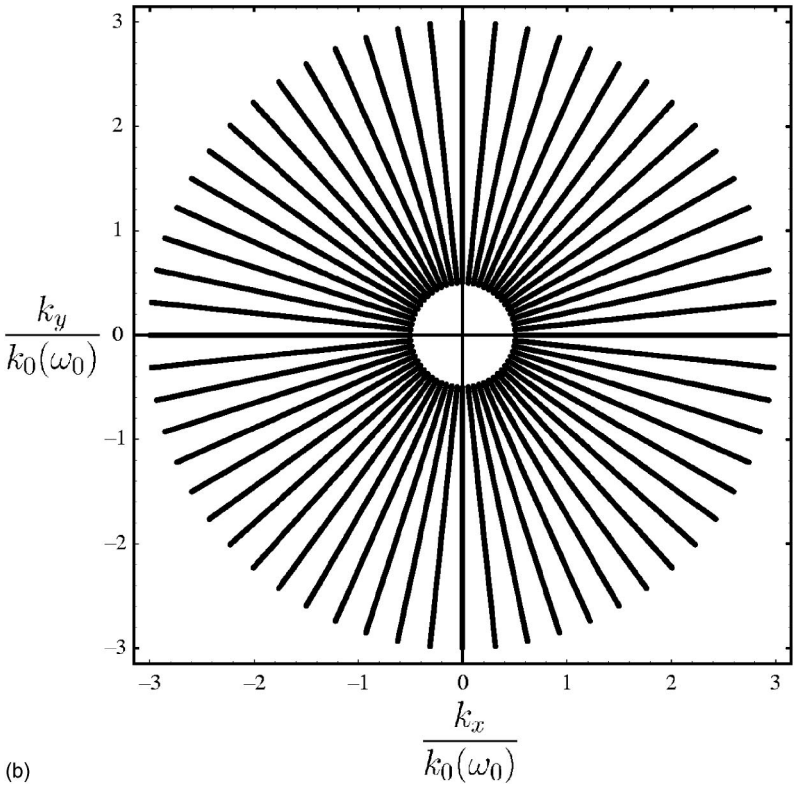


FIG. 4. Spatial frequency coverage for (a) an inwardly looking annular array where each line represents the locus of points sampled by a single multistatic measurement, and (b) an outwardly looking annular array (RRDT) where each line represents a single wideband monostatic measurement at a fixed angular location. The plots are normalized to $k_0(\omega)$, which is the narrow-band measurement frequency for the inwardly looking case, and the center of the band for the outwardly looking case.

$$\frac{1}{2\pi}\int_0^{2\pi} d\theta_0\, \phi(\mathbf{r}_0,2\omega)e^{-im\theta_0}$$

$$= -\frac{1}{4\pi}\sum_{n=-\infty}^{\infty} J_n(2k_0(\omega)R_0)\int_0^{\infty} r\,dr\,H_n^{(1)}(2k_0(\omega)r)$$

$$\times\int_0^{2\pi} d\theta o(r,\theta)e^{-in\theta}\int_0^{2\pi} d\theta_0 e^{i(n-m)\theta_0}. \qquad (22)$$

We use Eq. (18) and

$$\frac{1}{2\pi}\int_0^{2\pi} d\theta_0\, e^{i(m-n)\theta_0}=\delta_{mn}$$

to reduce Eq. (22) to

$$\phi_m(2\omega)=-\pi J_m(2k_0(\omega)R_0)\int_0^{\infty} r\,drH_m^{(1)}(2k_0(\omega)r)$$

$$\times o_m(r). \qquad (23)$$

Inverting Eq. (23) to obtain the transform of the object function, $o_m(\mathbf{r})$, is not possible because no orthogonality relation

exists for Hankel functions. If we make the assumption, however, that the object, $o(r,\theta)$, is real, we have from Eq. (18) that

$$o^*_{-m}(r)=o_m(r).$$

Using this and the property of Hankel functions

$$(-1)^m H^{(1)}_{-m}(2k_0(\omega)r)=H^{(1)}_m(2k_0(\omega)r),$$

we may solve for $o_m(r)$ by observing that

$$\phi_m(2\omega)+(-1)^m\phi^*_{-m}(2\omega)$$
$$=-2\pi J_m(2k_0(\omega)R_0)\int_0^\infty r\,dr J_m(2k_0(\omega)r)o_m(r), \quad (24)$$

results in a Bessel transform of the object

$$o_m(2k_0(\omega))\equiv\int_0^\infty r\,dr J_m(2k_0(\omega)r)o_m(r), \quad (25)$$

which is invertible. Using Eq. (25), we solve Eq. (24) for the object Bessel transform

$$o_m(2k_0(\omega))=-\frac{1}{2\pi J_m(2k_0(\omega)R_0)}[\phi_m(2\omega)$$
$$+(-1)^m\phi^*_{-m}(2\omega)]. \quad (26)$$

Explicitly expressing Eq. (26) in terms of the measured field, we obtain the "Fourier–Bessel diffraction theorem" for RRDT

$$o_m(2k_0(\omega))=-\frac{2v_0^3}{J_m(2k_0(\omega)R_0)}\frac{d}{d\omega}\left[\frac{1}{\omega^2}\left(\frac{\psi_m(\omega)}{P(\omega)}\right.\right.$$
$$\left.\left.+\frac{\psi^*_{-m}(\omega)}{P^*(\omega)}\right)\right]. \quad (27)$$

We believe it is more intuitive to relate the Fourier–Bessel transform of the object to its conventional two-dimensional Fourier transform. We do so in the next section.

A. The RRDT Fourier diffraction theorem

We wish to relate the Fourier–Bessel object transform of Eq. (27) in the (k,m) wave number/angular index space to a continuous variable wave number/angle Fourier space. The relation is developed in the Appendix. We substitute Eq. (27) into Eq. (A9) of the Appendix to obtain the Fourier diffraction theorem for RRDT

$$O(2k_0(\omega),\phi)=\sum_{m=-\infty}^{\infty}(-i)^m e^{im\phi}o_m(2k_0(\omega)),$$
$$=-2v_0^3\sum_{m=-\infty}^{\infty}\frac{(-i)^m e^{im\phi}}{J_m(2k_0(\omega)R_0)}\frac{d}{d\omega}$$
$$\times\left[\frac{1}{\omega^2}\left(\frac{\psi_m(\omega)}{P(\omega)}+\frac{\psi^*_{-m}(\omega)}{P^*(\omega)}\right)\right], \quad (28)$$

where the Fourier space polar pair $(2k_0(\omega),\phi)$ is related to the Fourier space Cartesian pair (k_x,k_y) via Eq. (A6).

This relationship shows that at each spatial spectrum angular location, ϕ, the locus covered in the object's spatial spectrum is a radial line, $k_0(\omega)$, that extends over the temporal spectral bandwidth of the pulse. When all angular locations are combined, the area is a torus whose inner and outer radii are proportional to the lower and upper pulse spectrum cutoff frequencies, respectively. If, for example, the pulse spectrum has a lower cutoff of ω_L and an upper cutoff of ω_U, the inner and outer radii of the torus are given by $2k_0(\omega_L)$ and $2k_0(\omega_U)$, respectively. This is shown in Fig. 3. The implication of this relationship is that radial resolution (alternatively, range resolution) is related to the spectral bandwidth of the incident field: the wider the bandwidth, the greater the resolution. Additionally, since the overall spatial spectral coverage is a torus omitting low frequencies about the origin, we expect the reconstruction to be hollow because it lacks the dc spatial field components.

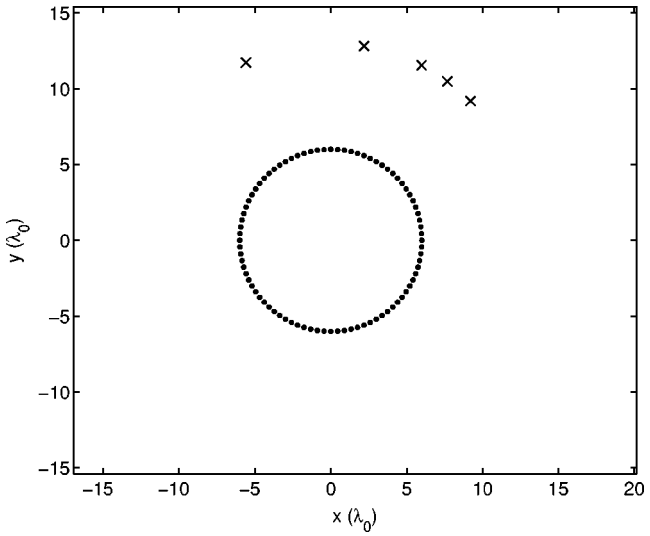


FIG. 5. Proof-of-principle geometry. Two simulations were performed: a multimonostatic run with 84 transducers which used 100 frequencies for the reconstruction, and a multistatic run with 20 transducers which used 21 frequencies.

It is informative to compare the spatial frequency space coverage, that is, the Fourier diffraction theorem, of the *outwardly* looking annular array to that of the *inwardly* looking annular array. André *et al.* derived this relation in[15] Eq. (5). For a single-frequency fully multimonostatic measurement system, the coverage is a disk centered at the origin. These two cases are shown in Fig. 4. The plots are normalized to $k_0(\omega)$, which is the narrow-band frequency for the inwardly looking array, and the center of the frequency band for the outwardly looking array.

IV. ANALYTIC MULTIMONOSTATIC INVERSE

We use the orthogonality of Bessel functions to invert Eq. (24) to obtain the transform of the object function

$$o_m(r)=-\frac{1}{\pi}\int_0^\infty k_0(\omega)dk_0(\omega)$$
$$\times\left[\frac{\phi_m(2\omega)+(-1)^m\phi^*_{-m}(2\omega)}{J_m(2k_0(\omega)R_0)}\right]J_m(2k_0(\omega)r). \quad (29)$$

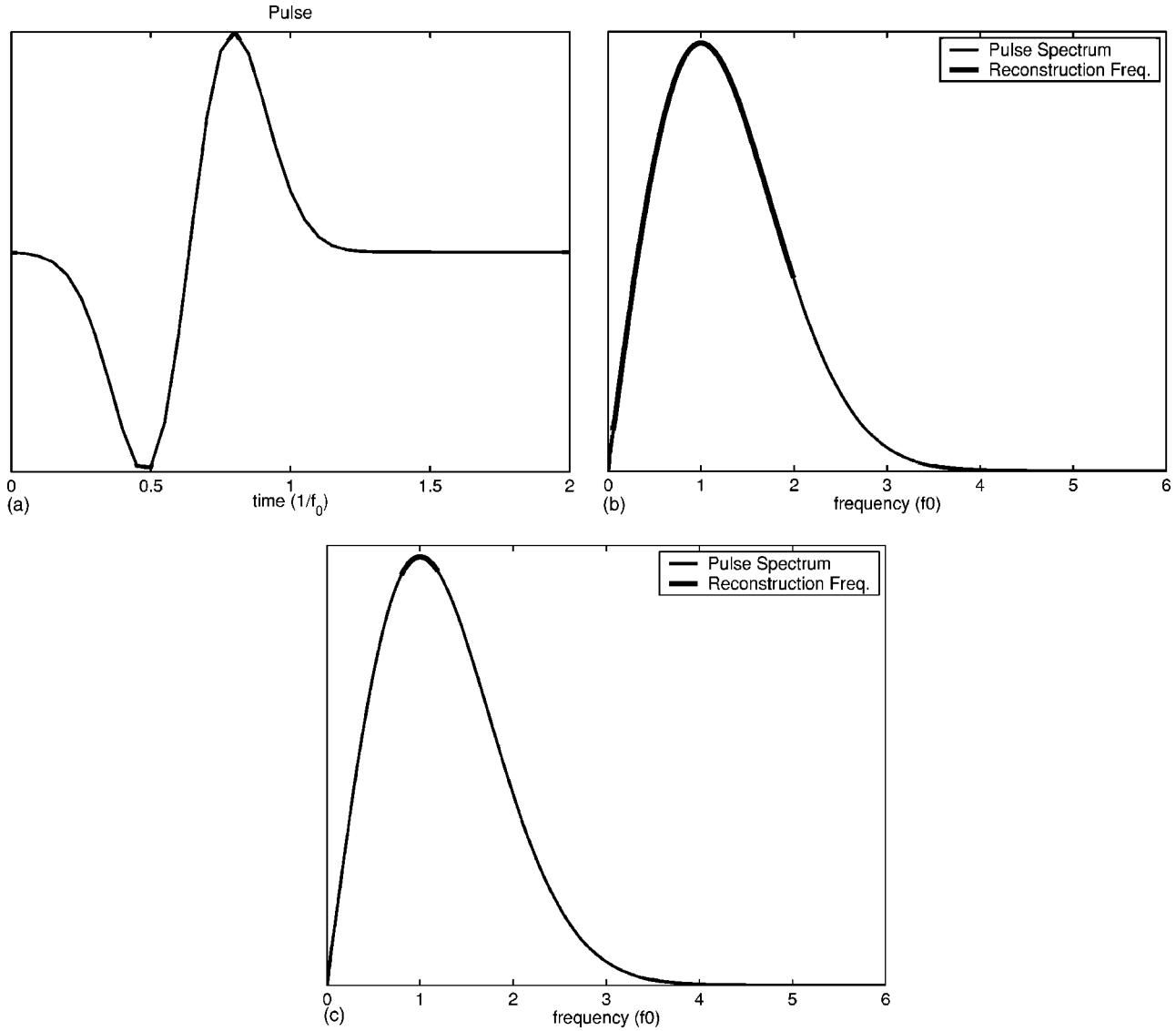


FIG. 6. Proof-of-principle pulse. (a) Time domain derivative of a Gaussian (DOG) pulse. (b) DOG pulse spectrum with the 100 frequencies used in the multimonostatic reconstruction highlighted. (c) DOG pulse spectrum with the 21 frequencies used in the multistatic reconstruction.

Combining Eq. (19) and Eq. (13) with Eq. (29) yields the reconstruction

$$\hat{o}(r,\theta) = -8v_0 \sum_{m=-\infty}^{\infty} e^{im\theta} \int_0^{\infty} \omega \, d\omega \frac{d}{d\omega} \left[\frac{1}{\omega^2} \left(\frac{\psi_m(\omega)}{P(\omega)} + \frac{\psi^*_{-m}(\omega)}{P^*(\omega)} \right) \right] \frac{J_m(2k_0(\omega)r)}{J_m(2k_0(\omega)R_0)}. \tag{30}$$

This analytic inversion has proved impractical to implement numerically. The reciprocal of the Bessel function results in poles along the real frequency axis obliging a contour integral. We were unsuccessful, however at identifying a contour in which every function within the integrand remained finite. This compelled us to seek a purely numerical solution to the inversion problem. We used the Hilbert space wave inversion algorithm developed by Devaney,[5] and summarize it in the following section.

V. HILBERT SPACE WAVE INVERSION

The Hilbert space inverse wave (HSIW) algorithm permits us to develop an inverse for any geometry with any combination of sources, receivers, and frequencies. Because we have this flexibility, we develop the HSIW theory for the most general multistatic, wideband case where we have spatial diversity in both the sources and receivers, as well as frequency diversity. In an actual radial reflection device such as an intravascular ultrasound probe, the data are collected at discrete angular locations. We denote by

$$\mathbf{R}_n^t \equiv R_0(\cos\theta_n, \sin\theta_n) \tag{31}$$

the transmitter locations where $\theta_n = n\Delta\theta_{\text{src}}$ for $n = 0,1,...,N_{\text{src}}-1$, where $N_{\text{src}} = 2\pi/\Delta\theta$. Similarly, the receiver locations are given by

$$\mathbf{R}_m^r \equiv R_0(\cos\theta_m, \sin\theta_m), \tag{32}$$

where $\theta_m = m\Delta\theta_{\text{rcv}}$ for $m = 0,1,...,N_{\text{rcv}}-1$ where $N_{\text{rcv}} = 2\pi/\Delta\theta_{\text{rcv}}$.

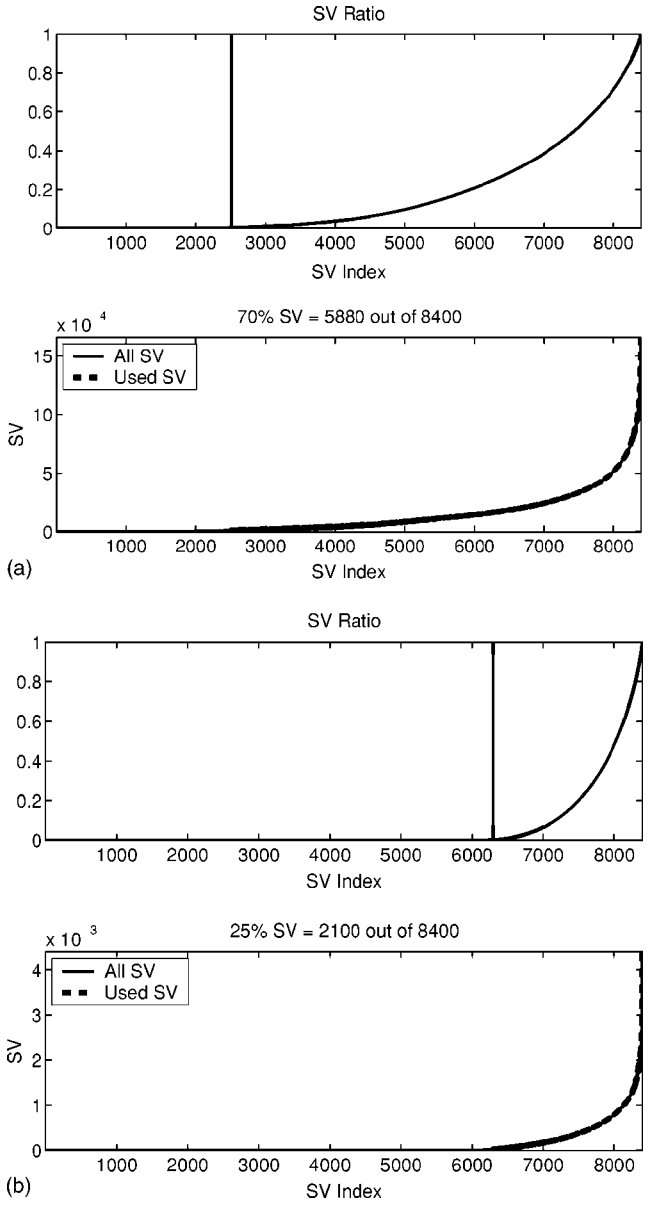


FIG. 7. Proof-of-principle singular values. The top plots show the eigenvalue ratios, $R(N)$, of Eq. (54). The vertical line shows where the selection was made "zero" and nonzero singular values. The bottom plots show the actual singular values. The highlighted shows those singular values used in the reconstructions. The (a) column is the multimonostatic case where 70% or 5880 out of 8400 singular values/vectors were used; the (b) column is the multistatic case where 25% or 2100 singular values/vectors were used.

For each source, the receiver(s) record the backscattered field as a time series that is digitized for processing. Fourier transforming the time-series data results in the spectrum of the measured wave form at discrete frequencies. The forward scattering equation under the Born approximation with both spatial and frequency diversity is

$$\psi_B^{\text{scatt}}(\mathbf{R}_m^r,\mathbf{R}_n^t,\omega_l)=P(\omega_l)k_0^2(\omega_l)\int d\mathbf{r}'G(\mathbf{R}_m^r,\mathbf{r}',\omega_l)\times o(\mathbf{r}')G(\mathbf{r}',\mathbf{R}_n^t,\omega_l), \tag{33}$$

where ω_l, $l=0,1,...,N_f-1$ are the discrete frequencies and N_f is the number of frequencies in the pulse bandwidth. For the multimonostatic case of Eq. (9), this reduces to

$$\psi_B^{\text{scatt}}(\mathbf{R}_n^t,\omega_l)=P(\omega_l)k_0^2(\omega_l)\int d\mathbf{r}'G^2(\mathbf{R}_n^t,\mathbf{r}',\omega_l)o(\mathbf{r}').$$

The HSIW interprets Eq. (33) as a mapping from a *continuous object space* to a *discrete measurement space*. The object space is the physical (x,y) space of the object function. The measurement space consists of the discrete angles and temporal frequencies at which the scattered data are collected. The scattering operator projects the object onto the measurement space. We define the forward propagation or projection kernel as

$$\Pi^*(\mathbf{r})\equiv P(\omega_l)k_0^2(\omega_l)G(\mathbf{R}_m^r,\mathbf{r},\omega_l)G(\mathbf{r},\mathbf{R}_n^t,\omega_l), \tag{34}$$

where $\Pi(\mathbf{r})$ is a $J\equiv(N_{\text{src}}\times N_{\text{rcv}}\times N_f)$ element column vector. Mathematically, the projection is represented as an inner product between the object function and the kernel via

$$D=\int d\mathbf{r}\,\Pi^*(\mathbf{r})o(\mathbf{r})\equiv\langle\Pi,o\rangle, \tag{35}$$

where D is a J element column vector where each element represents a particular source, receiver, and frequency combination. Symbolically, we define the forward scattering operator, K, as

$$K[\,\cdot\,]\equiv\int d\mathbf{r}\,\Pi^*(\mathbf{r})[\,\cdot\,]. \tag{36}$$

The HSIW is a method used to derive an inverse of this operator. Consider the singular value decomposition (SVD) of K

$$K=USV^\dagger, \tag{37}$$

where the columns of U form an orthonormal set of column vectors, u_j, which spans the measured data space, and the components of V form an orthonormal set of vectors, $v_j(\mathbf{r})$, which spans the object space. S is a diagonal matrix of singular values, σ_j. We wish to emphasize that the u_j are *column vectors*, whereas the $v_j(\mathbf{r})$ are *complex functions of* $\mathbf{r}$. The set of normal equations for this singular system is

$$Kv_j(\mathbf{r})=\sigma_j u_j, \tag{38}$$

$$K^\dagger u_j=\sigma_j v_j(\mathbf{r}), \tag{39}$$

$$KK^\dagger u_j=\sigma_j Kv_j(\mathbf{r})=\sigma_i^2 u_j, \tag{40}$$

$$K^\dagger Kv_j(\mathbf{r})=\sigma_j K^\dagger u_j=\sigma_i^2 v_j(\mathbf{r}). \tag{41}$$

The inversion goal is to estimate the object function of Eq. (35) given the measured data in D. We do so by expanding the object function in terms of the $v_j(\mathbf{r})$

$$\hat{o}(\mathbf{r})=\sum_{j=0}^{J-1}\alpha_j v_j(\mathbf{r}), \tag{42}$$

where the α_j are constant coefficients to be determined. Substituting the object expansion into Eq. (35), we obtain

$$D=\int d\mathbf{r}\,\Pi^*(\mathbf{r})\sum_{j=0}^{J-1}\alpha_j v_j(\mathbf{r}),$$

$$=\sum_{j=0}^{J-1}\alpha_j\int d\mathbf{r}\,\Pi^*(\mathbf{r})v_j(\mathbf{r}). \tag{43}$$

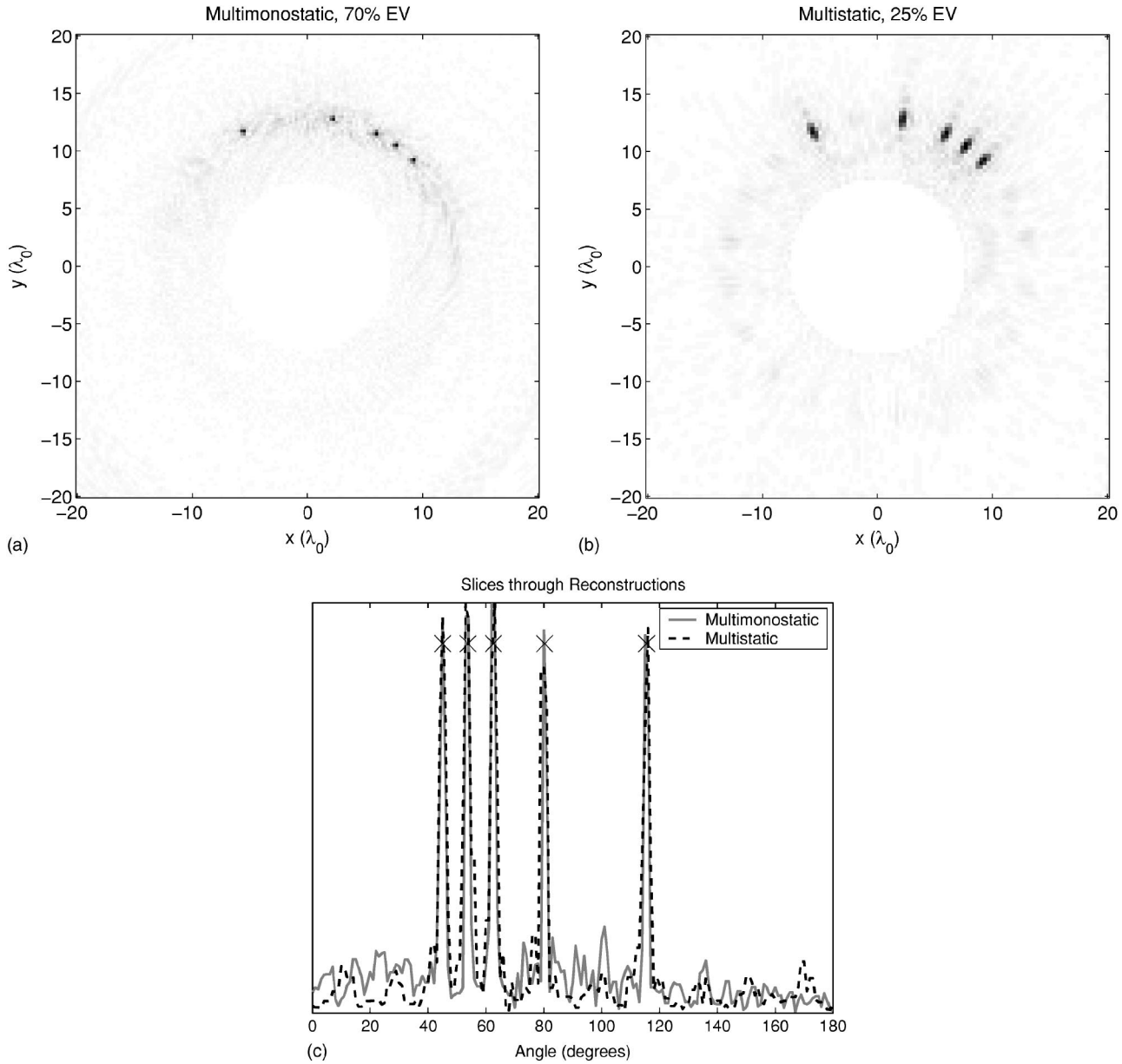


FIG. 8. Proof-of-principle reconstructions. (a) Multimonostatic. (b) Multistatic. (c) Angular slices through the reconstructions at the radius of the scatterers. The × indicate the actual location of the scatterers. The multimonostatic case shows better resolution and somewhat higher contrast than the multistatic case. This is a consequence of the greater number of frequencies used in the reconstruction which increased the coverage in the object's spatial spectrum.

Applying the definition of the K operator in Eq. (36) to Eq. (38) yields an expression for the integral of Eq. (43)

$$Kv_j = \int d\mathbf{r}\,\Pi^*(\mathbf{r})v_j(\mathbf{r}) = \sigma_j u_j\,, \tag{44}$$

which reduces Eq. (43) to

$$D = \sum_{j=0}^{J-1} \alpha_j \sigma_j u_j\,. \tag{45}$$

Using the orthogonality of the u_j vectors, we may solve for the unknown α_j as follows:

$$u_i^\dagger D = \sum_{j=0}^{J-1} \alpha_j \sigma_j u_i^\dagger u_j = \sum_{j=0}^{J-1} \alpha_j \sigma_j \delta_{ij} = \alpha_i \sigma_i\,, \tag{46}$$

resulting in

$$\alpha_i = \frac{u_i^\dagger D}{\sigma_i}. \tag{47}$$

We now require expressions for the adjoint of the forward scattering operator, $K^\dagger$, and the singular values and singular vectors, σ_j, u_j, and $v_j(\mathbf{r})$. Consider the following inner product equation which defines the adjoint:

$$\langle u, Kv\rangle = \langle K^\dagger u, v\rangle. \tag{48}$$

Using the definition of the forward scattering operator from Eq. (36), we have

$$u^\dagger \int d\mathbf{r}\,\Pi^*(\mathbf{r})v(\mathbf{r}) = \int d\mathbf{r}(u^\dagger \Pi^*(\mathbf{r}))v(\mathbf{r}). \tag{49}$$

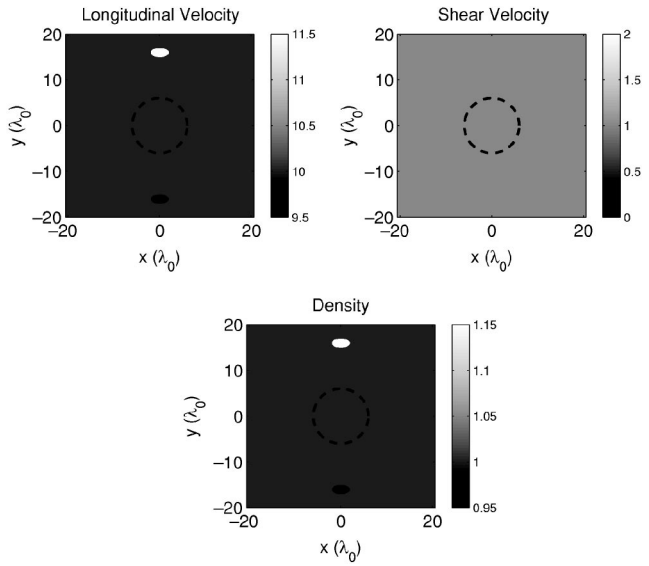


FIG. 9. Finite-difference time-domain simulation domains. The domains consist of a homogeneous background with two elliptical inclusions. One inclusion's longitudinal velocity and density are 15% above the background; the other's are 5% below. There is no variation in shear velocity. The dashed line indicates the transducer outline.

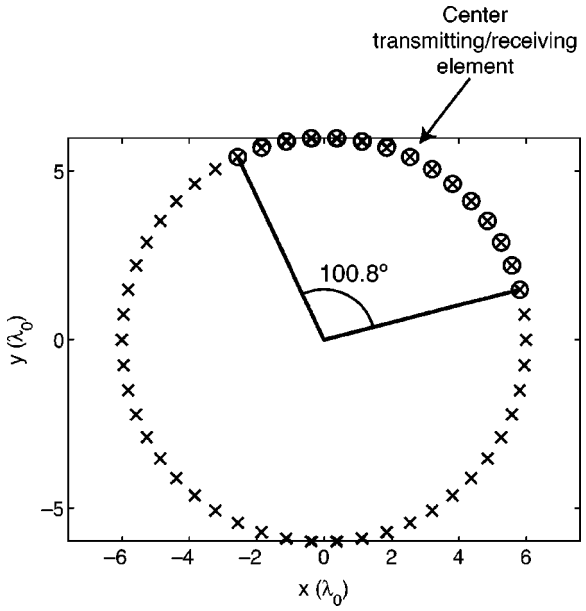


FIG. 10. The transducer geometry for the FDTD simulations. The × mark the 50 source positions. The circles indicate the active receivers within the rotating subaperture. For the 15-element aperture, the aperture angle is 100.8°.

Comparing the right-hand sides of Eq. (48) and Eq. (49), we obtain the following definition of the adjoint of the forward scattering operator:

$$K^{\dagger}[\,\cdot\,] \equiv [\,\cdot\,]\cdot \Pi^{T}(\mathbf{r}). \tag{50}$$

The σ_j and u_j are determined by solving the eigenvalue equation of Eq. (40) formed by the outer product of the forward scattering operator with its adjoint. Explicitly, the outer product is represented by

$$\left(\int d\mathbf{r}\, \Pi^{*}(\mathbf{r}) \Pi^{T}(\mathbf{r}) \right) u_j = \sigma_j^2 u_j, \tag{51}$$

which is a $J \times J$ eigenvalue equation of the form $Ax = \lambda x$. The $\Pi(\mathbf{r})$ vectors are known analytically and can be evaluated numerically. It follows that the elements of the outer product matrix can be computed numerically and the resulting system solved numerically for the σ_j^2 and u_j. Given these, we solve for $v_j(\mathbf{r})$ using Eq. (39)

$$v_j(\mathbf{r}) = \frac{1}{\sigma_j} \Pi^{T}(\mathbf{r}) u_j. \tag{52}$$

Substituting Eqs. (47) and (52) into Eq. (42) yields the final expression for the reconstruction

$$\hat{o}(\mathbf{r}) = \sum_{j=0}^{J-1} \frac{1}{\sigma_j^2} \Pi^{T}(\mathbf{r}) u_j u_j^{\dagger} D. \tag{53}$$

As described above, the $\Pi(\mathbf{r})$ vectors of Eq. (34) and outer products and eigenvalues of Eq. (51) are computed numerically. The system is inherently ill-conditioned due to the limited aperture of the measurement system which only measures part of the scattered field, and due to the loss of the evanescent field information. Thus, some of the eigenvalues, σ_j^2, are close to zero. Those eigenvalues and their corresponding eigenvectors determine the rank of the outer product matrix, and they must not be used in the reconstruction of Eq. (53). A decision must be made on the number of singular values/vectors to use. We have chosen to use the *best rank N* approximation. We compute the ratio

$$R(N) = \frac{\sum_{j=0}^{N-1} \sigma_j^2}{\sum_{j=0}^{J-1} \sigma_j^2}, \tag{54}$$

where we assume the singular values are arranged from smallest to largest: $\sigma_0^2 \leqslant \sigma_1^2 \leqslant \cdots \leqslant \sigma_{J-1}^2$. Plotting $R(N)$, we graphically identify the point at which the function starts to rise rapidly. The index of the singular value at which this

TABLE I. FDTD simulation configuration and reconstruction parameters.

Simulation configuration	Number of source locations	Number of receivers	Aperture angle	Number of reconstruction frequencies	Number of HSIW kernel elements, J
Multimonostatic	50	1	n/a	120	6000
Multistatic	50	15	100.8°	8	6000
Multistatic	50	15	100.8°	12	9000
Multistatic	50	9	57.6°	21	9450
Multistatic	50	5	28.8°	37	9520

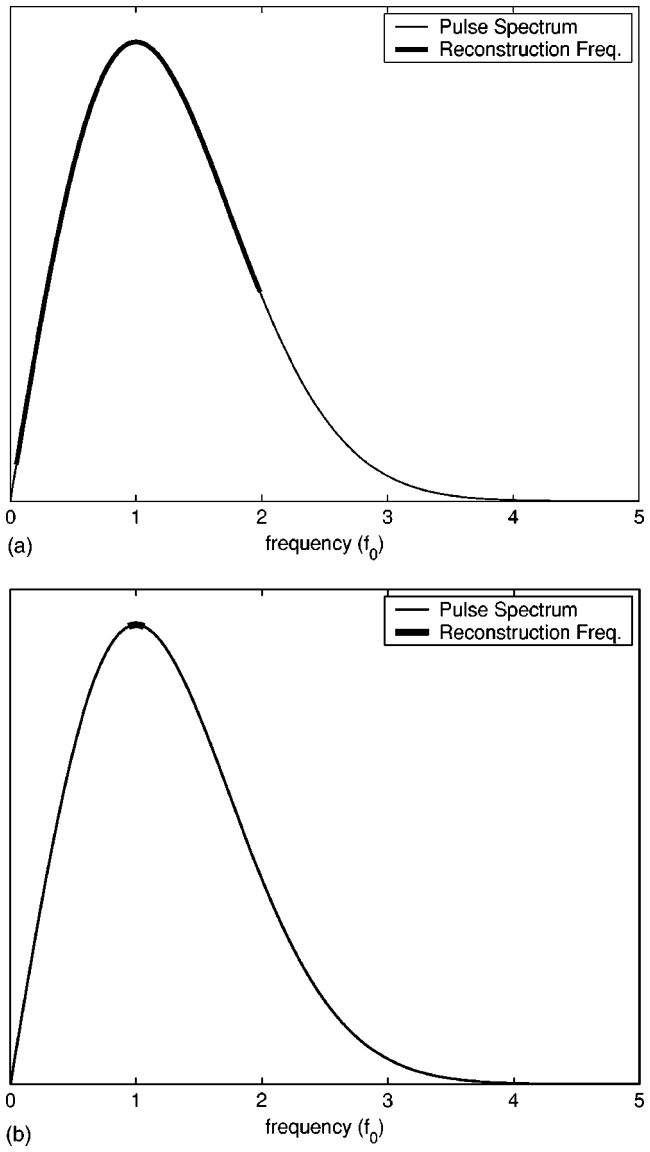


FIG. 11. FDTD pulse spectrum. (a) DOG pulse spectrum with the 120 frequencies used in the multimonostatic reconstruction highlighted. (b) DOG pulse spectrum with the 8 frequencies used in the multistatic reconstruction.

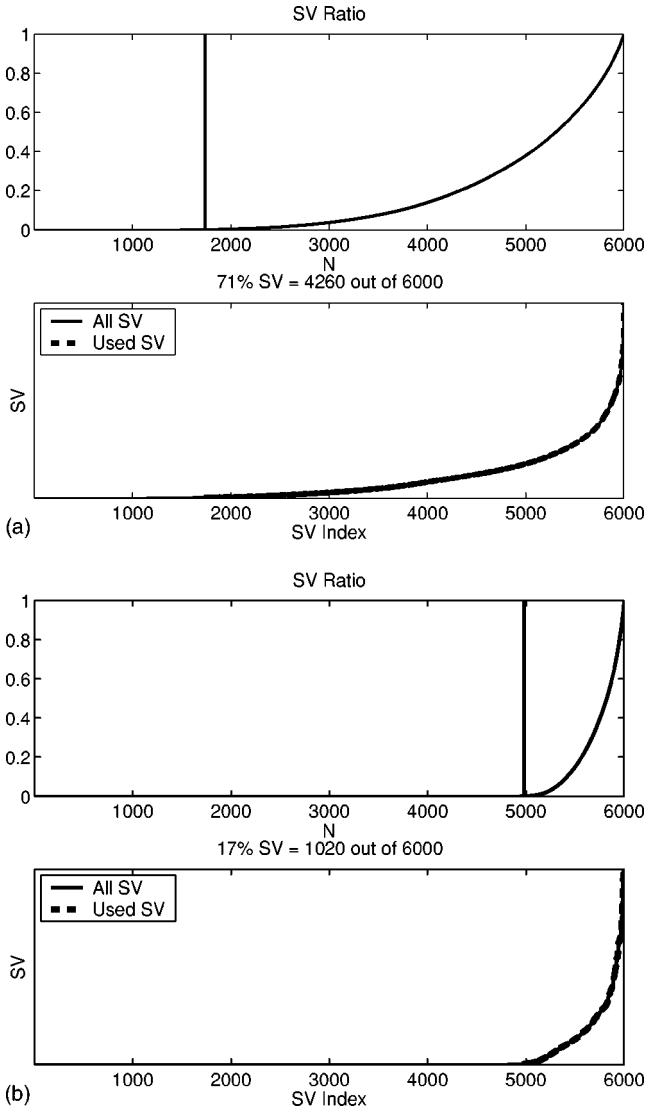


FIG. 12. FDTD simulation singular values. The top plots show the singular values ratios, $R(N)$, of Eq. (54). The vertical line shows where the selection was made distinguishing between the ''zero'' and nonzero singular values, thus determining the rank of the system. The bottom plots show the actual singular values. The highlighted trace shows those singular values used in the reconstructions. The (a) column is the multimonostatic case, the (b) column is the multistatic case.

occurs we label as J_0. With this value determined, our final reconstruction is

$$\hat{o}(\mathbf{r}) = \sum_{j=J_0}^{J-1} \frac{1}{\sigma_j^2} \Pi^T(\mathbf{r}) u_j u_j^\dagger D. \tag{55}$$

Our experience reconstructing both simulated and experimentally collected scattered field data has shown that this criterion works consistently well.

The HSIW is extremely flexible in that it allows any transducer configuration and any number of frequencies to be used in forming the reconstruction. The disadvantage is the potentially large amount of computer resources required for processing. Computing the outer product of Eq. (51) and its eigenvalues is time consuming for high-resolution reconstructions. Short-term disk storage space for the $\Pi(\mathbf{r})$ vectors of Eq. (34) was approximately 2 GB for the reconstructions of Sec. VII.

In the next section, we present the results of a proof-of-principle example based upon the exact Born approximation scattering model of Eq. (8).

VI. PROOF-OF-PRINCIPLE

Our proof-of-principle consists of two radial reflection simulations run under the Born approximation of Eq. (8). The geometry, shown in Fig. 5, consists of five point scatterers at a fixed radius but with increasing angular separation so as to achieve an angular resolution test. The first simulation was performed using a multimonostatic operating mode with 84 transducers and 100 reconstruction frequencies; the second used a multistatic operating mode with 20 sources colocated with 20 receivers and 21 frequencies. These numbers were selected so that in both cases the number of elements, J, in the projection kernel of Eq. (34), is 8400. This was done

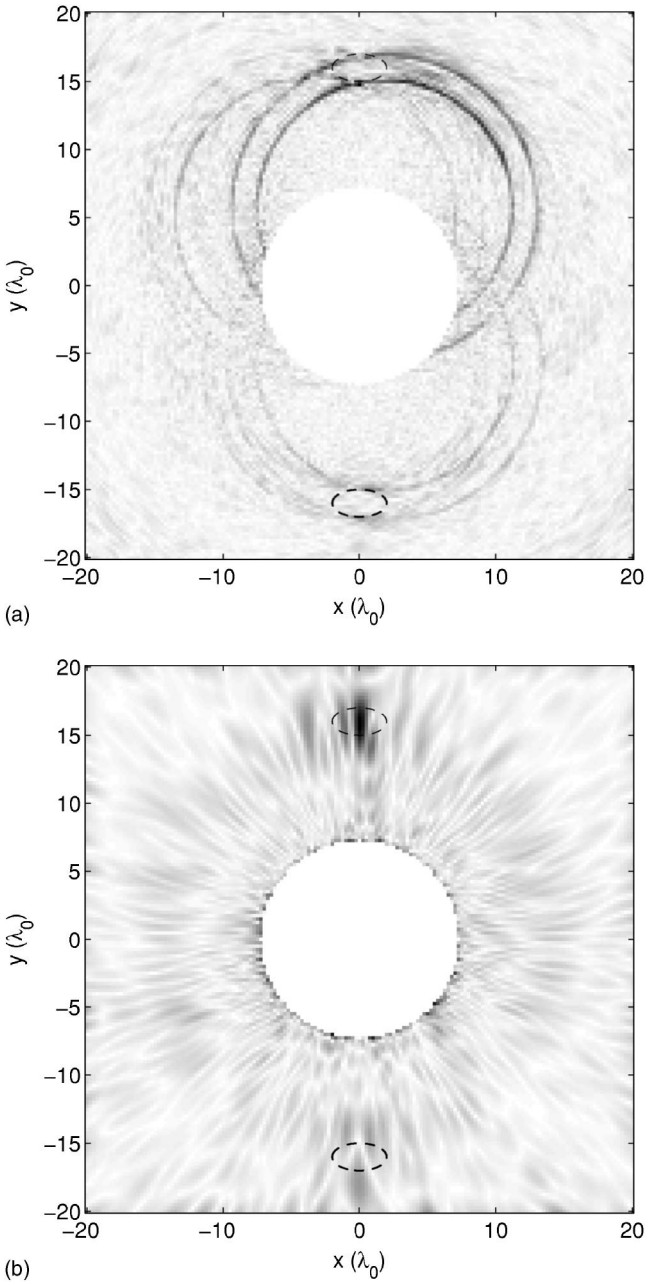


FIG. 13. Reconstructions of the FDTD simulations. (a) Multimonostatic using 71% of the singular values. (b) Multistatic using 17% of the singular values.

in order to equalize the comparison of the reconstruction results and minimize algorithmic differences. We did not model any shadowing due to the probe. That is, fields launched and/or received by transducers 180 deg away from a scatterer were able to propagated to/from the scatterer. The forward scattered field is computed via

$$\psi_B^{\text{scatt}}(\mathbf{R}_m^r,\mathbf{R}_n^t,\omega_l)=P(\omega_l)k_0^2(\omega_l)\sum_{s=0}^{S-1}G(\mathbf{R}_m^r,\mathbf{X}_s,\omega_l)\times G(\mathbf{X}_s,\mathbf{R}_n^t,\omega_l), \tag{56}$$

where S is the number of point scatterers, $\mathbf{X}_s$ is the location of the sth scatterer, and $G(\mathbf{r},\mathbf{r}',\omega)\equiv(i/4)H_0^{(1)}(k_0(\omega)|\mathbf{r}-\mathbf{r}'|)$ is the two-dimensional Green function.

The incident pulse used in the simulation is described by the first derivative of a Gaussian, or "DOG" pulse, given by

$$p(t)=-e^{1/2}\,\omega_0(t-t_0)e^{-(\omega_0(t-t_0))^2/2}, \tag{57}$$

where $\omega_0\equiv 2\pi f_0$ and f_0 is the peak frequency of the pulse in hertz. The pulse is presented in Fig. 6(a). Figures 6(b) and (c) show the pulse spectrum with the frequencies used in the multimonostatic and multistatic reconstructions, respectively. Figure 7 shows the singular value ratio, $R(N)$, of Eq. (54). The selection of the number of singular values to use was performed manually and the threshold is shown in the figure. The multimonostatic reconstruction used 5880 of the 8400 eigenvalues (70%), and the multistatic required 2100 (25%).

The reconstructions are shown in Fig. 8. The multimonostatic reconstruction reveals better resolution and somewhat better contrast over the multistatic reconstruction. This reflects the greater number of frequencies used in the reconstruction. As shown in Fig. 3, the more frequencies used results in a larger area in the object's spatial spectrum being reconstructed. For point scatterers, the multimonostatic operating mode is sufficient. The results of the next section demonstrate that this mode is insufficient for imaging extended scatterers under realistic wave propagation conditions.

VII. FULL WAVE SIMULATION

We performed a two-dimensional finite-difference time-domain (FDTD) simulation of wave propagation in a domain similar to that of the proof-of-principle example. We used E3D, an explicit 2D/3D elastic/acoustic propagation code designed for modeling seismic waves, developed at Lawrence Livermore National Laboratory.[6–11] The code simulates full wave scattering and requires as inputs a longitudinal velocity distribution, a shear velocity distribution, and a density distribution. In this case, unlike the proof-of-principle, the probe was included in the simulation, shadowing fields from transducers and scatterers with large angular separation.

The FDTD simulation domains are shown in Fig. 9. The domains consist of a homogeneous background with two elliptical inclusions. One inclusion's longitudinal velocity and density are 15% above the background; the other's are 5% below. There is no variation in shear velocity. The dashed line indicates the transducer outline.

As in the proof-of-principle example, we performed both a multimonostatic and multistatic simulation. The former used 50 transducer locations with 120 reconstruction frequencies resulting in $J=50\times 120=6000$ elements in the forward propagation kernel. The latter used 50 source locations and a receiving aperture of 15 transducers centered about the active source. As the transmitting element rotated, so did the receiving aperture, as indicated in Fig. 1(c). The exact multistatic configuration for this simulation is shown in Fig. 10. The multistatic reconstruction used 8 frequencies to achieve a $J=50\times 15\times 8=6000$ element kernel. As in the proof-of-principle example, we used kernels with identical number of elements to minimize variability when comparing

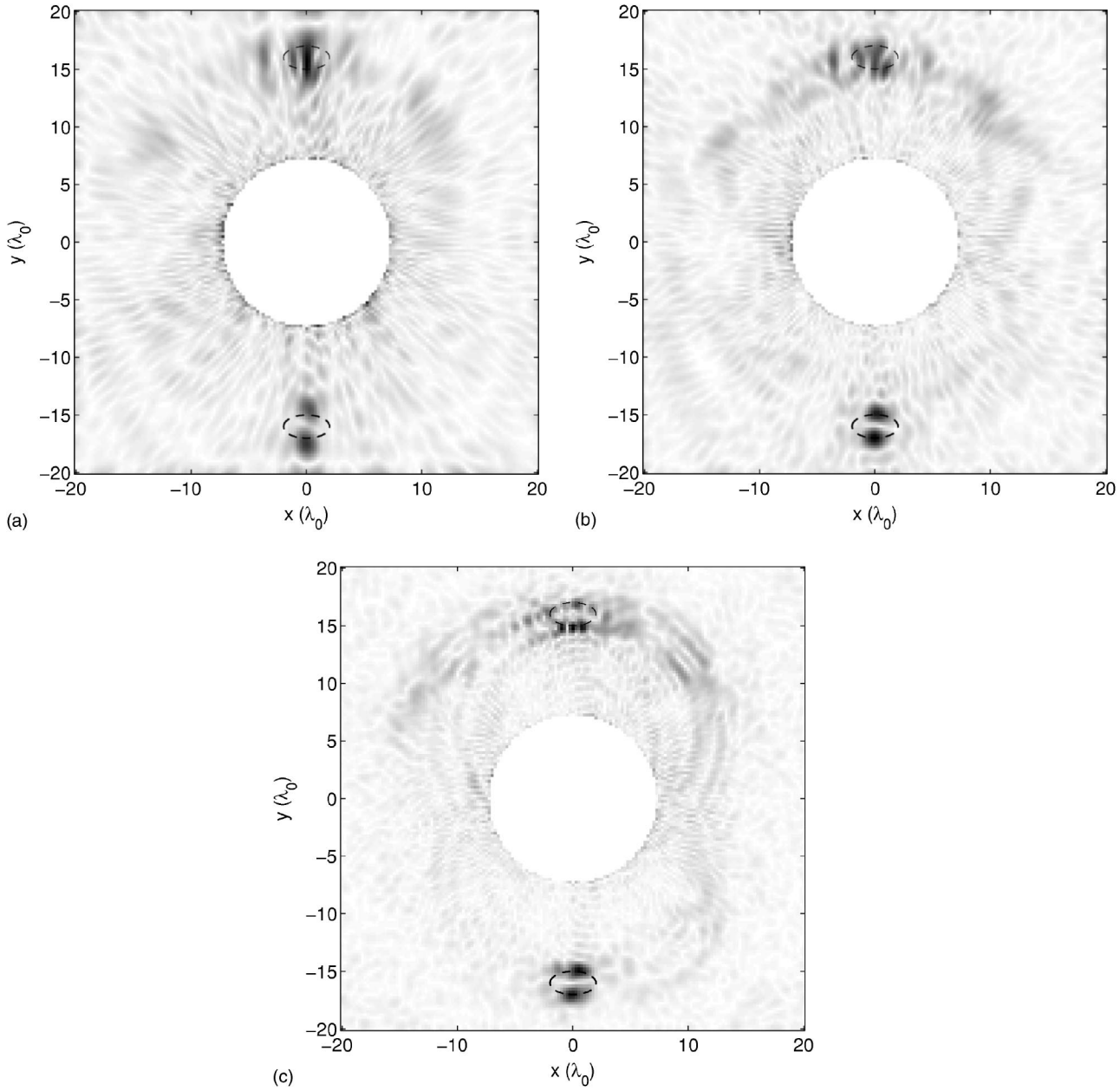


FIG. 14. Additional multistatic reconstructions to evaluate the trade-off between receiving aperture size and the number of frequencies used in the reconstruction. Radial resolution increases with the number of frequencies used, and azimuthal resolution increases aperture size.

the quality of the reconstructions. The simulation configurations and parameters are listed in the first two lines of Table I.

The full wave simulations used the same DOG pulse as the proof-of-principle example. The frequencies used in the reconstruction for each of the two cases are presented in Fig. 11. Again, we used the best rank N approximation to determine the number of singular values and vectors to use in the reconstructions. The singular value ratio, $R(N)$, and the singular value distribution are plotted in Fig. 12. The multimonostatic case required 4260 (71%) of the singular values, and the multistatic required 1020 (17%). The magnitude of the reconstructions is shown in Fig. 13. The high-pass nature of reflection mode diffraction tomography as represented in Fig. 3 is demonstrated in the multimonostatic reconstruction of Fig. 13(a), where the edges of the scattering objects are clearly highlighted. There is, however, an azimuthal ambiguity resulting from the use of only one transducer: off-radial axis scattering can only be resolved to within a fixed range from the transducer. Thus, off-radial axis scattering limits the usefulness of multimonostatic operation. Better azimuthal localization is achieved when more receiving transducers are included, widening the angular receiving aperture, permitting the resolution of the azimuthal ambiguity. This is demonstrated in the multistatic reconstruction of Fig. 13(b). Observe, however, when fewer frequencies are used, range resolution is lost.

We studied the receiver number/reconstruction frequency trade-off further in the three other multistatic simulations listed in Table I. We decreased the number of receivers and increased the number of reconstruction frequencies while keeping the size of the problem within our computer

resources. The reconstructions are shown in Fig. 14. We confirm that scatterer azimuthal localization improves with the number of receivers, but there is a loss of radial resolution when fewer frequencies are used. We conclude that azimuthal resolution is achieved by increasing the number of receivers, whereas range resolution increases with the number of frequencies.

VIII. CONCLUSIONS

We set out to develop a wave-based tomographic imaging tool based upon an intravascular ultrasound probe model in which a single rotating transducer both ensonifies the surrounding medium and measures the backscattered field. Our goal was to develop a wave-based inversion algorithm to image the surrounding medium rather than using the standard B-scan currently used that is based upon straight ray sound propagation.

We developed an analytic inverse for the multimonostatic configuration but were unable to implement it numerically. We then implemented a purely numerical inverse based upon the new Hilbert space inverse wave (HSIW) algorithm. The flexibility of the HSIW is such that it allows any configuration of transducers and frequencies to be used in the reconstruction. This permitted us to test not only a multimonostatic operating mode but also various multistatic modes.

We concluded that the multimonostatic mode lacks sufficient spatial diversity to image correctly scattering targets, that scatterer azimuthal localization improves as more receivers are used, and that range resolution improves as more frequencies are used.

ACKNOWLEDGMENT

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

APPENDIX: RELATIONSHIP BETWEEN THE FOURIER & FOURIER-BESSEL TRANSFORM

We wish to relate the Fourier–Bessel diffraction theorem of Eq. (27) to the Fourier diffraction theorem so that we may understand how the object's spatial spectrum is covered by the radial reflection measurement system. In order to do so, we require a relation between the Fourier–Bessel transform of the object to its Fourier transform. We consider a two-dimensional function represented in Cartesian coordinates as $f(x,y)$ or in polar coordinates as $f(r,\theta)$, where we have the following relationship between the coordinate pairs:

$$x = r\cos\theta,$$
$$y = r\sin\theta. \tag{A1}$$

The Fourier–Bessel transform pair is

$$F_m(k) = \frac{1}{2\pi}\int_0^{2\pi} d\theta \int_0^{\infty} r\,dr\, f(r,\theta) e^{-im\theta} J_m(kr), \tag{A2}$$

$$f(r,\theta) = \sum_{m=-\infty}^{\infty} \int_0^{\infty} k\,dk\, F_m(k) e^{im\theta} J_m(kr), \tag{A3}$$

and the Cartesian Fourier transform pair is

$$F(k_x,k_y) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, F(x,y) e^{-i(k_x x + k_y y)}, \tag{A4}$$

$$f(x,y) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y\, F(k_x,k_y) e^{i(k_x x + k_y y)}. \tag{A5}$$

The Cartesian Fourier transform pair is cast into polar coordinates using the following change of variables:

$$k_x = k\cos\phi,$$
$$k_y = k\sin\phi, \tag{A6}$$

to yield

$$F(k,\phi) = \int_0^{2\pi} d\theta \int_0^{\infty} r\,dr\, f(r,\theta) e^{-ikr\cos(\theta-\phi)}, \tag{A7}$$

$$f(r,\theta) = \frac{1}{(2\pi)^2}\int_0^{2\pi} d\phi \int_0^{\infty} k\,dk\, F(k,\phi) e^{ikr\cos(\theta-\phi)}. \tag{A8}$$

To determine the relationship between $F(k,\phi)$ and $F_m(k)$, we substitute the expansion

$$e^{-ikr\cos\theta} = \sum_{m=-\infty}^{\infty} (-i)^m J_m(kr) e^{im\theta}$$

into Eq. (A7) to find

$$F(k,\phi) = \sum_{m=-\infty}^{\infty} (-i)^m e^{im\phi} F_m(k). \tag{A9}$$

Using

$$\frac{1}{2\pi}\int_0^{2\pi} d\theta\, e^{i(m-n)\theta} = \delta_{mn},$$

we may invert Eq. (A9) to find

$$F_m(k) = \frac{(-i)^{-m}}{2\pi}\int_0^{2\pi} d\phi\, F(k,\phi) e^{-im\phi}. \tag{A10}$$

[1] A. J. Devaney, "A filtered backpropagation algorithm for diffraction tomography," Ultrason. Imaging **4**(4), 336–350 (1982).
[2] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging* (IEEE, New York, 1988).
[3] K. J. Langenberg, "Applied inverse problems for acoustic, electromagnetic and elastic wave scattering," in *Basic Methods of Tomography and Inverse Problems*, edited by P. C. Sabatier, Malvern Physics Series (Hilger, London, 1987), Chap. II, pp. 125–467.
[4] P. G. Yock, P. J. Fitzgerals, and Y. Honda, "Intravascular Ultrasound," in *Textbook of Interventional Cardiology*, edited by E. J. Topol (Saunders, Philadelphia, 1999).
[5] A. J. Devaney and M. Dennison, "Inverse scattering in inhomogeneous background media," Inverse Probl. **19**, 855–870 (2003).
[6] A. R. Levander, "Fourth-order finite-difference p-sv seismograms," Geophysics **53**, 1425–1436 (1988).
[7] S. Larsen and J. Grieger, "Elastic modeling initiative. III. 3-D computational modeling," Soc. Expl. Geophysi. Confer. Proc. **68**, 1803–1806 (1998).
[8] S. Larsen and M. Antolik, "3D simulations of scenario earthquakes in the

San Francisco Bay area,'' EOS Trans. Am. Geophys. Union **78**, 487 (1997).
[9] R. Madariaga, ''Dynamics of an expanding circular fault,'' Bull. Seismol. Soc. Am. **66**(3), 639–666 (1976).
[10] J. Virieux, ''P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method,'' Geophysics **51**(4), 889–901 (1986).
[11] S. Larsen and D. Harris, ''Seismic wave propagation through a low-velocity nuclear rubble zone,'' Technical report, Lawrence Livermore National Laboratory, Livermore, CA, 1993.
[12] G. Barton, *Elements of Green's Functions and Propagation—Potentials, Diffusion and Waves* (Oxford University Press, Oxford, 1991).
[13] S. J. Norton and M. Linzer, ''Ultrasonic reflectivity imaging in three dimensions: Exact inverse scattering solutions for plane, cylindrical, and spherical apertures,'' IEEE Trans. Biomed. Eng. **BME-28**(2), 202–220 (1981).
[14] A. J. Devaney, ''Geophysical diffraction tomography,'' IEEE Trans. Geosci. Remote Sens. **GE-22**(1), 3–13 (1984).
[15] M. P. André, P. J. Martin, G. P. Otto, L. K. Olson, T. K. Barrett, B. A. Spivey, and D. A. Palmer, ''A new consideration of diffraction computed tomography for breast imaging: Studies in phantoms and patients,'' Acoust. Imaging **21**, 379–390 (1995).

A phenomenological model of peripheral and central neural responses to amplitude-modulated tones

Paul C. Nelson and Laurel H. Carney[a)]
Department of Bioengineering and Neuroscience and Institute for Sensory Research, Syracuse University, Syracuse, New York 13244

(Received 22 March 2004; revised 22 June 2004; accepted 30 June 2004)

A phenomenological model with time-varying excitation and inhibition was developed to study possible neural mechanisms underlying changes in the representation of temporal envelopes along the auditory pathway. A modified version of an existing auditory-nerve model [Zhang *et al.*, J. Acoust. Soc. Am. **109**, 648–670 (2001)] was used to provide inputs to higher hypothetical processing centers. Model responses were compared directly to published physiological data at three levels: the auditory nerve, ventral cochlear nucleus, and inferior colliculus. Trends and absolute values of both average firing rate and synchrony to the modulation period were accurately predicted at each level for a wide range of stimulus modulation depths and modulation frequencies. The diversity of central physiological responses was accounted for with realistic variations of model parameters. Specifically, enhanced synchrony in the cochlear nucleus and rate-tuning to modulation frequency in the inferior colliculus were predicted by choosing appropriate relative strengths and time courses of excitatory and inhibitory inputs to postsynaptic model cells. The proposed model is fundamentally different than others that have been used to explain the representation of envelopes in the mammalian midbrain, and it provides a computational tool for testing hypothesized relationships between physiology and psychophysics. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1784442]

PACS numbers: 43.64.Bt [WPS] Pages: 2173–2186

I. INTRODUCTION

Physiological responses to amplitude-modulated (AM) stimuli have provided a basic description of how controlled fluctuations in a signal's temporal envelope are represented at several levels of the auditory pathway. The same class of stimulus has also been used extensively in psychophysical experiments, as a convenient means to investigate perceptual temporal processing capabilities and limitations. Sinusoidally amplitude-modulated (SAM) sounds are often used in both approaches because they can be described by a limited parameter set, and they are logical links between simple (tonal) stimuli and more spectro-temporally complex and biologically relevant signals such as speech. The abundance of data describing neural responses to SAM stimuli provides an opportunity to thoroughly test specific hypotheses concerning the physiological mechanisms that underlie envelope coding. Parallel psychological studies offer chances to better understand neural correlates of AM perception.

The current study uses computational modeling as a tool to quantitatively test realistic neural encoding hypotheses that may be used by the auditory system to code envelope modulations. Model responses will be directly compared to those reported in physiological studies of periodicity coding in the periphery, brainstem, and midbrain (auditory nerve, cochlear nucleus, and inferior colliculus, respectively). Some specific hypotheses that this approach will allow us to test are that (1) simple interactions between inhibition and excitation can give rise to modulation-tuned cells, and (2) physiologically realistic model parameter variations can account for the diversity of physiological AM responses. The broad goals of this study are to gain insights on coding strategies. As a result, the model is phenomenological in nature; we have not tried to include details of the underlying biophysical mechanisms in the model structure. Limitations of the predictive capabilities of the model we present here can provide useful clues for developing future experiments.

A. Extracellular physiological responses to AM stimuli

Typically, neural responses to AM stimuli are quantified based on (1) average firing rate and (2) synchrony (Goldberg and Brown, 1969) to the modulation period. These metrics are usually reported as a function of the stimulus modulation frequency (f_m), resulting in a physiological rate modulation transfer function (rate-MTF, average rate versus f_m), or a synchrony MTF (sync-MTF, synchrony versus f_m). For reference, a half-wave rectified sinusoid and a pure sinusoid (i.e., the envelope of a fully-modulated stimulus) are described by synchronization coefficients (SCs) of 0.784 and 0.5, respectively (or modulation gains of $+3.9$ and 0 dB for a fully modulated stimulus).

At the level of the auditory nerve (AN), the most complete set of data describing responses to SAM tones is that of Joris and Yin (1992) in the cat. High-characteristic frequency (CF) AN sync-MTFs for low to moderate stimulus levels are uniformly low-pass, with corner frequencies between 600 Hz and 1 kHz and peak modulation gains from 0 to $+4$ dB. Modulation corner frequencies are positively correlated with

[a)]Author to whom correspondence should be addressed. Electronic mail: lacarney@syr.edu

CF, indicating that temporal responses are dependent on the spectral interactions between AM signal components. Synchrony increases significantly with increasing stimulus modulation depth in AN fibers, and there is an optimal (moderate) stimulus level that elicits the most synchronous response (Joris and Yin, 1992; Smith and Brachman, 1980; Cooper *et al.*, 1993). High-spontaneous-rate (SR) fibers, which make up the majority of the AN-fiber population (Liberman, 1978), show little or no variation in rate with f_m, while average rates of most low-SR fibers tend to drop slightly as f_m increases (Joris and Yin, 1992). Encoding of AM in the AN is therefore widely assumed to be based on a temporal code, with stimulus modulation information carried in the periodicities of the poststimulus time (PST) histogram. However, in contrast to psychophysical measures, which are robust across level, AN PST histograms tend to flatten out at high stimulus levels, causing *decreased* response synchrony (Joris and Yin, 1992; Smith and Brachman, 1980). Central processing either recovers the diminished synchrony information at higher stimulus levels (possibly by combining information across audio frequencies), or uses some other feature of the AN neural response to encode envelopes.

Neurons in the cochlear nucleus (CN) can receive convergent information from multiple AN fibers, as well as from collaterals and descending inputs (both excitatory and inhibitory) (e.g., Ryugo and Parks, 2003; Schofield and Cant, 1996). The resulting MTFs are more complex than those measured in the AN. Sync-MTFs can be either low-pass or band-pass, and synchrony to the modulation period is often enhanced relative to AN fibers [Rhode and Greenberg (1994), cat; Frisina *et al.* (1990), gerbil]. Corner frequencies of CN sync-MTFs can be considerably lower than those measured in high-CF AN fibers, and are generally not correlated with the unit's CF. Most rate-MTFs remain relatively flat at the level of the ventral CN (VCN), with the exception of onset-choppers, which can exhibit rate-tuning to f_m (Rhode and Greenberg, 1994). Also, some AM response characteristics (i.e., shape of MTF and maximum synchrony values) are correlated with tone-based physiological characterizations of individual neurons (Rhode and Greenberg, 1994). For instance, many onset and chopper units are characterized by sync-MTFs that change from low-pass at low SPLs to band-pass at high SPLs (Frisina *et al.*, 1990; Rhode and Greenberg, 1994). Existing physiological models of AM processing rely on chopper cells in the VCN (e.g., Hewitt and Meddis, 1994); in the current study, we simulate bushy cells (which would be classified as primarylike or primarylike-with-notch based on their responses to tones) at the level of the VCN as an alternative pathway along which temporal envelope information could flow. The specific AM response characteristics that we have achieved in our physiologically realistic model bushy cells are (1) enhanced synchrony and (2) lower sync-MTF corner frequencies (*re*: AN inputs; Frisina *et al.*, 1990; Rhode and Greenberg, 1994). The low-pass to band-pass transition in the shape of the sync-MTF as sound level increases was not modeled, as this transition is not evident in VCN bushy cells (Frisina *et al.*, 1990; Rhode and Greenberg, 1994).

The inferior colliculus (IC), like the CN, is an almost obligatory synapse between lower brainstem nuclei and higher processing centers. Interactions between ascending and descending inputs potentially complicate the AM response properties of IC neurons. A systematic and thorough investigation of gerbil IC cell responses to SAM tones provided by Krishna and Semple (2000) has refined earlier work described by Langner and Schreiner (1988) in the cat. IC rate-MTFs often indicate rate tuning to specific modulation frequencies, and many cells exhibit suppressive f_m regions (Krishna and Semple, 2000). IC units are often highly synchronized to f_m, with most sync-MTFs taking low-pass or band-pass shapes. Best modulation frequencies (BMF, f_m eliciting largest response) based on rate are significantly lower (most are less than 100 Hz) than those observed in the VCN or AN based on synchrony. Increasing stimulus modulation depth generally increases the firing rate near rate-BMF, while changing the overall stimulus level has complicated and less systematic effects on rate-MTF shape (Krishna and Semple, 2000).

B. Intracellular responses and studies of inhibition

Inhibition could be an integral component of the neural processing that underlies the apparent transition from a temporal AM code in the periphery to a rate-based scheme at higher levels (Grothe, 1994; Faure *et al.*, 2003). Several aspects of the inhibitory circuitry in the VCN and IC have been previously investigated. In a pharmacological study of neurons in the chinchilla VCN, Caspary *et al.* (1994) selectively blocked inhibitory inputs and measured response areas (RAs) before and after neurotransmitter blockade. They found that response rates were most often affected near or at CF ($\sim$85% of neurons), and concluded that GABA and/or glycine-mediated inhibitory inputs are tonotopically aligned with excitatory inputs onto the same cell. Same-frequency inhibition and excitation (SFIE) is also suggested by morphological labeling studies, which indicate that the likely sources of inhibitory input (dorsal CN and superior olivary complex) are matched in frequency with their postsynaptic targets in the VCN (e.g., Wickesberg and Oertel, 1988).

Other studies have focused on understanding the time courses of inhibitory and excitatory influences in the VCN. Intracellular recordings in the brain-slice preparation of excitatory and inhibitory postsynaptic potentials (EPSPs and IPSPs) in response to current injections at the stump of the AN have provided details of these time courses that would not be possible to infer using extracellular methods (Oertel, 1983). Specifically, EPSPs of bushy cells in the VCN can be described as alpha functions with a time constant of approximately 0.5 ms, and IPSPs are slower, with time constants on the order of 2 ms. Differences in latencies between excitatory and inhibitory influences in these studies are also informative and consistent with the fact that all AN fibers are excitatory: inhibitory component onsets often begin 1–2 ms after excitation is observed and are therefore assumed to be disynaptic (Oertel, 1983). These observations can be built in to our model of neural processing between AN fibers and VCN neurons.

A similar framework of evidence exists for SFIE circuitry in the IC. Neurons originating in the dorsal nucleus of

the lateral lemniscus (DNLL) and projecting to the IC provide one source of GABA-ergic inhibitory input to the central nucleus of the IC (ICc) (Schneiderman *et al.*, 1988). Palombi and Caspary (1996) selectively blocked GABA in the ICc, and showed that rate increases occur mainly near CF in the absence of inhibition, similar to findings in the VCN. Brain slice intracellular recordings have suggested that synaptic modification of incoming spike trains in the IC lasts significantly longer than in the VCN (e.g., Wu *et al.*, 2002). Carney and Yin (1989) used extracellular responses of IC cells to clicks presented binaurally with an interaural time difference (ITD) to infer durations of inhibitory influence. They recorded from cells with a wide range of long-lasting inhibitory components that sometimes suppressed firing for tens of milliseconds following an initial click response.

C. Modeling

Previous modeling efforts of physiological AM coding have used a coincidence-detection mechanism at the level of the model IC cells to generate band-pass rate tuning for modulation frequency (Langner, 1981; Langner and Schreiner, 1988; Hewitt and Meddis, 1994). The elements of Langner's model, which is focused on pitch encoding instead of pure AM, are only loosely based on physiology. Neurons are hypothesized to perform a cross-correlation analysis between spike trains synchronized to the modulation frequency and the carrier frequency, with small delays between the inputs (Langner, 1981). The model loses its appeal, however, because it is only plausible at low carrier frequencies where phase-locking to the stimulus fine-structure is observed in inputs to the IC ($<$800 Hz, Yin, 2002).

The Hewitt and Meddis model for AM sensitivity is built around a population of intermediate ''chopper'' model VCN cells, whose parameters are adjusted to provide a set of envelope-locked inputs to postsynaptic model IC cells. In the framework of their model, AM-synchronized spikes in the VCN tend to fire at intervals related to their chopping period, and stimuli with AM fluctuations that are close in frequency to this inherent chopping elicit a more highly synchronous response. The synchrony-BMF (equivalently, the chopping period) of a model chopper cell is adjusted by varying a potassium-conductance time constant (i.e., Hewitt and Meddis, 1994) in a point-neuron model. One key assumption of this model structure is that the AM rate tuning observed in IC cells is determined by the synchrony tuning properties of VCN choppers that provide inputs to the IC. As a result, the range of rate-BMFs in model IC cells is determined by the range of synchrony-BMFs in their VCN inputs. By allowing the VCN potassium-conductance time constant to vary from 0.2 to 7 ms, Hewitt and Meddis (1994) showed responses from AM-tuned model cells with BMFs between 50 and 400 Hz. One issue for the Hewitt and Meddis model is the disparity in the ranges of VCN chopper synchrony-BMFs ($\sim$150–700 Hz; Rhode and Greenberg, 1994; Frisina *et al.*, 1990) and IC cell rate-BMFs ($\sim$1–150 Hz; Krishna and Semple, 2000).

Physiological work in the auditory system has suggested a possible role for envelope-locked inhibition in AM processing (i.e., Krishna and Semple, 2000; Caspary *et al.*, 2002), but the details of such a mechanism have not been agreed upon or quantitatively tested. We used the results from a modeling study in another modality as a starting point for the development of a new model that incorporates phasic inhibition. Krukowski and Miller (2001) implemented a model of temporal frequency tuning in the visual system in which inhibition dominates over excitation. These model neurons respond only to stimuli with envelope fluctuations because excitation and inhibition can occur out of phase with one another when the stimulus is modulated. Applied to the auditory system, we will show that an implementation of this mechanism can explain the transition from a temporal AM code in the periphery to a rate-based code at the level of the IC. A similar mechanism has recently been used to explain modulation coding at the level of the auditory cortex (Elhilali *et al.*, 2004).

In summary, the existence of physiological ''modulation filters'' has been experimentally shown. The details of the relationship between behavioral and neural AM responses remain unclear. We propose that one approach towards bridging this gap is to compare the predicted responses of a physiologically realistic computational model to actual AM responses. The specific model we have chosen to evaluate uses the convergence of long-duration inhibition and short-duration excitation as a mechanism for extracting modulation information from spike trains in lower levels.

II. METHODS

A. Auditory-nerve model

A new version of the physiologically based auditory-nerve (AN) model developed by Zhang *et al.* (2001) and modified by Heinz *et al.* (2001b) was used to simulate responses at the first level of neural coding. The modifications outlined below were included specifically to improve responses to AM tones, but they did not significantly affect responses to simple stimuli; model responses to tones are similar to those described in Heinz *et al.* (2001b, their model #3). Model code is available at web.syr.edu/$\sim$lacarney

1. New signal path filters

Effects of basilar membrane compression and two-tone-suppression were omitted in the modified linear version of the Zhang *et al.* (2001) AN model used in the present study. Spontaneous rate was adjusted as in the Heinz *et al.* (2001b) model to match the values described in the physiology, where available. We chose to use linear signal-path filters because of unresolved limitations of the nonlinear AN models at high CFs and high SPLs (discussed later). Tuning of the signal-path gammatone filters was adjusted based on properties of AN fibers in response to AM CF tones. Specifically, CF-dependent bandwidths of the sharp linear AN model (Heinz *et al.*, 2001b, model #3) were changed to account for the corner frequencies measured in low-pass sync-MTFs. For CFs above 1 kHz, the filter time constant was specified by

$$\tau = \frac{Q_{10}}{2\pi \mathrm{CF}}, \tag{1}$$

where the value of Q_{10} is estimated from a linear fit of measured values of Q_{10} (Miller *et al.*, 1997; see Zhang *et al.*, 2001). The value described here, which is an estimate of tuning at intermediate SPLs, is half the τ_{narrow} described in previous versions of the AN model (which was estimated from AN responses at low sound levels). The resulting high-CF model sync-MTFs have significantly higher corner frequencies than previous AN model versions, and are very close to those described in the cat by Joris and Yin (1992). For very low CFs (<1 kHz), the equation for τ_{narrow} described by Zhang *et al.* [2001, their Eq. (4)] is sufficient to account for the sync-MTF corner frequency in the small amount of data available from low-CF AN fibers.

2. Modified synapse model

Previous AN models have used a simplified implementation of Westerman and Smith's (1988) time-varying three-store diffusion model (Carney, 1993; Zhang *et al.*, 2001; Heinz *et al.*, 2001b) to describe the transformation of a stimulus-driven voltage to an instantaneous firing rate at the output of the IHC-AN synapse. Only the immediate permeability is assumed to depend on the acoustic stimulus (through the inner-hair-cell voltage). The parameters of these models are determined based on desired properties of high-CF PST histograms, including spontaneous rate (SR), steady-state discharge rate, and onset adaptation time constants at high sound levels. Once the initial parameters are set, the immediate permeability described in Westerman and Smith (1988) determines the (level-dependent and time-varying) effective time courses of the two adaptation processes. High-level stimuli result in smaller rapid and short-term time constants.

There is a basic limitation of these prior implementations that directly affects envelope-locked responses to AM tones. Recovery from prior stimulation (offset adaptation) in the models has two components and follows time courses that are set by the rapid and short-term components of onset adaptation. As a result, the final output of the previous synapse model (the product of an immediate permeability and concentration) recovers too quickly after tone stimulation. Physiological AN recordings (Smith, 1977; Smith and Zwislocki, 1975; Harris and Dallos, 1979) suggest that fibers with medium- or high-spontaneous rates recover more slowly than onset time constants would dictate, with a level-dependent dead time and a time constant longer than that of short-term onset adaptation.

Zhang (2004) has developed a strategy that effectively avoids these constraints on the synapse model. A SR-dependent shift is added to the desired rate response before implementing the difference equations that describe the diffusion model. After the differential equations are simulated on the up-shifted version of the presynaptic response, the resulting output is shifted back down by the same amount and half-wave rectified (so the final synapse output is never negative). By including the shift, offset recovery is effectively slowed because the early offset response is set to zero when the down-shifted output is half-wave rectified. The modification can be physiologically interpreted as a constant neurotransmitter leak in the synaptic cleft (Zhang, 2004).

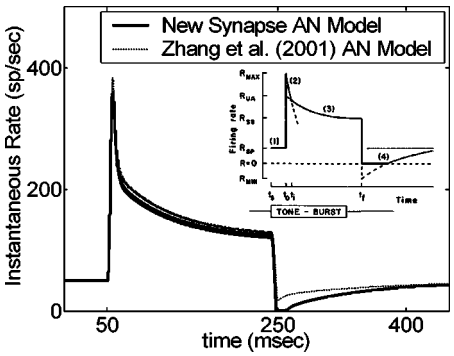


FIG. 1. AN model pure-tone responses (main panel) and schematic of AN responses from Harris and Dallos' (1979) physiological forward masking study (inset; with permission). Modification of the IHC-AN synapse portion of the Zhang *et al.* (2001) AN model resulted in longer (and more physiologically realistic) offset suppression. Model waveforms are responses of an 8-kHz CF fiber to a 25-dB SPL pure tone at CF.

For all of the simulations presented here, the shift (which was the same in amplitude for the upward shift as for the downward shift) was equal to twice the desired SR of the fiber. This value was a compromise: larger shifts further increase the modulation gain, but also caused systematic variations in average firing rate with modulation frequency. A SR-independent shift was also considered, but steady-state rates of low-SR fibers at high sound levels were unrealistically high when a high shift value (e.g., shift$=120$ sp/s) was used for all model AN fibers. Figure 1 illustrates the effect of the $2\times$SR shift on a response to a 200-ms (8-ms $\cos^2$ rise/fall time), 25-dB SPL pure tone presented at CF for an 8-kHz AN model fiber with a SR of 50 sp/s. The final synapse outputs are shown for the current model version, as well as for the Zhang *et al.* (2001) AN model (their complete nonlinear model with compression and suppression). The effect most relevant to increasing modulation gain can be seen after the tone offset, where the response is suppressed below SR for about 150 ms in this example. If a fully modulated AM tone is thought of as a train of tone pips, the corollary effect would be to suppress firing in the envelope valleys or troughs, which is observed physiologically (Joris and Yin, 1992).

3. Input and output signals and other details

Inputs to the AN model were the instantaneous pressure waveforms of the stimuli in Pascals. The corresponding output was the time-varying discharge rate of the synapse model, $r_{\text{AN}}(t)$, in spikes/second. Characteristic frequency and spontaneous rate were adjusted and matched to single-fiber physiology (where available) to generate an individual fiber response. Simulation sampling rates varied between 75 and 200 kHz, depending on the CF of the AN fiber and the stimulus frequency. Stimuli had 25-ms rise-fall times, and both carrier and modulator were started in sine phase, unless explicitly stated otherwise. Detailed responses of a previous version of this phenomenological model have been compared to physiological data describing AN responses to tones, noises, and clicks (Zhang *et al.*, 2001). One goal of the current study was to compare responses of the new AN model to general trends and absolute values of en-

velope synchrony and average firing rate observed in cat AM physiology (neglecting the initial onset response, as in Joris and Yin 1992).

B. Model cochlear nucleus cells

Some basic physiological observations are incorporated into the generation of model responses at the first levels of AN fiber convergence. First, membrane and channel properties of bushy cells in the VCN have the effect of low-pass filtering incoming spike trains (Oertel, 1983). VCN postsynaptic potentials were approximated by alpha functions of the form $P(t)=te^{-t/\tau}$, where values of τ were chosen based on Oertel's intracellular recordings of bushy cells. Inhibitory influences are slower and longer-lasting (i.e., IPSP time constants are longer, on the order of 2 ms) than excitatory input influences, which can be characterized by EPSP time constants of about 0.5 ms. Computationally, smoothed inputs to the model VCN cells were obtained by convolving $r_{\mathrm{AN}}(t)$ with the appropriate alpha function. The relative strengths of inhibition and excitation were adjusted by varying the area under the alpha function (or, equivalently, the gain of the low-pass filter pass-band). At this first level of convergence, the strength of excitation was assumed to be greater than that of inhibition. Because all AN fibers are excitatory, model inhibitory inputs to VCN cells were also slightly delayed ($D_{\mathrm{CN}}=1$ ms) relative to the excitation, to simulate a disynaptic delay along an intrinsic inhibitory pathway (Oertel, 1983).

Another physiologically based observation that is included in the simulated VCN cells is that many neurons in the anteroventral cochlear nucleus (AVCN) receive same-frequency inhibitory and excitatory (SFIE) inputs (see Sec. I). This mechanism was realized simply by subtracting a membrane-modified, inhibitory AN response from a corresponding smoothed excitatory AN response with the same CF. In terms of the AN input $r_{\mathrm{AN}}(t)$ and the approximated post-synaptic potentials, the VCN response is given by

$$r_{\mathrm{CN}}(t)=A[te^{-t/\tau_{\mathrm{exc}}}*r_{\mathrm{AN}}(t)-S_{\mathrm{CN,inh}}(t-D_{\mathrm{CN}}) \times e^{-(t-D_{\mathrm{CN}})/\tau_{\mathrm{inh}}}*r_{\mathrm{AN}}(t-D_{\mathrm{CN}})], \quad (2)$$

where different time courses of inhibition and excitation ($\tau_{\mathrm{exc}}=0.5$ ms, $\tau_{\mathrm{inh}}=2$ ms) are represented in the two convolution ($*$) terms, and same-frequency inhibition and excitation (SFIE) is included by using the same AN instantaneous discharge rate for both inputs. The resulting difference was half-wave rectified so that negative rates caused by the subtraction were represented by zero firing in the resulting model VCN cell instantaneous firing rate. The scalar $S_{\mathrm{CN,inh}}$ sets the strength of inhibition relative to the excitation. The scalar A was taken into account to achieve realistic average rates; this value was set to 1.5 for the simulations presented here.

C. Model inferior colliculus cells

Responses of model VCN cells were used as inputs to higher levels of convergence in a hypothetical processing cascade. Synaptic modifications, input frequency convergence, and relative delays between the inputs were implemented as in the first level of processing. Mathematically, Eq. (2) was used to derive IC model responses, with $r_{\mathrm{AN}}(t)$ being replaced by $r_{\mathrm{CN}}(t)$, and $A=1$. One important difference between the two levels of processing was the strength of inhibition: model IC neurons had stronger inhibition than excitation ($S_{\mathrm{IC,inh}}>1$), while model VCN cells were dominated by their excitatory inputs ($S_{\mathrm{CN,inh}}<1$). As a result, the IC model cell responded strongly only when the timing of the two inputs was such that a peak of excitation coincided with a valley of the envelope-locked inhibition.

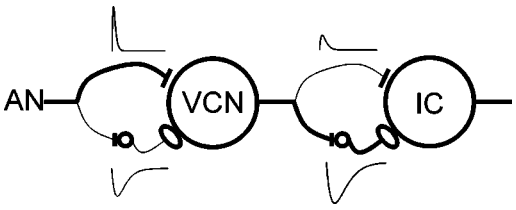


FIG. 2. Schematic diagram of the same-frequency inhibition and excitation (SFIE) model. A single model AN fiber provides the postsynaptic cell with both excitatory and inhibitory input, via an inhibitory interneuron. The thickness of the lines corresponds to the relative strength of the inhibition and excitation at each level. Alpha functions representing the assumed membrane and synaptic properties are also shown.

Because the time courses of inhibitory influences at the level of the IC are more variable than those observed in the VCN (e.g., Carney and Yin, 1989), the inhibitory time constant in the second level of model cells was systematically varied. Consequences of this variation are described in Sec. III. The single inhibitory time constant of the model should not necessarily be interpreted as a simple time constant of a single IPSP, but rather as an effective time constant describing the overall time-course of the inhibitory influence (which could be, for example, a train of IPSPs rather than a unitary IPSP). A schematic diagram of the two-layer SFIE model is shown in Fig. 2.

III. RESULTS

A. AN model responses

Joris and Yin's (1992) systematic study of cat AN responses to SAM tones provides an excellent template for detailed evaluation of AN model responses. Physiological and model AN sync-MTFs and rate-MTFs of high-CF fibers are stereotyped in their shape: sync-MTFs are low-pass at low to moderate signal levels with -3-dB cutoffs between 600 and 1000 Hz, and rate-MTFs of high-SR AN fibers show little variation in average rate with f_m. Also, both model and actual AN responses at low f_m can be more modulated than a half-wave rectified version of the stimulus (modulation gain>0).

Physiology and model predictions for a single high-CF AN fiber are illustrated in Fig. 3 in terms of modulation gain, average firing rate, and PST histogram shape. Model fiber parameters were matched to the physiological fiber description (CF and SR), and stimulus conditions (carrier frequency and relative level) were also duplicated. The main effect of modifying the AN-IHC synapse portion of the AN model was to increase the envelope synchrony (or equivalently, modulation gain) to physiologically realistic values. Previous

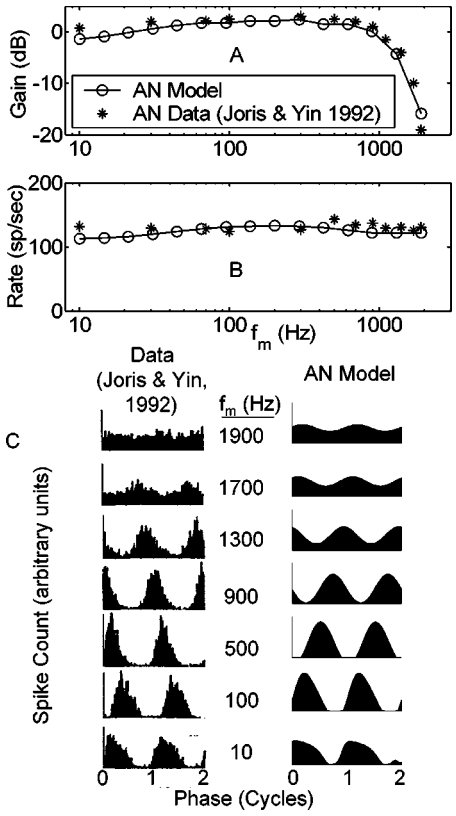


FIG. 3. Comparison of model and actual modulation transfer functions and period histograms. (a) Physiological and model AN sync-MTFs for a high-CF fiber. (b) Rate-MTFs for the same fiber (onset response neglected). (c) Period histograms for the actual AN fiber (left column) and model AN fiber with matched parameters. Two cycles of the response are shown starting at a zero sine phase point of the modulating waveform. CF=21 kHz; SR=61 sp/s. Stimulus carrier frequency=CF; SPL set at ~20 dB above rate threshold. Cat data from Joris and Yin (1992, their Fig. 10; with permission).

AN model versions had maximum modulation gains of approximately −5 dB at low SPLs [the full nonlinear version of the Zhang *et al.* (2001) model predicts a SC of 0.29 for a fully modulated stimulus], while the version described here has a peak gain of +2.5 dB (SC=0.66). The low-pass shape and corner frequency of the sync-MTF are consistent with the data [Fig. 3(a)]; the flat rate-MTF and steady-state average rate values are also well-predicted [Fig. 3(b)]. Shapes of the PST histograms are not strictly determined by measurements of synchrony and rate; however, the model responses are also qualitatively similar in PST shape to those measured physiologically [Fig. 3(c)].

Another characteristic of AN fiber responses to AM can be observed by varying the modulation depth (m) at a given f_m. The effects of changes in m are important to test with the model for our long-term purposes because psychophysical data is often reported in terms of a threshold m that is always less than $m=1$ (i.e., less than 100% stimulus modulation). Physiological and modeled response synchrony and modulation gain are plotted against stimulus modulation depth for a high-CF, high-SR AN fiber in Fig. 4. Predictions based on the modified synapse AN model fit the data quite well, and are a significant improvement over the previous model version. In addition, the model responses (not shown)

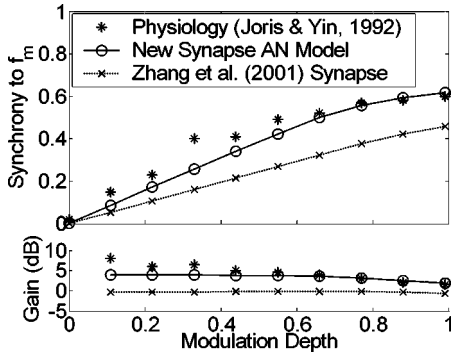


FIG. 4. Changes in synchronization to the modulation period with stimulus modulation depth for model and actual AN fibers [physiological responses from Joris and Yin, (1992), their Fig. 1(b); with permission]. Predictions from two AN model versions are shown; the version with the modified synapse model will be used for all subsequent simulations. Fiber parameters: CF=20.2 kHz; SR=53 sp/s. Stimulus parameters: carrier frequency=CF; f_m=100 Hz; SPL set to elicit the maximum synchrony value (20 dB SPL for the model fiber; 49 dB SPL for the actual fiber).

are very similar in shape to the PSTs illustrated in Joris and Yin (1992, their Fig. 1). For direct evaluation of the contribution of the new version of the AN-IHC synapse model, predictions are also shown using the same presynaptic model structure in combination with the Zhang *et al.* (2001) synapse model. Synchrony using the new ''constant neurotransmitter-leak'' model is significantly enhanced over the previous version for all modulation depths.

In addition to comparing measured and simulated response characteristics for different modulation frequencies and depths, we have also evaluated the level-dependence of the AN-model synchrony and rate. Physiologically, modulated synchrony-level functions (synchrony versus level for a given f_m) are non-monotonic, with a peak at some best modulation level (BML, usually 10–20 dB above rate threshold) over a wide range of modulation frequencies (Joris and Yin, 1992). The AN model used in this study is characterized by the same level-dependent synchrony trends. Figure 5 compares physiology (unconnected symbols) and simula-

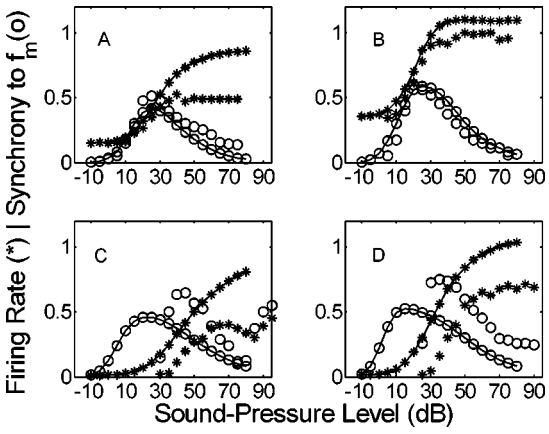


FIG. 5. Comparison of actual (unconnected symbols) and model (connected symbols) rate-level (*) and synchrony-level (○) functions for four combinations of AN parameters. Firing rates are normalized by 200 sp/s. CF, SR, and f_m: (a) 510 Hz, 30 sp/s, 10 Hz, (b) 12 kHz, 71 sp/s, 100 Hz, (c) 710 Hz, 2 sp/s, 10 Hz, (d) 8.1 kHz, 2.6 sp/s, 100 Hz. The carrier frequency was set to the fiber's CF. Physiological responses replotted from Joris and Yin (1992, their Fig. 5; with permission).

tions (connected symbols) of responses to four combinations of fiber CF, SR, and stimulus f_m in terms of average rate (*) and synchrony (○) to the modulation period. The general shapes of both rate-level (monotonic and saturating) and sync-level functions (non-monotonic) are accurately predicted by the AN model. Thresholds in the model fibers are set near 0 dB SPL, regardless of SR; as a result, the simulated low-SR rate-level and sync-level functions (lower two panels) are shifted toward lower sound levels. Peak firing rates are higher in the AN model low-SR and low-CF groups in comparison to Joris and Yin's (1992) physiological results [panels (a), (c), and (d) of Fig. 5]. This is an inherent property of the model, and it is one that we chose not to focus on in our efforts to match the AM response properties. However, the rates are within a reasonable range when population data from both Joris and Yin (1992) and Liberman (1978) are considered. Low-SR model fibers exhibit slightly lower peak envelope synchrony and broader synchrony-level functions than the actual AN fibers [panels (c) and (d)]. The high-SR, high-CF model fiber shown in panel (b) of Fig. 5 is accurate in terms of its rate threshold, dynamic range, maximum rate, and synchrony-level function. It is this class of AN fibers that is described most thoroughly in Joris and Yin's study, and it is responses from these fibers that were used to test and validate the AN model.

Joris and Yin (1992) also quantified the relationship between a fiber's sync-MTF corner frequency and its tuning-curve bandwidth. High-frequency AN fibers have broader tuning than low-CF fibers, which allows more AM sideband energy to pass for relatively higher stimulus f_m. Accordingly, -3-dB cutoffs increase with CF. Interestingly, at very high CFs (>15 kHz in the cat) the positive relationship between CF and cutoff frequency saturates (the corner frequency becomes independent of CF). The AN model predicts similar trends, although the absolute values of the model cutoff frequencies are slightly lower than the average physiologically measured values (but model values lie within the scatter of the data; Joris and Yin, 1992, their Fig. 14). Figure 6 shows examples of model sync-MTFs at three CFs along with those from comparable AN fibers (from Joris and Yin, 1992).

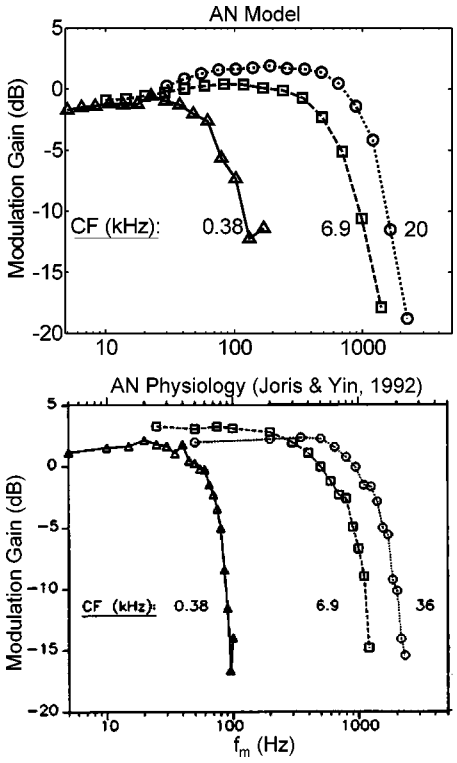


FIG. 6. Sync-MTFs ($m=0.99$) for three fibers with different CFs: AN model and actual data [from Joris and Yin (1992), their Fig. 13a; with permission]. SR of each fiber; triangles: 24 sp/s; squares: 6.3 sp/s; circles: 39 sp/s. Model SPL=24 dB.

B. Model CN cell responses

Figure 7 illustrates simulations of the effect of the convergence of slow inhibition and fast excitation with the same CF on a postsynaptic model VCN cell. The model's synchronous response to SAM is affected in two ways that are consistent with physiology (Rhode and Greenberg, 1994; Frisina *et al.*, 1990): (1) synchrony is enhanced with respect to AN inputs at low modulation frequencies, and (2) the upper frequency limit of phase-locking to the envelope is lower in the model VCN cells. The model parameter that most directly contributes to these response characteristics is the strength of inhibition (relative to excitation). MTFs for model VCN cells with a range of inhibitory to excitatory strength ratios between 0 and 0.6, along with the input AN MTFs, are shown in Fig. 7. Average rates decrease for model cells with stronger inhibition (for all f_m), and rate-MTFs can exhibit a shallow peak for the model cells receiving the strongest inhibitory inputs (lower panel). Rates in VCN model cells with no inhibition are higher than the AN inputs because the excitatory alpha-function area (or low-pass filter gain) was greater than unity (area=1.5). Also, as inhibition increases, VCN sync-MTFs systematically have higher corner frequencies, and maximum modulation gain values increase (upper panel).

What is the specific mechanism that gives rise to enhanced envelope synchrony in the model VCN cells? The effect can be understood by considering the differences in assumed membrane and/or channel properties for inhibition

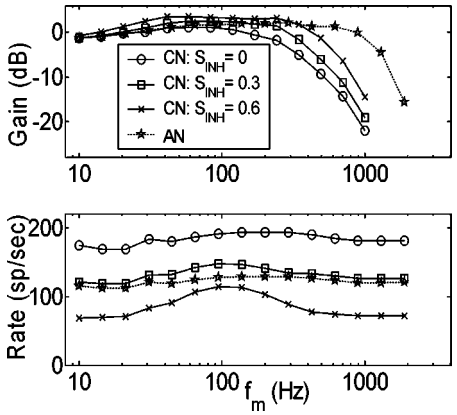


FIG. 7. Sync-MTFs (top panel) and rate-MTFs (bottom panel) for an 8-kHz CF AN model fiber, and three model VCN cells with varying strengths of inhibition (*re*: strength of excitation). Model AN fiber SR=50 sp/s. Model VCN $\tau_{exc}=0.5$ ms; $\tau_{inh}=2$ ms; $D_{CN}=1$ ms. Stimuli were presented at 24 dB SPL.

and excitation. Stronger and faster excitatory inputs interact with weaker and slower inhibitory inputs in ways that depend on the stimulus f_m to produce the final model VCN cell response. For low and intermediate modulation frequencies (i.e., 10 Hz$<f_m<$200 Hz), the more sluggish inhibitory inputs ($\tau_{\text{inh}}=2$ ms) are not able to follow the fast fluctuations in the AN responses as faithfully as the excitatory inputs ($\tau_{\text{exc}}=0.5$ ms). Higher synchrony results because excitation dominates for a more focused time interval (near the modulation period onset). For $f_m>200$ Hz, the slow inhibitory component acts to effectively remove a DC offset from the excitatory component. This causes higher modulation gain values and higher sync-MTF cutoff frequencies for model cells that receive stronger inhibition. VCN sync-MTF corner frequencies are lower than those in the model AN fibers (regardless of the relative strength of inhibition) because they are limited by the excitatory synaptic properties that modify the ascending VCN model cell inputs.

For nearly equal-strength inhibition and excitation (CN $S_{\text{inh}}>0.6$, not shown), model VCN rate-MTFs are more band-pass in shape than has been reported in most physiological studies of bushy cells. When the overall strengths of the cell's two inputs are nearly the same, there is a narrow range of f_m for which the temporal interactions between inhibition and excitation cause high firing rates in the postsynaptic cell. At very low f_m and very high f_m, the two inputs are similar in their ability to keep up with the stimulus modulations, resulting in a low discharge rate since inhibition is nearly as strong as excitation. To avoid band-pass rate-MTFs, the strength of inhibition in the model VCN bushy cells was set to 0.6 when they were used to provide inputs to model IC cells.

C. Model IC cell responses

1. MTFs and effect of varying time constants and delays

A second layer of model cells receiving inhibition-dominated SFIE inputs is hypothesized to represent IC units that integrate information from many convergent inputs. The most basic results observed in the model responses are that IC cells fire only over some narrow range of f_m (i.e., they are rate-tuned to f_m), and their AM responses are highly synchronized to the modulation period. This is consistent with physiological responses in the gerbil (Krishna and Semple, 2000) and cat (Langner and Schreiner, 1988). The BMF of a given IC model cell is determined mainly by the time constants of the inhibitory and excitatory influences: fast-acting inputs give rise to high BMFs; slower time constants result in lower BMFs. We constrain the inhibitory τ to always be equal to or longer than the excitatory τ when generating model responses. The range of BMFs that can be obtained by varying these parameters over a physiologically realistic range is illustrated in Fig. 8. Each rate-MTF describes the responses of a model cell with a given combination of $\tau_{\text{IC,exc}}$ and $\tau_{\text{IC,inh}}$. The same model AN fiber and VCN cell provided inputs to each of these model IC cells (i.e., rate-tuning in the IC is not determined by synchrony-tuning in the VCN). Absolute rate (top panel) and normalized rate-MTFs (bottom panel) are shown. Note that a wide range of rate-BMFs

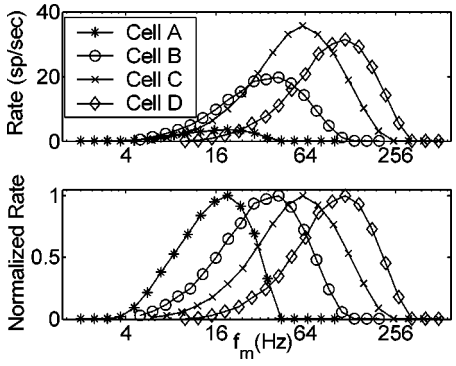


FIG. 8. Absolute and normalized rate-MTFs of four model IC cells with different combinations of excitatory and inhibitory time constants. $\tau_{\text{IC,exc}}$ and $\tau_{\text{IC,inh}}$; cell A: 5 ms and 10 ms; cell B: 2 ms and 6 ms; cell C: 1 ms and 3 ms; cell D: 1 ms and 1 ms. Common parameters to all four cells: AN CF=8 kHz; AN SR=50 sp/s; VCN $\tau_{\text{exc}}=0.5$ ms; VCN $\tau_{\text{inh}}=2$ ms; $D_{\text{CN}}=1$ ms; $S_{\text{CN,INH}}=0.6$; $D_{\text{IC}}=2$ ms; $S_{\text{IC,INH}}=1.5$. Stimulus carrier frequency = AN CF; $m=1$; SPL=24 dB.

(~40–120 Hz) can be obtained with a fixed $\tau_{\text{IC,exc}}$, and variation of $\tau_{\text{IC,inh}}$ from 1 to 7 ms. Tuning to even slower envelope frequencies is achieved by assuming longer time constants (cell A in Fig. 8, with $\tau_{\text{IC,exc}}=5$ ms and $\tau_{\text{IC,inh}}=10$ ms, is tuned to ~20 Hz). The upper limit of model BMFs (~120 Hz) is consistent with that observed in the gerbil IC (Krishna and Semple, 2000). This boundary is determined in the model by restricting the time constants of the excitatory and inhibitory inputs to be longer than 1 ms (Wu *et al.*, 2002; Carney and Yin, 1989).

The exact f_m that elicits the largest rate-based model response can be further adjusted by changing the delay between excitation and inhibition. When the inhibitory delay is kept within a physiologically realistic range (<10 ms), rate-based BMF shifts at low modulation frequencies are relatively small but systematic. Longer delays result in a shift of rate-MTF peaks to lower f_m and increases in rate at BMF. Short delays between inhibition and excitation cause maximal overlap in the envelope-locked inputs, and thus decrease the overall firing rate and increase the cell's BMF. Grothe (1994) has proposed a scheme for AM tuning in the bat medial superior olive (MSO) that uses a pure delay between inhibition and excitation to set the model cell's BMF. We chose not to rely on such a mechanism because of one specific consequence that is not observed in the physiology: a multi-peaked rate-MTF results, with rebounds in rate at envelope frequencies that are multiples of the ''fundamental'' f_m. This problem is avoided in the current model by the use of strong inhibition that is desynchronized at high f_m.

2. Effect of varying stimulus modulation depth

Responses of model IC cells change for stimuli with different modulation depths in a way that is consistent with physiological observations (Krishna and Semple, 2000). Figure 9 shows an example of physiological MTFs (top panels, from Krishna and Semple, 2000) and model MTFs (bottom panels); the parameter in the figure is stimulus m. Firing rate increases with modulation depth, most significantly near BMF. Synchrony saturates rapidly as m is increased. The possible shapes of the model IC sync-MTFs are limited be-

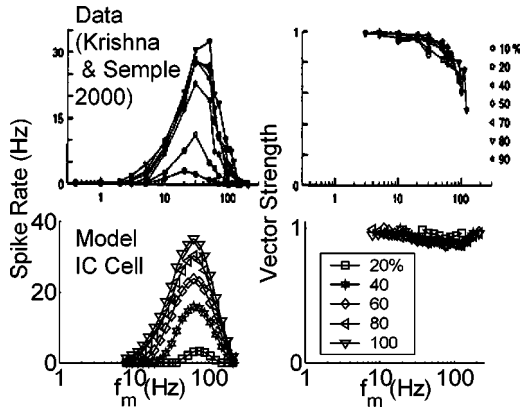


FIG. 9. Effect of varying stimulus m on actual (upper panels) and model (lower panels) IC unit MTFs. Model and stimulus parameters were the same as those describing cell C in Fig. 8. Rate-MTFs are shown on the left; sync-MTFs on the right. Gerbil physiological data from Krishna and Semple (2000, their Fig. 2C; with permission).

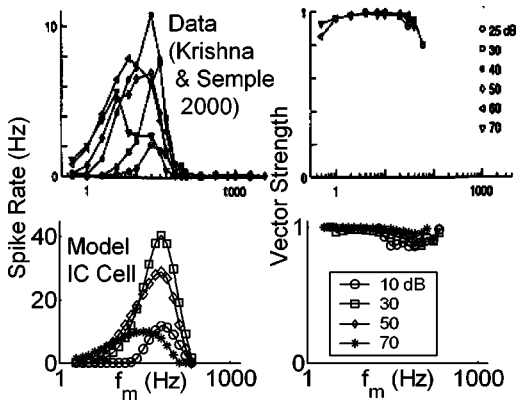


FIG. 10. Effect of varying stimulus SPL on actual (upper panels) and model (lower panels) IC unit MTFs. Model and stimulus parameters were identical to those described in Fig. 9 (with $m=1$). Data from Krishna and Semple (2000, their Fig. 5D; with permission).

cause a narrow range of f_m's elicits a high enough rate to compute a significant SC. Envelope locking persists in the response at modulation rates well above the high-frequency rate-MTF cutoff frequency. Each of these effects of varying m has been observed in physiological recordings (e.g., Krishna and Semple, 2000). One clear discrepancy between model and data is in the lack of a synchrony roll-off in the model at high f_m. This is due to the fact that the model is deterministic, and any response will be synchronized, even at very low rates. The addition of some amount of noise (i.e., jitter in the discharge times) to the model IC cell output would avoid the artificially high synchrony values for conditions with very low average rates (e.g., at high modulation frequencies for the model cell in Fig. 9). Figure 9 also illustrates the stereotyped shapes of model cell rate-MTFs. The high-frequency roll-off is more abrupt than the low-frequency transition on a logarithmic frequency axis, and rates go to zero at very high f_m. These features have also been described in gerbil IC recordings, and have been hypothesized (but not tested) to be caused by inhibitory inputs (Krishna and Semple, 2000).

3. Effect of varying stimulus SPL

IC cell rate-MTFs can change with sound level (e.g., Krishna and Semple, 2000). This point has important implications for relating physiological responses to psychophysical measures, which are almost always robust across a wide range of SPLs. Predicted rate-MTFs based on the IC model cells also depend on SPL in ways broadly consistent with some of the single unit physiology [although the physiological level-dependence can be very different from cell to cell (Krishna and Semple, 2000)]. Figure 10 compares responses from one cell in Krishna and Semple's (2000) study to those of the model IC cell illustrated in Fig. 9. Several trends are predicted by the model. (1) The largest peak rate response is elicited by a medium-SPL stimulus (40 dB in the physiology; 30 dB in the model). (2) The rate-BMF shifts to lower f_m with increasing sound level. (3) The bandwidth of the rate-MTF tends to increase with sound level. (4) Synchrony remains relatively high despite these changes in the rate-based characterization.

4. Model mechanisms and PSTs

In the present model, mechanisms underlying specific features of model rate-MTFs and sync-MTFs can be determined directly by considering the intermediate waveforms that shape the model IC cell response. At low f_m, neither the inhibitory nor the excitatory inputs are significantly modified by their low-pass membrane and synaptic properties. That is, inputs to the model IC cell are able to closely follow the VCN response. As a result, excitation and inhibition are overlapping for most of the stimulus cycle (depending on their relative delays), and since these cells are inhibition-dominated, the average firing rate is very low. For stimulus frequencies near the cell's BMF, interactions between inhibition and excitation become more interesting. The slower inhibitory inputs start to lose much of their AC response, and the weaker excitation will cause higher postsynaptic rates when its stimulus-locked peaks line up in time with the stronger inhibitory "valleys." At high f_m, the inhibitory inputs are nearly tonic, and the excitatory components are also smoothed by their effective synaptic and membrane filtering properties. In the extreme, at very high f_m, both inputs to the model IC cell are time-independent. The result is little postsynaptic activity because inhibition dominates, and temporal interactions can no longer provide stimulus-locked firing in the final response.

Despite the good agreement between available physiology and model IC cell responses in terms of their rate- and sync-MTFs, the predictive power of the model structure in describing PST histogram shapes has not yet been demonstrated. Examples of physiological (from Langner and Schreiner, 1988) and modeled temporal responses for a single IC unit at various modulation frequencies (with a stimulus rise-fall time of 5 ms) are shown in Fig. 11. The model cell was chosen such that the rate-BMF was similar to that of the physiological cell (BMF=60 Hz). A strong onset is present in both data and model responses, regardless of the stimulus f_m. This feature is not part of any of the model

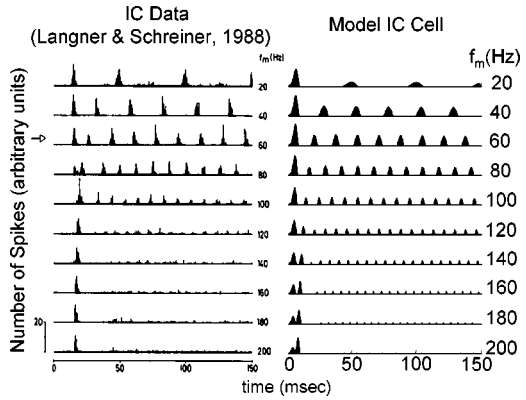


FIG. 11. Comparison of actual (left) and model (right) IC cell PST histograms for a range of stimulus f_m. Model parameters were the same as the cell illustrated in Fig. 10 (with SPL=20 dB). Physiological PSTs from Langner and Schreiner (1988, their Fig. 2, unit IC115; with permission).

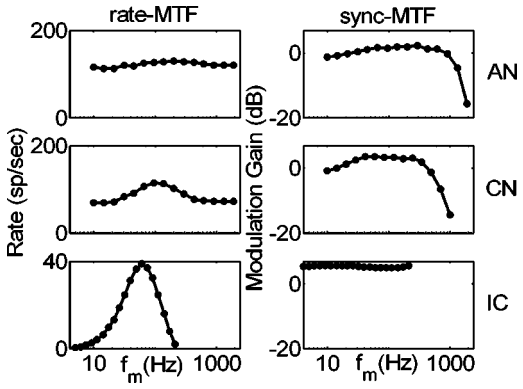


FIG. 12. Summary of the effects of SFIE on rate- and sync-MTFs for two model layers of convergent excitation and inhibition. Model and stimulus parameters match those of the previous three figures (with SPL=24 dB and $m=1$).

quantifications presented up to this point (all rate and synchrony calculations were made in the steady-state portion of the response). In the model, the onset response is caused by the relative sluggishness of the inhibitory input with respect to excitation. These model cells would be classified as onset units in terms of their pure-tone responses.

Other subtle similarities are present that are not reflected in rate- or sync-MTFs. The slow build-up at high f_m after onset that is observed in IC physiology (e.g., at $f_m = 120-200$ Hz) is captured by the model. This phenomenon is a result of the slow inhibitory input in the model cell, which determines how quickly the cell response reaches steady-state. Also, two peaks in the onset response at certain modulation frequencies are present in both data and model (caused by fixing the starting phase of the signal modulation). In general, the qualitative similarities in the PST histograms suggest that the simple mechanism proposed here can predict more than gross rate and synchrony changes with modulation frequency.

D. Summary

The apparent transition from a temporal AM code in the periphery to a rate-based scheme at the level of the IC is predicted by a simple neural encoding hypothesis. Specifically, two levels of convergent slow inhibition and fast excitation reproduce several established response characteristics of VCN and IC neurons, at least qualitatively. The most fundamental difference between the processing strategies at the two levels is the strength of the inhibitory input relative to the excitatory component. A summary of the responses of two-stage SFIE is shown in Fig. 12.

IV. DISCUSSION

Several important features of neural AM encoding in the auditory brainstem and midbrain were predicted by the cascaded convergence of same-frequency inhibitory and excitatory inputs with the physiologically realistic time courses described in this study. A modified version of the Zhang *et al.* (2001) AN model that is consistent with physiological recordings in the cat provided inputs to this processing mechanism. At the level of the AN, changes in envelope synchrony and rate with modulation frequency and depth were accurately captured by the new AN model. A single layer of convergent inhibition and excitation yielded model responses with increased synchrony relative to AN firing patterns, similar to trends seen in VCN bushy cell physiology. In a second level of convergent SFIE, model cell responses were rate-tuned to stimulus modulation frequency by allowing the inhibitory inputs to be stronger than the corresponding excitatory inputs. Rate tuning to f_m has been reported in both gerbil and cat IC (i.e., Krishna and Semple, 2000; Langner and Schreiner, 1988).

A. Limitations of the AN model

A significant shortcoming of the AN model used in the present study becomes apparent in its response to high-SPL, high-f_m AM stimuli, which elicit unrealistically high synchrony in the model. This results in strongly band-pass model AN sync-MTFs at high sound levels. Evaluation of the model in this study was limited to low and moderate SPLs, to avoid using physiologically inconsistent AN responses as inputs to higher processing centers. A possible source for this discrepancy is the speed and strength with which the onset properties of the AN model act on a cycle-by-cycle basis. At high levels and low f_m, there is a clear onset component with each AM stimulus cycle, but synchrony is reduced compared to lower-SPL stimuli because the remainder of the cycle histogram (after the onset) is saturated. In contrast, high-SPL, high-f_m SAM stimuli cause a highly synchronized response that is completely dominated by the strong onset, with no saturated portion of the cycle histogram. It should be noted that at very high f_m (i.e., $f_m > 1000$ Hz), envelope synchrony still rolls off at high SPLs; it is for stimulus f_m between 20 and 500 Hz that the time course of onset adaptation causes the most significant increase in synchrony (that has not been observed physiologically). A clear direction of future work is to understand and possibly modify the role that the synapse model onset components play in shaping the high-SPL AM cycle histograms.

This inconsistency between the model and data at high levels is tempered by the fact that the AN model fibers always have a pure-tone rate threshold of approximately 0 dB SPL, while real AN fibers generally have thresholds that vary

between fibers. Much of the AN physiology has been collected at levels that are on the steepest part of the rate-level function, usually 10–15 dB above rate threshold. The AN model can therefore be considered to provide physiologically realistic AM responses at levels below pure-tone rate saturation.

B. Model VCN cells: Alternative mechanisms

Synchronous AM response features in the VCN are diverse and correlated with PST classification type (Rhode and Greenberg, 1994). Part of the diversity is a result of the fact that VCN neurons receive different numbers of inputs that act with various strengths and latencies. Despite the additional complications, some broad observations are clear: (1) VCN sync-MTFs can have low-pass or band-pass shapes, (2) sync-MTFs have lower corner frequencies than high-CF AN fibers, and (3) rate-MTFs remain relatively flat [with the exception of onset choppers (Rhode and Greenberg, 1994)]. We have shown that a simple two-input model bushy cell that receives membrane and synapse-modified inhibition and excitation with the same CF can exhibit low-pass or weakly band-pass sync-MTFs with corner frequencies lower than the model AN fiber sync-MTFs, depending on the relative strengths of excitation and inhibition. Also, the firing rate of the model VCN neurons is nearly independent of modulation frequency when the inhibitory strength is less than 50%–60% of the excitatory strength. Allowing the inhibition to become stronger than this threshold results in tuned rate-MTFs. Interestingly, Joris and Yin (1998) show that many of the globular bushy cells in their population exhibit systematic variation in average rate with f_m (their Fig. 13C).

The observation that model cells could be rate-tuned in the first layer of convergent SFIE leads to an important question: is the cascade of two postsynaptic model cells necessary to explain the transition from a synchrony code to a rate code? While the two-layer structure is not strictly required to understand rate tuning, it is useful as a tool to understand the flow of information along the auditory pathway. Our physiologically based approach incorporates empirical data into the details of the model at each level of processing. By using data obtained from bushy cells in the VCN, we have described an alternative envelope information pathway that does not require populations of VCN choppers to understand the synchrony-to-rate transformation. However, interpretation of the current model structure is complicated by the fact that bushy cells in the VCN do not send axons directly to the IC; they project to the superior olivary complex (e.g., Oliver and Huerta, 1992). Neurons in the lateral superior olive (LSO) have low-pass sync-MTFs, similar to those of AN fibers and VCN bushy cells (Joris and Yin, 1998). The stereotypical rate-MTF in the LSO is also low-pass in shape, which is in contrast to the flat (all-pass) rate-MTFs measured in LSO afferents (Joris and Yin, 1998). However, the corner frequencies of monaural LSO rate-MTFs are significantly higher than those found in the IC; this suggests that the transformation that occurs at the level of the LSO is not crucial to generating the details of rate tuning in the IC. Physiological data in response to monaural AM tones in the medial superior olive (MSO) of cat and gerbil are not available.

Is the SFIE convergence mechanism for generating increased synchrony and band-pass sync-MTFs a unique solution? No; it is only a possible explanation for the physiological observations. While our model is physiologically plausible, it is also phenomenological in nature. An alternative hypothesis with significant physical support is that a coincidence-detection mechanism could also enhance envelope synchrony. A recent modeling study has suggested that VCN neurons could act as cross-frequency coincidence detectors on a shorter relative time scale (Heinz *et al.*, 2001a; Carney *et al.* 2002), and work continues to investigate the possibility that slower fluctuations could also be modified by such a mechanism in a manner consistent with physiology.

C. Model IC cells

The SFIE model described in this study is attractive for several reasons. Its primary appeal lies in its simplicity: a population of model IC cells, each receiving only two inputs, can span a wide range of rate-based BMFs. Individual cell rate tuning is determined primarily by differences in synaptic dynamics between excitatory and inhibitory inputs, and rate-MTFs can be further adjusted by varying the inhibitory delay. For the basic model with a single excitatory input and a single inhibitory input, overall rates are lower in cells with low BMFs, and peak rates drop again at high BMFs; this trend could be adjusted simply by changing the number of inputs to each cell. In other words, there is no inherent relationship between maximum rate and BMF. The high-synchrony and low-pass or all-pass sync-MTF features of the model cells are also consistent with physiological measurements in units that respond strongly at the onset of pure-tone stimulation.

It is worthwhile to consider some details of the shapes of simulated IC cell rate-MTFs, and the model features that underlie them. One such detail is the sharp roll-off on the high-frequency side of the rate-MTF, and the more gradual rate increase with f_m on the low-frequency end. Both stereotyped traits of the model cells are attributable to the strength and speed of inhibition relative to excitation. For f_m well above BMF, both inhibition and excitation tend towards constant (tonic) values because membrane and synaptic time constants limit the extent to which envelope locking can occur. Since inhibition is stronger in this level of the model, zero firing is observed on the output for the high-f_m stimulus condition. For very slow amplitude fluctuations, the overall rate is low because both inhibitory and excitatory components can accurately follow the modulations of the VCN cell response, resulting in a high degree of overlap between the two inputs. As f_m is increased, the temporal interactions gradually change until an optimal frequency causes phase-locked excitatory peaks to line up in time with inhibitory valleys (at BMF).

A more quantitative way to compare rate-MTF shapes between model cells is to measure their Q value (Q = BMF/bandwidth at $\frac{1}{2}$ of the cell's peak rate). For the parameter spaces investigated, model cells were all observed to have rate-tuning Q values $\leqslant 1.2$ at low to medium SPLs, with most measurements near 1. Although the metric was not reported in Krishna and Semple (2000), approximate calcu-

lations using their published rate-MTFs suggest similar physiological values. In the present model, effectively broader tuning can be caused by allowing inhibitory synaptic filtering properties to be significantly different than those of excitatory inputs ($\tau_{\mathrm{IC,inh}} \gg \tau_{\mathrm{IC,exc}}$). If we assume that the excitatory time constant is relatively fast and consistent across cells, this mechanism would predict that cells with lower BMFs would also have broader tuning.

Krishna and Semple (2000) found that a large proportion ($\sim$45%) of IC neurons exhibited a *suppressive* region in the rate-MTF. That is, the firing rate was depressed over some narrow range of f_m relative to rates at higher and lower stimulus modulation frequencies. The model presented here does not directly predict such a trend, but a simple scheme can be imagined in which suppressive regions can be created. Rate-tuned collaterals or ascending inputs [possibly from the VNLL (Batra, 2004)] converging on a unit in the IC could provide inhibitory inputs to a postsynaptic cell. Krishna and Semple (2000) point out that for such a scheme to work, the net inhibitory effect would simply be proportional to the mean spike rate of the inputs. Another population of cells in the IC is characterized by a low-pass to band-pass sync-MTF shape transition with increasing SPL that is similar to some onset and chopper responders in the VCN (i.e., Krishna and Semple, 2000). The model presented here does not predict such a transformation; all model IC cell responses are highly synchronized to the modulation period, even at high SPLs.

Sinex *et al.* (2002) described recordings in the chinchilla IC in response to mistuned complex tones that have beating periodicities which are not present in the stimulus envelope. They put forth a simplified processing model that could account for some of these periodicities by combining information across different CFs. The model presented here would not predict such interactions, because model inputs are all from a single CF. Given the complex connections (both ascending and descending) at the level of the IC, it is reasonable to assume that information is integrated across different frequencies. We have restricted excitatory and inhibitory inputs on a postsynaptic cell to have the same CF in an effort to keep the model as simple as possible, but this approach could be extended to include model IC cells that receive inhibition and excitation with different CFs.

D. Future directions

This study was a necessary first step in an effort to quantitatively relate physiological responses to human psychophysical performance in AM detection tasks. Recent psychophysical models of AM perception assume that a population of hard-wired filters tuned to f_m provides information about a signal's temporal envelope to higher processing centers (e.g., Dau *et al.*, 1997; Ewert *et al.*, 2002). This ''modulation filterbank'' model structure is fundamentally different than that of previous models which assume that the net effect of central processing is to low-pass filter, or smooth, the envelope (e.g., Viemeister, 1979). The model IC cells presented here can be thought of as a physiological implementation of a modulation filter, and work continues to understand the relationship between model responses and perceptual modulation tuning.

Long-duration inhibitory influences in the model IC cells may also provide a physiologically realistic mechanism for understanding psychophysical forward-masking phenomena. Offset suppression measured in AN fibers is too short to account for the time course of forward masking in humans (Smith, 1977; Harris and Dallos, 1979; Plack and Oxenham, 1998). Preliminary work suggests that the same f_m-tuned model IC cells presented here may effectively suppress responses to a probe signal that is presented after a masking stimulus for significantly longer masker-probe intervals. A physiological framework that integrates our understanding of modulation detection and forward masking is appealing.

An important test of the AN model will be to implement the newly derived sharp human filter bandwidths (Shera *et al.*, 2002; Oxenham and Shera, 2003), and observe the effects on AM response properties. Sharper filters will result in lower envelope synchrony for a given carrier frequency, as AM sidebands will be more strongly attenuated. In this respect, some of the modulation encoding ability seems to be lost with narrower-band filters. From a different perspective, a sharp peripheral filter will result in significant modulation of wideband stimuli, emphasizing fluctuations in the filter's frequency pass band. Taken together, these observations illustrate the importance of using many stimuli and response quantifications when evaluating the performance of any model (especially one that simulates a highly nonlinear system such as the auditory system).

The main focus of future work on model VCN neurons will be to see if alternative processing mechanisms (i.e., coincidence detection) could also increase synchrony and underlie band-pass sync-MTFs. Coincidence detection can be approximated computationally using the simplified approach described in this work by taking the product of multiple input instantaneous firing rates as the response of a postsynaptic cell. Integrate-and-fire model neurons based on dynamic channel properties would provide a higher-order approximation of information integration at the level of VCN cells, which receive multiple excitatory AN inputs, as well as slower inhibitory inputs. Physiological observations that estimate the frequency range of a cell's inputs, as well as the number of inputs, would be very useful in building a realistic VCN model cell.

Our model of processing in the IC should be extended to include cells that receive and integrate excitatory and inhibitory inputs with different CFs. This may allow the model described here to predict responses in the IC to mistuned complex tones [as in Sinex *et al.* (2002)] by combining resolved spectral information from different frequency bands. We should also consider whether long-lasting inhibition can be modeled simply with a single long IPSP, as opposed to a train of shorter and weaker IPSPs. This could be accomplished by directly comparing responses to the two IPSP configurations. Since the computational cost of convolution is quite high, carrying out this experiment should also motivate the development of a faster, frequency-domain algorithm that performs the same function of convolution in the time domain. Another possible direction for studying the fundamental mechanism of an additive combination of envelope-locked excitation and inhibition with different time courses

and relative strengths would be to simplify the structure of the model presented here (as in Krukowski and Miller, 2001); a reduced parameter space might allow for more systematic variation of the key components of the second-order (IC) model cells.

The techniques we have used to simulate the effects of membrane and synaptic filtering provide a first approximation of how information in the central auditory system may be modified as it ascends. While the computational methods described here are physiologically based, they are by no means exhaustive. Perhaps most importantly, we assume that both inhibitory and excitatory contributions to a postsynaptic model cell are simply proportional to the instantaneous firing rate of a lower-level input. One way to refine the model would be to take into account the effects of membrane and synapse properties on a spike-by-spike basis. Implementing a conductance-based integrate-and-fire model neuron would allow us to include other realistic properties of cells in the auditory brainstem and midbrain. For example, some neurons may act as coincidence detectors; low-threshold potassium channels have been implicated as candidate mechanisms that could allow for such interactions between sequential inputs (Svirskis and Rinzel, 2003). Due to computational demands, it was not possible to include these effects in the present version of the model, but this is another clear direction for future work.

ACKNOWLEDGMENTS

We thank Xuedong Zhang for his help in implementing modifications to the AN model. Ray Meddis and an anonymous reviewer provided useful comments on an earlier version of this paper. This research was supported by NIH-NIDCD R01-01641 and the Jerome R. and Arlene L. Gerber Fund.

Batra, R. (**2004**). ''Responses to amplitude-modulated tones of neurons in the ventral nucleus of the lateral lemniscus of the unanesthetized rabbit,'' Assoc. Res. Otolaryngol. Abstr. **27**, 914.

Carney, L. H. (**1993**). ''A model for the responses of low-frequency auditory-nerve fibers in cat,'' J. Acoust. Soc. Am. **93**, 401–417.

Carney, L. H., and Yin, T. C. T. (**1989**). ''Responses of low-frequency cells in the inferior colliculus to interaural time differences of clicks: Excitatory and inhibitory components,'' J. Neurophysiol. **62**, 144–161.

Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S. (**2002**). ''Auditory phase opponency: A temporal model for masked detection at low frequencies,'' Acust. Acta Acust. **88**, 334–347.

Caspary, D. M., Palombi, P. S., and Hughes, L. F. (**2002**). ''GABAergic inputs shape responses to amplitude modulated stimuli in the inferior colliculus,'' Hear. Res. **168**, 163–173.

Caspary, D. M., Backoff, P. M., Finlayson, P. G., and Palombi, P. S. (**1994**). ''Inhibitory inputs modulate discharge rate within frequency receptive fields of anteroventral cochlear nucleus neurons,'' J. Neurophysiol. **72**, 2124–2132.

Cooper, N. P., Robertson, D., and Yates, G. K. (**1993**). ''Cochlear nerve fiber responses to amplitude-modulated stimuli: Variations with spontaneous rate and other response characteristics,'' J. Neurophysiol. **70**, 370–386.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). ''Modeling auditory processing of amplitude modulation. I. Modulation detection and masking with narrow-band carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.

Elhilali, M., Fritz, J. B., Klein, D. J., Simon, J. Z., and Shamma, S. A. (**2004**). ''Dynamics of precise spike timing in primary auditory cortex,'' J. Neurosci. **24**, 1159–1172.

Ewert, S. D., Verhey, J. L., and Dau, T. (**2002**). ''Spectro-temporal processing in the envelope-frequency domain,'' J. Acoust. Soc. Am. **112**, 2921–2931.

Faure, P. A., Fremouw, T., Casseday, J. H., and Covey, E. (**2003**). ''Temporal masking reveals properties of sound-evoked inhibition in duration-tuned neurons of the inferior colliculus,'' J. Neurosci. **24**, 3052–3065.

Frisina, R. D., Smith, R. L., and Chamberlain, S. C. (**1990**). ''Encoding of amplitude modulation in the gerbil cochlear nucleus. I. A hierarchy of enhancement,'' Hear. Res. **44**, 99–122.

Goldberg, J. M., and Brown, P. B. (**1969**). ''Responses of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization,'' J. Neurophysiol. **22**, 613–636.

Grothe, B. (**1994**). ''Interaction of excitation and inhibition in processing of pure tone and amplitude-modulated stimuli in the medial superior olive of the mustached bat,'' J. Neurophysiol. **71**, 706–721.

Harris, D. M., and Dallos, P. (**1979**). ''Forward masking of auditory nerve fiber responses,'' J. Neurophysiol. **42**, 1083–1107.

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001a**). ''Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection,'' J. Acoust. Soc. Am. **110**, 2065–2084.

Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (**2001b**). ''Auditory-nerve model for predicting performance limits of normal and impaired listeners,'' JARO **2**, 91–96.

Hewitt, M. J., and Meddis, R. (**1994**). ''A computer model of amplitude-modulation sensitivity of single units in the inferior colliculus,'' J. Acoust. Soc. Am. **95**, 2145–2159.

Joris, P. X., and Yin, T. C. T. (**1992**). ''Responses to amplitude-modulated tones in the auditory nerve of the cat,'' J. Acoust. Soc. Am. **91**, 215–232.

Joris, P. X., and Yin, T. C. T. (**1998**). ''Envelope coding in the lateral superior olive. III. Comparison with afferent pathways,'' J. Neurophysiol. **79**, 253–269.

Krishna, B. S., and Semple, M. N. (**2000**). ''Auditory temporal processing: Responses to sinusoidally amplitude-modulated tones in the inferior colliculus,'' J. Neurophysiol. **84**, 255–273.

Krukowski, A. E., and Miller, K. D. (**2001**). ''Thalamocortical NMDA conductances and intracortical inhibition can explain cortical temporal tuning,'' Nat. Neurosci. **4**, 424–430.

Langner, G. (**1981**). ''Neuronal mechanisms for pitch analysis in the time domain,'' Exp. Brain Res. **44**, 450–454.

Langner, G., and Schreiner, C. E. (**1988**). ''Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms,'' J. Neurophysiol. **60**, 1799–1822.

Liberman, M. C. (**1978**). ''Auditory-nerve responses from cats raised in a low-noise chamber,'' J. Acoust. Soc. Am. **63**, 442–455.

Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). ''Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers,'' J. Acoust. Soc. Am. **101**, 3602–3616.

Oertel, D. (**1983**). ''Synaptic responses and electrical properties of cells in brain slices of the mouse anteroventral cochlear nucleus,'' J. Neurosci. **3**, 2043–2053.

Oliver, D. L., and Huerta, M. F. (**1992**). ''Inferior and superior colliculi,'' in *The Mammalian Auditory Pathway: Neuroanatomy*, edited by D. R. Webster, A. N. Popper, and R. R. Fay (Springer Verlag, New York), pp. 168–221.

Oxenham, A. J., and Shera, C. A. (**2003**). ''Estimates of human cochlear tuning at low levels using forward and simultaneous masking,'' JARO **4**, 541–554.

Palombi, P. S., and Caspary, D. M. (**1996**). ''GABA inputs control discharge rate primarily within frequency receptive fields of inferior colliculus neurons,'' J. Neurophysiol. **75**, 2211–2219.

Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**, 1598–1608.

Rhode, W. S., and Greenberg, S. (**1994**). ''Encoding of amplitude modulation in the cochlear nucleus of the cat,'' J. Neurophysiol. **71**, 1797–1825.

Ryugo, D. K., and Parks, T. N. (**2003**). ''Primary innervation of the avian and mammalian cochlear nucleus,'' Brain Res. Bull. **60**, 435–456.

Schneiderman, A., Oliver, D. L., and Henkel, C. K. (**1988**). ''Connections of the dorsal nucleus of the lateral lemniscus: an inhibitory parallel pathway in the ascending auditory system?'' J. Comp. Neurol. **276**, 188–208.

Schofield, B. R., and Cant, N. B. (**1996**). ''Projections from the ventral cochlear nucleus to the inferior colliculus and the contralateral cochlear nucleus in guinea pigs,'' Hear. Res. **102**, 1–14.

Shera, C. A., Guinan, Jr., J. J., and Oxenham, A. J. (**2002**). ''Revised esti-

mates of human cochlear tuning from otoacoustic and behavioral measurements,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Sinex, D. G., Sabes, J. H., and Li, H. (**2002**). ''Responses of inferior colliculus neurons to harmonic and mistuned complex tones,'' Hear. Res. **168**, 150–162.
Smith, R. L. (**1977**). ''Short-term adaptation in single auditory nerve fibers: Some post-stimulatory effects,'' J. Neurophysiol. **40**, 1098–1111.
Smith, R. L., and Brachman, M. L. (**1980**). ''Response modulation of auditory-nerve fibers by AM stimuli: Effects of average intensity,'' Hear. Res. **2**, 123–133.
Smith, R. L., and Zwislocki, J. J. (**1975**). ''Short-term adaptation and incremental responses of single auditory-nerve fibers,'' Biol. Cybern. **17**, 169–182.
Svirskis, G., and Rinzel, J. (**2003**). ''Influence of subthreshold nonlinearities on signal-to-noise ratio and timing precision for small signals in neurons: Minimal model analysis,'' Network Comput. Neural Syst. **14**, 137–150.
Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.
Westerman, L. A., and Smith, R. L. (**1988**). ''A diffusion model of the transient response of the cochlear inner hair cell synapse,'' J. Acoust. Soc. Am. **83**, 2266–2276.
Wickesberg, R. E., and Oertel, D. (**1988**). ''Tonotopic projection from the dorsal to the anteroventral cochlear nucleus of mice,'' J. Comp. Neurol. **268**, 389–399.
Wu, S. H., Ma, C. L., Sivaramakrishnan, S., and Oliver, D. L. (**2002**). ''Synaptic modification in neurons of the central nucleus of the inferior colliculus,'' Hear. Res. **168**, 43–54.
Yin, T. C. T. (**2002**). ''Neural mechanisms of encoding binaural localization cues in the auditory brainstem,'' in *Integrative Functions in the Mammalian Auditory Pathway*, edited by D. Oertel, A. N. Popper, and R. R. Fay (Springer Verlag, New York), pp. 99–159.
Zhang, X. (**2004**). Personal communication.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). ''A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression,'' J. Acoust. Soc. Am. **109**, 648–670.

Simultaneous measurement of middle-ear input impedance and forward/reverse transmission in cat

Susan E. Voss[a)]
Picker Engineering Program, Smith College, 51 College Lane, Northampton, Massachusetts 01063 and Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114

Christopher A. Shera
Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street Boston, Massachusetts 02114 and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 29 January 2004; revised 2 July 2004; accepted 6 July 2004)

Reported here is a technique for measuring forward and reverse middle-ear transmission that exploits distortion-product otoacoustic emissions (DPOAEs) to drive the middle ear ''in reverse'' without opening the inner ear. The technique allows measurement of DPOAEs, middle-ear input impedance, and forward and reverse middle-ear transfer functions in the same animal. Intermodulation distortion in the cochlea generates a DPOAE at frequency $2f_1-f_2$ measurable in both ear-canal pressure and the velocity of the stapes. The forward transfer function is computed from stapes velocities and corresponding ear-canal pressures measured at the two primary frequencies; the reverse transfer function is computed from velocity and pressure measurements at the DPOAE frequency. Middle-ear input impedance is computed from ear-canal pressure measurements and the measured Thévenin equivalent of the sound-delivery system. The technique was applied to measure middle-ear characteristics in anesthetized cats with widely opened middle-ear cavities (0.2–10 kHz). Stapes velocity was measured at the incudo-stapedial joint. Results on five animals are reported and compared with a published middle-ear model. The measured forward transfer functions and input impedances generally agree with previous measurements, and all measurements agree qualitatively with model predictions. The reverse transfer function is shown to depend on the acoustic load in the ear canal, and the measurements are used to compute the round-trip middle-ear gain and delay. Finally, the measurements are used to estimate the parameters of a two-port transfer-matrix description of the cat middle ear. [DOI: 10.1121/1.1785832]

PACS numbers: 43.64.Bt, 43.64.Jb, 43.64.Ha [WPS] Pages: 2187–2198

I. INTRODUCTION

The middle ear's primary function of coupling acoustic signals from the ear canal to the cochlea has been well recognized for over 100 years (reviewed in Merchant and Rosowski, 2003). With the discovery of otoacoustic emissions (OAEs) the recognized role of the middle ear expanded: not only does the middle ear couple external sounds to the cochlea, but it also couples sounds generated within the cochlea back to the ear canal. Describing the transmission characteristics of the middle ear in both the forward and the reverse directions is critical for many purposes, including understanding the middle-ear's effects on OAEs and understanding middle-ear function in both normal and diseased ears in order to develop better therapeutic and diagnostic approaches for pathological middle ears.

Recent work has focused on middle-ear transmission in both the forward and the reverse directions (e.g., Puria and Rosowski, 1996; Magnan *et al.*, 1997, 1999; Avan *et al.*, 2000; Puria, 2003, 2004). However, quantitative descriptions (measurement or model) of both forward- and reverse-transfer functions of the middle ear are not sufficient to completely describe the middle ear's function, as these transfer functions depend on terminating impedances (i.e., cochlear impedance and ear-canal impedance directed outward from the ear canal). Instead of describing the middle ear's function through specific transfer functions, the middle ear can be regarded as a black box whose input–output relations can be described without reference to its specific components or to its termination (Shera and Zweig, 1991, 1992b; Puria, 2003, 2004). Mathematically, these input–output relations can be described using a two-port model characterized by a 2×2 transfer matrix [sometimes called an ''$ABCD$ matrix'' after its four matrix elements, which are traditionally denoted $\left(\begin{smallmatrix}A & B\\ C & D\end{smallmatrix}\right)$]. If the load impedances are known, measurement of four independent, complex functions of frequency completely determines the values of the four matrix elements. In practice, at least one of these four measurements must be obtained while driving the middle ear ''in reverse'' (i.e., from within the inner ear).

Despite its considerable importance both for testing models of middle-ear mechanics and for understanding the effects of middle-ear transmission on OAEs, a complete two-port characterization of the input–output relations of the

a)Electronic mail: svoss@email.smith.edu

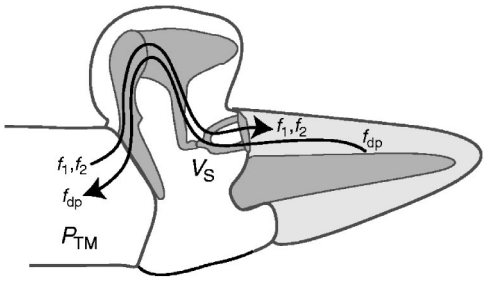


FIG. 1. Schematic diagram illustrating the idea behind the method. Primary stimulus tones (at frequencies f_1 and f_2) drive the middle ear in the forward direction while the distortion-product (at frequency f_{dp}) drives the middle ear ''in reverse.'' Forward and reverse middle-ear transfer functions can be computed from measurements of ear-canal pressure (P_{TM}) and stapes velocity (V_S).

middle ear has so far only been obtained in human temporal bones (Puria, 2003). Much of the technical difficulty in obtaining this characterization comes from the need to measure the cochlear response (e.g., stapes velocity or intracochlear pressure) while driving the middle ear in reverse (e.g., with a hydrophone in the vestibule). For example, inserting transducers into the inner ear can introduce tiny air bubbles into the scalae that, unless carefully controlled for, significantly alter the measured responses (Puria *et al.*, 1997). As a step towards obtaining a two-port characterization of a living middle ear, we have developed a method for measuring forward and reverse middle-ear transmission that exploits distortion-product otoacoustic emissions (DPOAEs) to drive the middle ear in reverse without opening the inner ear. The technique allows for the measurement of DPOAEs, middle-ear input impedance, and forward and reverse middle-ear stapes-velocity transfer functions in the same animal. The ability to measure all of these quantities in the same preparation allows for experimentally based estimates of the transfer matrix that are not compromised by interanimal variations in the measurements. A preliminary account of this work has been presented elsewhere (Voss and Shera, 2002).

II. METHODS

A. Overview of the method

The idea behind the measurement is illustrated in Fig. 1. When the cochlea is stimulated by primary tones at frequencies f_1 and f_2 (with $f_2 > f_1$), intermodulation distortion in the cochlea generates energy at the combination-tone frequency $f_{dp} = 2f_1 - f_2$ that propagates back along the cochlear partition. When it reaches the base of the cochlea, some of this energy vibrates the stapes and is subsequently transmitted through the middle ear to the ear canal, where it can be measured as a DPOAE. Energy at the three frequencies f_1, f_2, and f_{dp} is therefore measurable in both the stapes velocity and the ear-canal pressure; in effect, the primary tones drive the middle ear in the forward direction while the distortion-product drives the middle ear ''in reverse.'' Simultaneous measurements of forward transmission (at f_1 and f_2) and reverse transmission (at f_{dp}) can therefore be performed by extracting the appropriate frequency components from the measured ear-canal pressure and stapes velocity using Fourier analysis. Note that for the method to yield reliable measurements of reverse transmission it is crucial that the dominant sources of energy at f_{dp} measured in the ear canal lie on the cochlear side of the middle ear (i.e., within the cochlea or annular ligament).

B. Definitions of transfer functions

The forward stapes-velocity transfer function, $\overrightarrow{T}_S(f)$, is a measure of the transmission from the ear canal through the middle ear to the stapes. $\overrightarrow{T}_S$ is defined as the ratio of stapes velocity (V_S) to the ear-canal pressure at the tympanic membrane (P_{TM}) when the middle ear is driven from the ear canal (i.e., at frequencies f_1 and f_2):

$$\overrightarrow{T}_S(f) \equiv \frac{V_S(f)}{P_{TM}(f)}, \quad f \in \{f_1, f_2\}. \tag{1}$$

When the middle ear is driven in the forward direction, we adopt the convention that a positive displacement moves the stapes into the cochlea.

The reverse stapes-velocity transfer function, $\overleftarrow{T}_S(f)$, measures the transmission from the stapes back through the middle ear to the ear canal. $\overleftarrow{T}_S$ is defined as the ratio of P_{TM} to V_S when the middle ear is driven from within the cochlea (i.e., at the frequency f_{dp}):

$$\overleftarrow{T}_S(f_{dp}) \equiv \frac{P_{TM}(f_{dp})}{V_S(f_{dp})}. \tag{2}$$

In this case, a positive displacement moves the stapes out of the cochlea. The left- and rightward pointing arrows atop the transfer functions $\overrightarrow{T}_S$ and $\overleftarrow{T}_S$ indicate both the forward and reverse directions and the assumed polarity of positive displacements.

C. Animal preparation

Measurements were made on one ear in each of five anesthetized cats. Treatment of animal subjects accorded with protocols approved by the animal care committee at the Massachusetts Eye and Ear Infirmary.

Young cats weighing between 2.15 and 2.50 kg were anesthetized with intraperitoneal injections of Dial (75 mg/kg). Booster doses of Dial (10% of the initial dose) were given throughout the experiments, as indicated by either a withdrawal response to a toe pinch or an increase of 20% in heart rate. Cats were also connected to a saline drip in order to keep them hydrated for the duration of the experiment. All measurements were performed in a humidified sound-proof booth maintained at approximately 38 °C.

The pinna and most of the cartilaginous ear canal were removed to allow the sound source to be placed close to the tympanic membrane. The ventral and lateral walls of the bulla and most of the bony septum were removed so that the middle ear was opened widely. As illustrated in Fig. 2, access to the stapes was obtained by drilling the bone lateral to the superior-posterior quadrant of the tympanic membrane (Tonndorf and Tabor, 1962). As described by Tonndorf and Tabor (1962):

> ''[The hole's] location corresponds to McEwen's triangle in man: posterior to the rim of the eardrum, in-

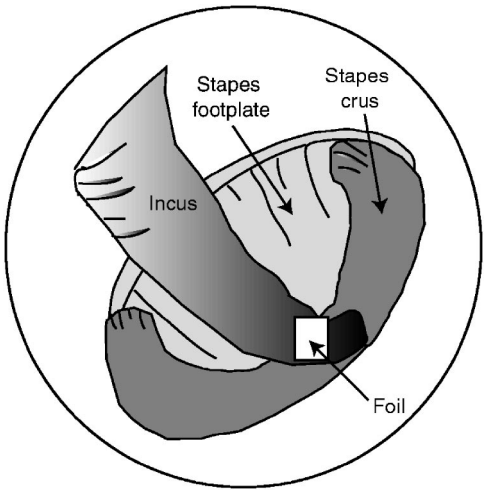


FIG. 2. Schematic of the view of the stapes after drilling the location in cat that corresponds to McEwen's triangle in man (adapted from Fig. 4 of Tonndorf and Tabor, 1962). The white rectangle labeled ''Foil'' indicates the location of the reflective tape used during velocity measurements to reflect the laser beam.

ferior to the inferior temporal line and anterior-superior to the mastoid ridge. This perforation opens directly over the incudo-stapedial joint, giving good access to the anterior two-thirds of the footplate and to the anterior crus. ... (Care must be taken not to cut into the underlying soft tissues as bleeding from the mastoid vessels is very annoying.)''

To determine the effects of drilling on cochlear sensitivity we measured ear-canal DPOAEs both before and after drilling McEwen's triangle in three of the five ears. Since differences between pre- and postdrilling emissions differed by less than 5 dB at most frequencies, we conclude that the drilling had little effect on the emissions generated by the cochlea. While changes in emissions before and after drilling have no effect on our measured transfer functions, it was important to determine that after drilling the cochlea was still generating a substantial and robust distortion product.

D. Stimulus generation and response measurement

Ear-canal pressures (P_{EC}) were generated and measured with calibrated transducers positioned within 2–3 mm of the tympanic membrane and controlled by a computer running LabVIEW (detailed in Shera and Guinan, 1999). The earphones were two $\frac{1}{4}$-in. Larson-Davis microphones used as sound sources; the microphone was an Etymotic ER-10C. The stimulus was either a broadband chirp (used to measure impedance) or two pure tones at primary frequencies f_1 and f_2 (used to measure forward and reverse transfer functions). To reduce the possibility of distortion in the earphone system, each of the two primary tones was produced by its own earphone. The primary-tone frequency ratio f_2/f_1 and level difference L_1-L_2 were chosen with the goal of maximizing the magnitude of the resulting distortion product (f_2/f_1 was fixed at the value 1.2 and L_1-L_2 was typically 5 to 10 dB). Response magnitudes and angles were obtained from the 4096-point discrete Fourier transform of the time-domain average of N responses ($64 \leqslant N \leqslant 1024$) sampled at 44.04 kHz.

E. Measurement of stapes velocity

After exposing the stapes via McEwen's triangle, we placed a small reflective foil on the long process of the incus, at the incudo-stapedial joint (see Fig. 2). The velocity of this foil was measured with a laser vibrometer (Polytex OFV-501). Guinan and Peake (1967) have shown that no significant slippage occurs between the incus and the stapes at the incudo-stapedial joint (i.e., the joint is effectively rigid in cat); this assumption is consistent with the more recent work of Decraemer and colleagues (Decraemer, 2004a,b). We therefore refer to the measured velocity as the stapes velocity, V_S. The Doppler-shifted reflected signal was detected and decoded by the vibrometer to produce an output voltage proportional to stapes velocity (detailed in Voss *et al.*, 2000). The vibrometer output voltage was amplified by a factor of 10.

The laser vibrometer system measures velocity in the direction of the laser beam. Here, the laser beam was focused on the incudo-stapedial joint, and the angle between the laser beam and the pistonlike stapes motion was judged visually to be close to zero. Any small nonzero angle would have negligible effects on the results reported here. For example, if the angle had been 20°, the measurements reported here would systematically underestimate the true velocity by only 0.5 dB (i.e., $20\log_{10}[\cos(20\pi/180)]$).

We assume that the mass of the foil (0.05 mg) had a negligible effect on the motion of the ossicular system since the mass of the foil was substantially smaller than the mass of the stapes (0.530 ± 0.05 mg) and the incus (4.313 ± 0.328 mg) (Lynch, 1981). Additional support for this assumption comes from control measurements reported by Voss (1998, Figs. 1–4) where stapes velocity measurements are shown to be similar when measured using either a single foil (mass 0.05 mg) or with three foils (0.15 mg). Thus, these measurements are consistent with the assumption that the foil has little effect on the mechanics of the middle-ear system.

F. Noise floor for stapes-velocity measurements

We obtained estimates of the noise floor for the velocity measurements by (1) measuring the velocity of the skull in response to the ear-canal stimulus and/or (2) measuring the stapes velocity in the absence of the stimulus. Both methods produced similar estimates of the noise floor. One measured noise floor for the stapes-velocity magnitude is plotted as circles in the top-center plot of Fig. 4. Here, and in all experiments, the measured noise floor has been filtered with a median filter in order to smooth the noise floor across the measured frequency range. The effect of the median filter on the noise floor is illustrated by the line labeled ''filtered noise floor.'' The magnitude of the stapes velocity at the distortion-product frequency ($|V_S(f_{dp})|$) was substantially smaller than that at the primary frequencies ($|V_S(f_1)|$ and $|V_S(f_2)|$) and was often corrupted by noise. In our plots of the reverse transfer function (Fig. 5), we do not plot data points for which the noise floor was within 10 dB of the measured stapes velocity. The noise floor limited the measurement of

the reverse transfer function at the lower frequencies. Measurements of the forward transfer function were always more than 10 dB above the noise floor.

G. Source of intermodulation distortion

The method described here assumes that the dominant sources of intermodulation distortion originate on the cochlear side of the middle ear (i.e., within the cochlea and/or the annular ligament). Although it is well documented that the cochlea produces intermodulation distortion, there is little experimental evidence about the magnitude of middle-ear intermodulation distortion, although the documented linearity of middle-ear mechanics at the driving frequency suggests that it is small. Here we assume that middle-ear intermodulation distortion at $2f_1-f_2$ is small compared to that generated within the cochlea or annular ligament.

Postmortem measurements made in one of our preparations support this assumption for frequencies above 2 kHz. Specifically, at ear-canal sound pressure levels of 85 and 80 dB SPL for f_1 and f_2, respectively, ear-canal distortion products measured 2 h postmortem decreased by 15 to 30 dB between 0.7 and 7 kHz. The stapes velocity at f_{dp} decreased by the same factor for frequencies above 2 kHz; for many measurements above 2 kHz the decrease was equivalent to a reduction to the level of the noise floor. For measurements between 0.7 and 2 kHz, the results are not so clear cut. Stapes velocity at f_{dp} did not always decrease proportionally with the ear-canal pressure at f_{dp}. About half of these data points were within the noise floor. Further study is clearly needed to quantify the magnitude and sources of any intermodulation distortion originating within the middle ear.

H. Impedance measurements

The middle-ear input impedance, $Z_{EC}(f)$, was calculated from the ear-canal pressure (P_{EC}) and the Thévenin-equivalent impedance (Z_{Th}) and pressure (P_{Th}) of the transducer (e.g., Allen, 1986; Keefe *et al.*, 1992; Voss and Allen, 1994; Neely and Gorga, 1998). Pressure measurements made in six cylindrical tubes provided six complex equations for the two unknown Thévenin equivalents, Z_{Th} and P_{Th}. Acoustic estimates of the tube lengths were obtained by minimizing the error function in the overdetermined system of six equations (Allen, 1986; Keefe *et al.*, 1992), and the optimized lengths were used to compute Z_{Th} and P_{Th}. Results were checked by comparing measured and theoretical impedances in five additional tubes. At frequencies in the range 0.2–10 kHz the measured impedances were within 1 dB in magnitude and 0.01 cycles in angle of the corresponding theoretical impedances, except at maxima and minima in the impedance, where the estimates depend heavily on the precise length of the tube; at these frequencies the differences approached 5 dB in magnitude and 0.05 cycles in angle.[1]

I. Effect of higher-order modes

Our description of the measurements (i.e., impedance and forward and reverse transmission) assumes plane-wave propagation and that any higher-order spatial modes are negligible. Two theoretical sources for higher-order modes exist: (1) complex wave motion on the tympanic membrane that produces evanescent pressure modes in the ear canal near the tympanic membrane and (2) evanescent modes near the probe microphone that result from the stimulus pressure generation. Using both theoretical and measurement-based explanations, Lynch (1981) (pp. 146–148) argues that evanescent pressure modes generated by the tympanic membrane are insignificant for a probe tube placed within a few mm of the tympanic membrane at frequencies up to 22.4 kHz. Although evanescent modes produced by the earphone presumably contribute to the total pressure recorded by the microphone, our ability to accurately measure the impedance of test cavities using the Thévenin equivalents of our sound source suggests that the total contribution of these modes is small at frequencies below 10 kHz. Although our transducer assembly had a short probe-tube extension (<1 mm), the work of Huang *et al.* (2000a,b) suggests that a longer ''probe-tube extension'' is needed above about 2 kHz to eliminate evanescent-wave contributions from the pressure source. However, an important difference between our work and the work of Huang *et al.* (2000a,b) is that we work in domestic cat near the tympanic membrane where the cross-sectional dimensions of the ear canal are significantly smaller than the range of dimensions explored by Huang *et al.* (2000a,b).

J. Accounting for the residual ear-canal air space

The probe microphone that measured the ear-canal sound pressure was positioned about 3 mm from the tympanic membrane. We model the residual ear-canal air space between the probe microphone and the tympanic membrane as a rigid-walled cylindrical tube with uniform, plane-wave propagation occurring for frequencies below 10 kHz. In this case, the pressure and volume velocity at the probe microphone (P_{EC}, U_{EC}) are related to their counterparts at the tympanic membrane (P_{TM}, U_{TM}) by the transfer matrix

$$\begin{pmatrix} P_{EC} \\ U_{EC} \end{pmatrix} = \begin{pmatrix} \cosh(ikl) & Z_0 \sinh(ikl) \\ 1/Z_0 \sinh(ikl) & \cosh(ikl) \end{pmatrix} \begin{pmatrix} P_{TM} \\ U_{TM} \end{pmatrix}, \qquad (3)$$

where $Z_0=\rho c/A$ is the characteristic impedance of the tube, l is the length of the cylindrical air-filled tube, A is the cross-sectional area of the tube, $k=2\pi f/c$ is the wavenumber, ρ is the density of air, c is the velocity of sound in air, and f is the frequency (e.g., Møller, 1965; Rabinowitz, 1981; Lynch *et al.*, 1994; Huang *et al.*, 1997; Voss *et al.*, 2000). We assume that the effective area of the equivalent coupling tube is equal to the area of the source tube (radius $a=2.8$ mm). Although we did not measure the exact volume or dimensions of the air space in our cats, the equivalent volume was measured by Lynch *et al.* (1994) in six animals prepared in a similar manner to ours. Their volumetric measurements ranged from 62 to 75 mm^3; we approximate the volume in our preparations by their mean value of 68.5 mm^3 and the distance l from the probe microphone to the tympanic membrane by $l=68.5/\pi a^2=2.74$ mm, a value consistent with our estimate of 3 mm.

In the forward direction, application of Eq. (3) is consistent with the findings of Lynch *et al.* (1994), who found

only small differences between Z_{EC} and Z_{TM} at frequencies below about 6 kHz but larger differenences at higher frequencies, where the impedance ratio was approximately 3 dB in magnitude and up to 0.20 cycles in angle. In contrast, application of Eq. (3) results in ratios P_{TM}/P_{EC} that are nearly 1 at all frequencies up to 10 kHz; across all ears and in narrow bands the magnitude of the pressure ratio ranges from 0.85 to 1.20, but at most frequencies it is within 0.95 to 1.02 (i.e., a variation of less than 0.5 dB). The angle difference is generally within 0.02 cycles of zero.

In the reverse direction, the ear-canal air space also influences our measurements and model. In this case, the load on the ear-canal air space is the Thévenin impedance of the source, Z_{Th}. Thus, the total load at the tympanic membrane in the reverse direction is the ear-canal air space terminated by Z_{Th}. We define this load as $Z_{SRC}=P_{TM}/U_{TM}$. Consistent with the findings in the forward direction, in the reverse direction Z_{Th} and Z_{SRC} differ by substantial amounts while P_{EC} and P_{TM} are similar. In the reverse direction, the ratio Z_{SRC}/Z_{Th} has a magnitude near one and an angle near zero for frequencies below 1 kHz, but the magnitude varies by a factor of 1.5–3 at most frequencies between 1 and 10 kHz with corresponding angle variations between 0.1 and 0.25 cycles. The ratio P_{TM}/P_{EC} in the reverse direction is between 0.9 and 1.1 in magnitude and within 0.025 cycles of zero in angle.

All data presented here use P_{TM} and Z_{TM} obtained from measurements of P_{EC} and Z_{EC} using Eq. (3).

K. Stability of the preparation

We observed substantial variation in the stability of our preparations. In all cases, preparations were sensitive to drying-out effects (e.g., Voss *et al.*, 2000); although we humidified the warm chamber air, exposing the middle-ear system apparently caused the ossicular system to dry out and stiffen over time. This effect was manifest as an increase in the ear's impedance and a decrease in the low-frequency magnitude of the stapes velocity. In some, but not all cases, these effects were reversed by moistening the middle ear with saline. In several of the experiments, middle-ear bleeding ultimately led to problems with the stability of the preparation; in some of these cases the bleeding was controlled for several hours through periodic gentle suction and bone wax, but a large blood clot would ultimately form over the stapes, making further measurements impossible. Further confounding the problem was the fact that a single measurement session that swept a wide frequency range typically took several hours, since a large number of averages were needed to reduce the noise floor. Changes in middle-ear impedance were readily observable when the computer-controlled voltage to the earphone no longer produced the expected sound pressure levels. (The expected sound pressure was based upon an in-the-ear calibration performed periodically during the experiments; changes in ear-canal impedance resulted in changes in this calibration and thus in the ear-canal pressure produced.) We used deviations in the ear-canal sound levels L_1 and L_2 from their expected values as an indication that the impedance of the system had changed. In all measurements reported here, L_1 and L_2 are within 1.5 dB of their expected values. When changes in either the impedance magnitude or the low-frequency stapes-velocity magnitude occurred, measurements were aborted and attempts to return the system to its original response were made via moistening the middle ear and tympanic membrane with saline. If these attempts failed, measurements were ceased. Although the measurements shown in Fig. 4 at several stimulus levels were made over a 4-h period when the preparation was stable, the results presented here are generally the measurements made at the beginning of the measurement session on each ear.

III. RESULTS

We measured DPOAEs, middle-ear input impedance, and forward and reverse stapes-velocity transfer functions in each of five ears. Figure 3 illustrates the calculation of forward and reverse transfer functions. The transfer functions $\overrightarrow{T}_S(f)$ and $\overleftarrow{T}_S(f)$ (right panel) were computed from eardrum pressures P_{TM} (left panel) and stapes velocities V_S (center panel) measured simultaneously in response to primary tones at frequencies f_1 and f_2. Results are plotted for frequencies in the range 0.2–10 kHz. Measurements of the forward and reverse transmission are plotted for all frequencies that they were measured at within the 0.2–10 kHz range; in some cases the measurements only cover part of the frequency range.

A. Linearity of forward and reverse transmission

Consistent with previous findings (e.g., Wever and Lawrence, 1954; Guinan and Peake, 1967; Buunen and Vlaming, 1981; Voss *et al.*, 1996), forward middle-ear transmission appears to be linear. In our results, stapes velocities at f_1 and f_2 grow linearly with ear-canal sound pressure over the range of stimulus levels used (40–100 dB SPL). Figure 4 shows that reverse transmission also appears linear. The left-hand column plots the ear-canal DPOAEs $P_{TM}(f_{dp})$ measured at several different primary levels. The middle column plots the corresponding distortion-product components of the stapes velocity, $V_S(f_{dp})$. The right-hand column plots the reverse middle-ear transfer function $\overleftarrow{T}_S(f_{dp})$, defined as the ratio P_{TM}/V_S. Although both $P_{TM}(f_{dp})$ and $V_S(f_{dp})$ depend nonlinearly on the primary stimulus levels (L_1 and L_2), the reverse transfer function $\overleftarrow{T}_S(f)$ appears approximately independent of V_S, consistent with linear behavior. Note, however, that in the reverse direction the range of stapes velocities explored (15 dB) is considerably smaller than in the forward direction (60 dB). The range we were able to explore is bounded from below by the measurement noise floor and from above by the magnitude of the distortion generated within the cochlea.

B. Forward and reverse transfer functions

Figure 5 shows our measurements of $\overrightarrow{T}_S(f)$ and $\overleftarrow{T}_S(f)$ on five ears. The figure also compares our results with the measurements of Guinan and Peake (1967) and with the predictions of the middle-ear model of Puria and Allen (1998).

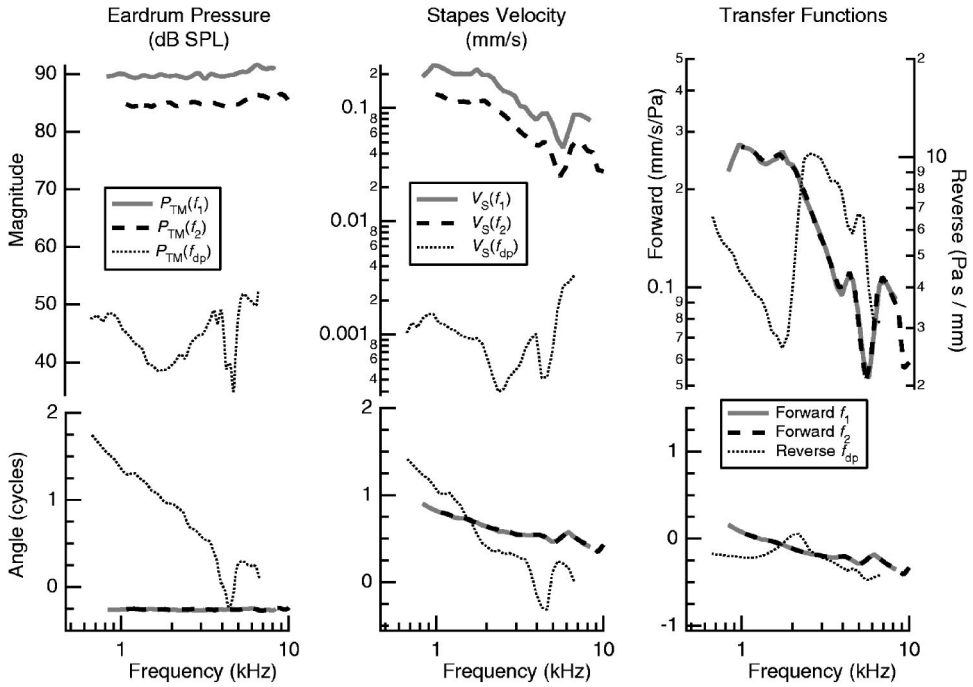


FIG. 3. Magnitudes (upper) and angles (lower) of simultaneous measurements of eardrum pressure P_{TM} (left panel) and stapes velocity V_S (center panel) at the frequencies f_1, f_2, and $f_{dp}=2f_1-f_2$. The right-hand panel shows the corresponding forward and reverse middle-ear transfer functions, $\overrightarrow{T}_S$ and $\overleftarrow{T}_S$. Note that the forward transfer functions computed from measurements at f_1 and f_2 superimpose. Measurements are from cat 58.

The five forward transfer functions $\overrightarrow{T}_S(f)$ share some features (Fig. 5, left). Transmission magnitudes increase with frequency at low frequencies, reach a maximum between 1 and 2 kHz, and generally decrease at higher frequencies. Additionally, all angles decrease as frequency increases. The data are similar to corresponding measurements of Guinan and Peake (1967), which were made using stroboscopic illumination, and to the model of Puria and Allen (1998).

The five reverse transfer functions $\overleftarrow{T}_S(f)$ also share some features (Fig. 5, right). Reverse transmission has a magnitude minimum between 1 and 3 kHz, followed by a local magnitude maximum within an octave of the minimum. The angles have a local maximum that corresponds with the rapid increase in the magnitude that occurs between the magnitude minimum and maximum. Similar features are also predicted by the model of Puria and Allen (1998) when the ear canal is terminated by Z_{SRC} (Sec. II J).

C. Impedance measurements

Figure 6 shows that the impedances at the eardrum $Z_{TM}(f)$ on all five ears share many features. These include a

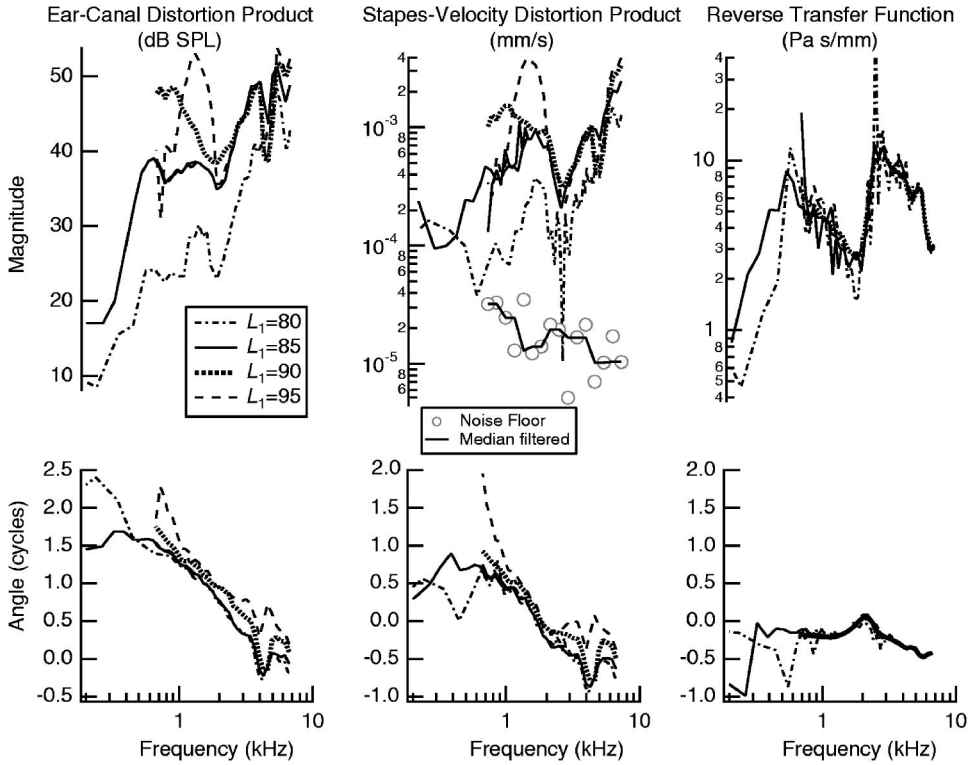


FIG. 4. Left: Measurements of the $2f_1-f_2$ component of the eardrum pressure, $P_{TM}(f_{dp})$. In all cases, $L_2=L_1-5$ dB. Center: The distortion-product component of the stapes velocity, $V_S(f_{dp})$. Right: The reverse transfer function, $\overleftarrow{T}_S(f)$. The data points marked ''noise floor'' on the magnitude plot of the stapes velocity result from a velocity measurement made with no stimulus. Primary levels are indicated by L_1 and L_2, corresponding to the tones at f_1 and f_2, respectively. All measurements are from cat 58, which was the only preparation stable enough to permit a series of measurements at several stimulus levels.

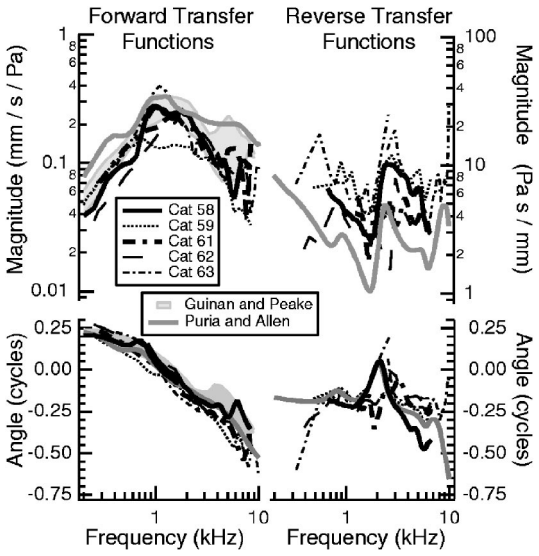


FIG. 5. Magnitudes (upper) and angles (lower) of the simultaneous measurements of $\overrightarrow{T}_S(f)$ (left) and $\overleftarrow{T}_S(f)$ (right) in five ears. The gray shaded region indicates the range of the forward transfer function measured by Guinan and Peake (1967) on 4 ears. Transfer functions computed using the model of Puria and Allen (1998) are shown for comparison.

stiffness-dominated behavior below 1 kHz (i.e., an impedance magnitude that decreases at about 6 dB per octave and an angle of roughly -0.25 cycles) and a mixed impedance at higher frequencies. Our impedance measurements are similar to those of Puria and Allen (1998) and Lynch *et al.* (1994), except that on average our low-frequency magnitudes are a few dB lower than the Lynch *et al.* (1994) measurements. These differences may reflect the relatively small animals employed in our study: According to Lynch *et al.* (1994, Fig. 17), low-frequency impedance magnitudes decrease with body mass. The Puria and Allen (1998) model and measurements are closer to our measurements: Both their model and data have low-frequency magnitudes that are similar to ours and both have angles that are not mass dominated but instead noodle about zero.

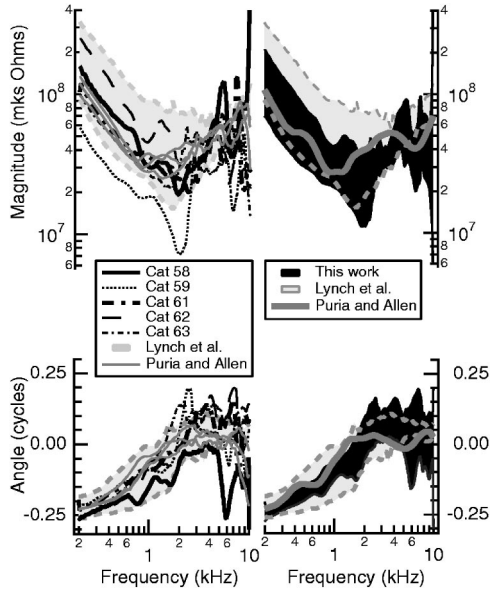


FIG. 6. Magnitudes (upper) and angles (lower) of the middle-ear impedance, $Z_{TM}(f)$. Left: The impedances measured on the five ears discussed here. Shown for comparison are those reported by Lynch *et al.* (1994) (lightly shaded region enclosed by the dashed lines) and Puria and Allen (1998) (three individuals in gray). Right: The mean and standard deviation for the impedances measured here (dark shaded region). Magnitude means and standard deviations were computed on a log scale. The solid gray line shows Z_{TM} computed using the Puria and Allen (1998) model. The range of values measured by Lynch *et al.* (1994) is enclosed by the dashed gray lines.

IV. DISCUSSION

A. Towards a two-port description of the middle ear

The linearity of middle-ear mechanics below the acoustic-reflex threshold (e.g., Guinan and Peake, 1967; Nedzelnitsky, 1980; Cooper and Rhode, 1992) implies that the middle ear can be completely characterized in terms of its response to pure tones. Since the cochlear contents appear essentially incompressible at audio frequencies (Voss *et al.*, 1996; Shera and Zweig, 1992a), the complex pressures and volume velocities on either side of the middle ear are related by a "transfer matrix," $\mathbf{T}_{ME}(f)$ (Shera and Zweig, 1992b). The 2×2 matrix $\mathbf{T}_{ME}(f)$ relates the input and output of the middle ear. The two input variables are the pressure at the tympanic membrane (P_{TM}) and the volume velocity at the tympanic membrane (U_{TM}), and the two output variables are the pressure across the cochlear partition (P_C) and the volume velocity of the stapes (U_S). The matrix $\mathbf{T}_{ME}(f)$ is defined by the equation

$$\begin{pmatrix} P_{TM} \\ U_{TM} \end{pmatrix} = \mathbf{T}_{ME} \begin{pmatrix} P_C \\ U_S \end{pmatrix}, \tag{4}$$

with the four complex matrix elements of $\mathbf{T}_{ME}(f)$ denoted by $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$. The matrix elements of $\mathbf{T}_{ME}(f)$ have simple interpretations obtained by considering specific loading conditions (Shera and Zweig, 1992b). If the stapes is immobilized so that $U_S=0$, then $A=P_{TM}/P_C$ and $C=U_{TM}/P_C$. In other words, A^{-1} is the "infinite-load" forward pressure transfer ratio and C^{-1} is the "infinite-load" forward transfer impedance. If $P_C=0$ (e.g., if the cochlear fluids are drained), then $B=P_{TM}/U_S$ and $D=U_{TM}/U_S$. In other words, B^{-1} is the "no-load" forward transfer admittance and D^{-1} is the "no-load" forward velocity transfer ratio.

The matrix $\mathbf{T}_{ME}(f)$ provides a meaningful description of the middle ear whenever the four variables defining the transformation constitute the effective input and output of the system. So long as the input and output are effectively one-dimensional, the vibration of the eardrum and ossicles can be arbitrarily complicated, involving complex motions in all three spatial dimensions (e.g., Decraemer *et al.*, 1991). On the input side, the pressure in the cat ear canal a few millimeters from the eardrum is approximately uniform in any cross section at frequencies less than roughly 20 kHz (Lynch, 1981; Rosowski *et al.*, 1988). On the output side, measurements near the oval window in the basal turn are consistent with the "long-wavelength approximation," indicating that the pressure is essentially uniform across the stapes footplate, at least for frequencies much less than the local characteristic frequency (Nedzelnitsky, 1980). In addition, the motion of the stapes is largely one-dimensional ("piston-like") in cat (Guinan and Peake, 1967; Decraemer, 2004b),

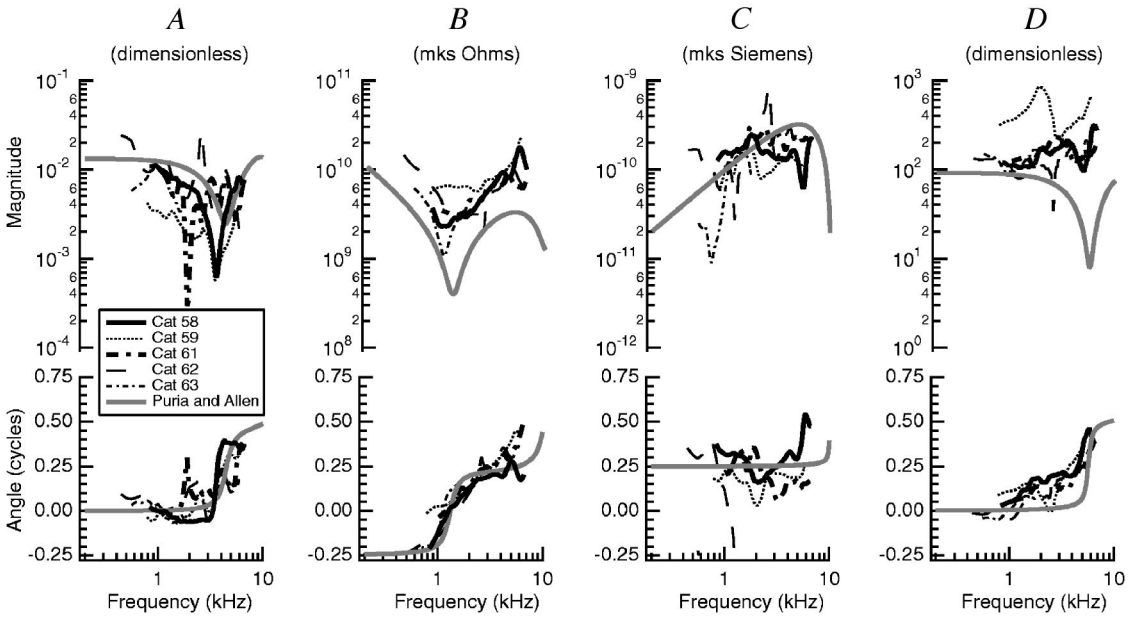


FIG. 7. Magnitudes (upper) and angles (lower) of the matrix elements *A*, *B*, *C*, and *D* computed [Eq. (A4)] using our data and Lynch *et al.*'s (1982) model of Z_C. Matrix elements computed from the Puria and Allen (1998) model are shown for comparison.

allowing us to estimate stapes volume velocity as the product of measured stapes velocity and the area of the footplate.

Although the matrix $\mathbf{T}_{ME}(f)$ characterizes the transmission properties of the middle ear in a manner independent of any sources or loads presented to it, experimental determination of the elements of $\mathbf{T}_{ME}(f)$ requires knowledge and/or manipulation of the loads at both ends of the middle-ear system. The two loads in our preparation are (1) the combination of the Thévenin impedance of the transducer inserted in the ear canal and the air space between the transducer and the tympanic membrane, Z_{SRC}, and (2) the cochlear input impedance, Z_C. Although Z_{SRC} is known (Sec. II J), Z_C cannot be measured directly without inserting pressure transducers into the cochlear vestibule, a procedure that can introduce bubbles or other artifacts that modify the effective value of Z_C one seeks to measure. Fortunately, the value of Z_C can be determined without direct measurement if its value can be manipulated in some way (e.g., if the impedance can be reduced effectively to zero by draining the cochlear fluids). At every frequency five independent, complex measurements are then needed to determine the four matrix elements of $\mathbf{T}_{ME}$ and the value of Z_C (five equations determine five unknowns). At least one of these measurements must be obtained while driving the middle ear ''in reverse'' (i.e., from within the inner ear). Perhaps the five most convenient measurements are (1) the middle-ear input impedance; (2) the forward and (3) reverse stapes-velocity transfer functions; (4) the ''no-load'' (or ''short-circuit'') middle-ear input impedance; and (5) the ''no-load'' forward stapes transfer function. The ''no-load'' conditions refer to measurements made with the cochlear fluids drained (e.g., Allen, 1986).

B. Estimates of the matrix elements

The measurements presented here (i.e., forward and reverse transmission and ear-canal impedance) provide only three of the five measurements[2] necessary to determine $\mathbf{T}_{ME}$ and Z_C. Nevertheless, we can use our measurements to estimate all four matrix elements of $\mathbf{T}_{ME}$ by (1) applying the principle of reciprocity to obtain the additional constraint $\det \mathbf{T}_{ME}=1$ (e.g., Shera and Zweig, 1991) and (2) using values of Z_C measured in other preparations (e.g., Lynch *et al.*, 1982). Assuming that the middle ear is indeed a reciprocal mechanical system, this procedure should yield accurate estimates of $\mathbf{T}_{ME}$ at frequencies where the matrix elements are not especially sensitive to the value of Z_C. Expressions for the four matrix elements $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ in terms of Z_{TM}, $\overrightarrow{T}_S$, $\overleftarrow{T}_S$, Z_{SRC}, and Z_C can be found in the Appendix.

Figure 7 shows our estimates of the matrix elements of $\mathbf{T}_{ME}(f)$ obtained in this way. For simplicity, we approximate Lynch *et al.*'s (1982) measurements of Z_C by their circuit model, which provides a good description of their averaged data. The Appendix demonstrates that the estimates of *A* and *C* [Eq. (A4)] do not depend on the assumed value of Z_C. The measurements and the Puria and Allen (1998) model show similar patterns for both the magnitude and angle of *A* at most frequencies. The model and measurements of *C* are similar in their order of magnitude and overall form, but there are no clear similarities in the finer details.

The estimates for *B* and *D* [Eq. (A4)] assume the Lynch *et al.* (1982) form for the cochlear impedance Z_C. To assess the sensitivity of our estimates of *B* and *D* to the assumed value of Z_C, we calculated *B* and *D* for all preparations and from the model[3] (Puria and Allen, 1998) using a range of values for Z_C. Lynch *et al.* (1982) demonstrate interanimal variations in measured impedance magnitude that span a range of roughly 20 dB (from about $|Z_C|/3$ to $3|Z_C|$). The effects on *B* and *D* of similar variations in $|Z_C|$ are shown for cat 58 and for the model by the shaded regions in Fig. 8. (We did not vary the phase of Z_C; Lynch *et al.*'s model Z_C is essentially resistive over the frequency region explored here.) The effects in other preparations were similar to those in cat 58. At most frequencies the parameters *B* and *D* de-

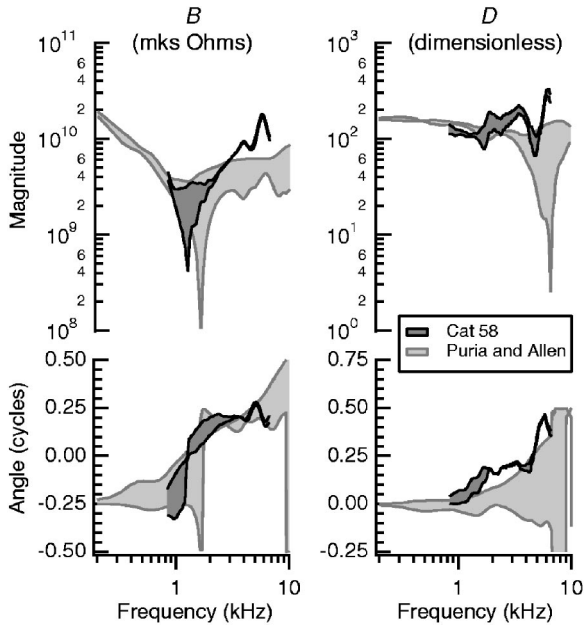


FIG. 8. Sensitivity of the calculations for the matrix elements B (left) and D (right) to the value of the cochlear impedance magnitude $|Z_C|$. The shaded regions enclose the maximum and minimum of the range of magnitude (top) and angle (bottom) values for B and D calculated when $|Z_C|$ was varied up and down by a factor of 3 about its base magnitude. Variations are plotted from measurements on cat 58 (shaded dark gray between black lines) and from the Puria and Allen (1998) model (shaded between gray lines).

rived from the measurements appear more sensitive to individual variations among ears (Fig. 7) than to the imposed variations in $|Z_C|$ (Fig. 8). Thus, at most frequencies, our estimates of B and D appear not to depend substantially on our choice of Z_C. The measurement-based estimates for B are generally similar to the results from the model, while the measurement-based estimates for D differ systematically from the model in both magnitude and angle. The model-based parameters appear more sensitive than the measurement-based parameters to imposed variations in $|Z_C|$.[4]

C. Reverse transmission and the ear-canal load impedance

Just as forward middle-ear transmission depends on the cochlear input impedance, reverse transmission depends on the impedance that loads the ear canal (e.g., Matthews, 1983; Rosowski *et al.*, 1984; Zwicker, 1990; Puria and Rosowski, 1996; Puria, 2003). The two-port description of the middle ear makes this dependence explicit:

$$\overleftarrow{T}_S = \frac{A_S Z_{REV}}{A + C Z_{REV}}, \tag{5}$$

where Z_{REV} is the reverse impedance at the tympanic membrane directed into the ear canal towards the outer ear, and A_S is the area of the stapes footplate. [Equation (5) is derived in the Appendix as Eq. (A3).] Equation (5) implies that $\overleftarrow{T}_S$ is approximately independent of the reverse ear-canal impedance Z_{REV} only if $|Z_{TM}^{\infty}| \ll |Z_{REV}|$, where $Z_{TM}^{\infty} \equiv A/C$. (Note that $|A/C|$ is the value of Z_{TM} when the cochlear load is infinite; e.g., when the stapes is fixed.) Figure 9 shows that

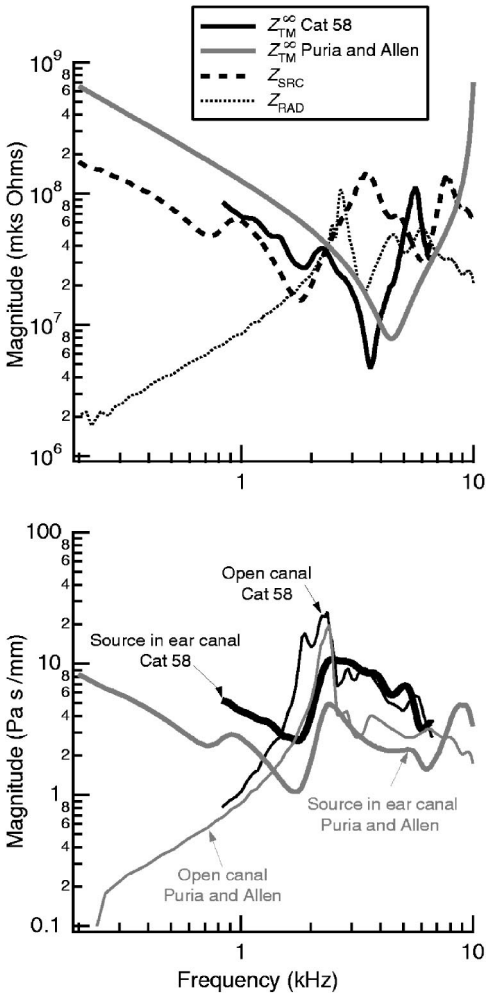


FIG. 9. Upper: Magnitudes of relevant impedances that determine the reverse-transfer function with both the source in the ear canal and the ear canal open to the environment. Z_{SRC} is the Thévenin source equivalent of the earphone and the ear-canal air space; Z_{RAD} is the radiation impedance from the cat ear canal measured by Rosowski *et al.* (1988); and Z_{TM}^{∞} is the ratio A/C calculated from the measurements and the model. Lower: Magnitude of the reverse transfer function as measured with the source in the ear canal, as predicted from the model with the source in the ear canal, and as predicted for the ear canal open to the environment from both cat 58 data and the model.

this condition is almost never satisfied for representative values of Z_{REV}. The upper panel plots $|Z_{TM}^{\infty}|$ estimated from the data and also from the Puria and Allen (1998) model. Also plotted are two different values of the reverse ear-canal load impedance Z_{REV}: (1) the ear-canal air space and the Thévenin source impedance of the acoustic transducer used here (Z_{SRC}) and (2) the radiation impedance of a cat's ear measured at the tympanic membrane (Z_{RAD}) (Rosowski *et al.*, 1988). With the exception of a narrow frequency interval near 3.5 kHz, $|Z_{TM}^{\infty}|$ is always comparable to or greater than $|Z_{REV}|$. Indeed, when the ear canal is open to the environment the opposite limit (i.e., $|Z_{TM}^{\infty}| \gg |Z_{REV}|$) pertains at frequencies less than about 1.5 kHz. Consequently, the reverse transfer function is always strongly dependent on the ear-canal load. We illustrate this dependence in the lower panel of Fig. 9, which plots $|\overleftarrow{T}_S|$ for both terminating impedances. The transfer-function magnitudes are similar in the two cases at frequencies above 2 kHz (where $|Z_{SRC}|$ and $|Z_{RAD}|$ are

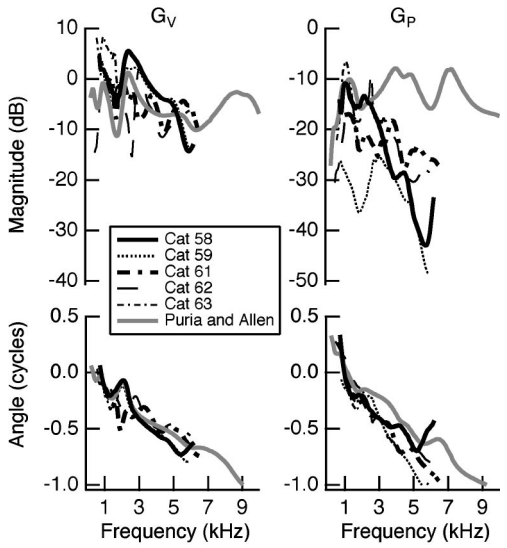


FIG. 10. Left: Magnitude (upper) and angle (lower) of G_V, the round-trip middle-ear velocity gain. A linear fit to all angle data (1 to 5.5 kHz) has a slope corresponding to a delay of 110 μs. Right: Magnitude (top) and angle (bottom) of G_P, the round-trip middle-ear pressure gain. A linear fit to all angle data (1 to 5.5 kHz) has a slope corresponding to a delay of 162 μs.

similar) but differ substantially at lower frequencies (where $|Z_{\mathrm{SRC}}|$ and $|Z_{\mathrm{RAD}}|$ diverge from one another).

D. Round-trip middle-ear gain

The product of the forward and reverse middle-ear pressure transfer functions provides a measure of the total middle-ear gain for otoacoustic emissions reemitted at the stimulus frequency (e.g., SFOAEs and TEOAEs). In these cases, a signal in the ear canal travels through the middle ear (with a gain and phase shift described by the forward transfer function), is reemitted by the cochlea, and travels in reverse through the middle ear back to the ear canal (described by the reverse transfer function). Thus, the product of the forward and reverse middle-ear pressure transfer functions describes the middle ear's influence on these emissions measured in the ear canal. Puria (2003) names this product the ''round-trip pressure gain.''

Here we define G_V as the product of forward and reverse stapes velocity transfer functions:

$$G_V = \vec{T}_S(f)\,\overleftarrow{T}_S(f). \tag{6}$$

The round-trip pressure gain, G_P, can be written in terms of G_V as[5]

$$G_P = \left(\left.\frac{P_C}{P_{EC}}\right|_{\mathrm{forward}}\right)\left(\left.\frac{P_{EC}}{P_C}\right|_{\mathrm{reverse}}\right) = \frac{\vec{Z}_C}{\overleftarrow{Z}_{ME}} G_V, \tag{7}$$

where $\vec{Z}_C$ is the input impedance of the cochlea (here written with an arrow to emphasize that the system is driven in the forward direction) and $\overleftarrow{Z}_{ME} = P_C/U_S$ is the reverse middle-ear (or cochlear output) impedance measured looking out from the cochlea toward the ear canal.

Figure 10 plots G_V and G_P for our five preparations along with those calculated from the model. The magnitude of the velocity gain $|G_V|$ waxes and wanes between -15 and 5 dB, and the magnitude of the pressure gain $|G_P|$ is about 20 dB below that of the velocity gain $|G_V|$. The mean group delays of G_V and G_P provide two different measures of the round-trip delay through the middle ear. Fitting a straight line to the phase data (from 1 to 5.5 kHz) yields mean round-trip delays of 110 ± 4 and 162 ± 6 μs, respectively, where the uncertainties are approximate 95% confidence intervals. In each case, the linear fits account for approximately 80% of the total variance of the data. The mean round-trip group delays found here are similar to those predicted by the Puria and Allen (1998) model of 125 ± 4 and 145 ± 6 μs for G_V and G_P, respectively (again computed from 1 to 5.5 kHz).[6]

V. SUMMARY

We have demonstrated a technique for measuring middle-ear impedance and forward/reverse middle-ear transmission in the same cat ear (with widely opened middle-ear cavities). The method uses DPOAEs as an intracochlear sound source to drive the middle ear ''in reverse'' without opening the inner ear. The measured forward transfer functions and input impedances generally agree with previous measurements, and the measured forward and reverse transfer functions and the input impedances agree qualitatively with model predictions. In addition, we have developed a measurement-based framework for determining the equivalent two-port network (i.e., transfer or $ABCD$ matrix) that characterizes the transmission properties of the cat middle ear. Using this framework we have estimated the matrix elements by combining the measurements presented here with published measurements of the cochlear input impedance.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01 DC03687 from the NIDCD. We gratefully acknowledge the help of Leslie Liberman, who assisted with animal care and preparation, and thank Willem F. Decraemer, John J. Rosowski, and William T. Peake for useful discussions. We thank Sunil Puria for sharing his computer code to compute the responses of his model. William T. Peake, Douglas H. Keefe, and an anonymous reviewer provided helpful comments on the manuscript.

APPENDIX: SOLVING FOR THE MATRIX ELEMENTS OF $\mathbf{T}_{\mathrm{ME}}$

In this Appendix we derive equations that express the four matrix elements of $\mathbf{T}_{\mathrm{ME}}$ in terms of the middle-ear input impedance (Z_{TM}) and the forward and reverse stapes-velocity transfer functions ($\vec{T}_S$ and $\overleftarrow{T}_S$). We assume that the middle ear is a reciprocal mechanical system ($\det \mathbf{T}_{\mathrm{ME}} = 1$) and that both the Thévenin-equivalent impedance of the measurement transducer, including the residual ear-canal space (Z_{SRC}), and the cochlear input impedance (Z_C) are known.

We begin by expressing the three measured quantities in terms of the matrix elements of $\mathbf{T}_{\mathrm{ME}} \equiv \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ defined by Eq. (4):

$$Z_{\mathrm{TM}} = \frac{AZ_C + B}{CZ_C + D}; \tag{A1}$$

$$\overrightarrow{T}_S = \frac{1}{A_S}\frac{1}{AZ_C + B}; \tag{A2}$$

and

$$\overleftarrow{T}_S = \frac{A_S}{A/Z_{SRC} + C}. \tag{A3}$$

Note that the area of the stapes footplate, $A_S = 1.15 \times 10^{-2}\,\mathrm{cm}^2$ as reported by Wever and Lawrence (1954), has been used to convert stapes volume velocity (used in the definitions of $\mathbf{T}_{ME}$ and Z_C) to stapes velocity (used in the definitions of $\overrightarrow{T}_S$ and $\overleftarrow{T}_S$). When the middle ear is driven in reverse we adopt the convention that positive volume velocity flows into the middle ear at the stapes and out at the eardrum. [The signs of U_S and U_{EC} are therefore reversed relative to the convention adopted in Eq. (4); when the middle ear is driven in the forward direction, positive volume velocity flows into the middle ear at the ear drum and out at the stapes.]

When combined with the constraint of reciprocity (which requires $\det \mathbf{T}_{ME} = 1$), Eqs. (A1)–(A3) provide four equations that can be solved for the four matrix elements $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$. The solution is

$$\mathbf{T}_{ME}(f) = \begin{pmatrix} P/Q & 1/\overrightarrow{T}_S A_S - Z_C P/Q \\ R/Q & (1 - \overrightarrow{T}_S A_S Z_C Z_{TM} R/Q)/\overrightarrow{T}_S A_S Z_{TM} \end{pmatrix}, \tag{A4}$$

where

$$Q \equiv \frac{\overleftarrow{T}_S}{A_S}(Z_{SRC} + Z_{TM}); \tag{A5}$$

$$P \equiv Z_{TM} Z_{SRC}(1 + \overrightarrow{T}_S \overleftarrow{T}_S); \tag{A6}$$

and

$$R \equiv Z_{SRC} - \overrightarrow{T}_S \overleftarrow{T}_S Z_{TM}. \tag{A7}$$

[1]The optimized lengths of the six closed brass cylindrical tubes used to determine the Thévenin equivalent were 17.64, 23.45, 28.21, 37.52, 46.48, and 64.08 mm, and the lengths of the five additional tubes used to check the Thévenin equivalent were 20.25, 26.37, 33.97, 40.47, and 53.64 mm. The lengths on these five additional tubes were determined using the same procedures used to determine Z_{Th}. The radii of all tubes were $r = 2.8$ mm.

[2]We made no measurements in the drained-cochlea condition because by the end of each experiment reported here substantial blood clots had formed around the incudo-stapedial joint. Future attempts at measurements with a drained cochlea should be made within a few hours of the start of the experiment, rather than 12–24 h after the start.

[3]The model of Puria and Allen (1998) uses the theoretical Z_C estimated by Puria and Allen (1991) rather than the measurements of Lynch *et al.* (1982).

[4]When the Lynch *et al.* (1982) measurement-based Z_C is used with the Puria and Allen (1998) model, variations in B and D are somewhat smaller than with the model-based Z_C from Puria and Allen (1991).

[5]Puria (2003) defines the round-trip pressure gain in terms of the pressure P_V measured in the vestibule rather than the pressure P_C measured across the cochlear partition. Since the pressure in the vestibule is much larger than the pressure in the scala tympani (Nedzelnitsky, 1980), the difference between these two pressures is minor.

[6]In gerbil Olson (1998) reports a forward delay of about 25 μs.

Allen, J. B. (**1986**). "Measurement of eardrum acoustic impedance," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 44–51.

Avan, P., Büki, B., Maat, B., Dordain, M., and Wit, H. P. (**2000**). "Middle ear influence on otoacoustic emissions. I: Noninvasive investigation of the human transmission apparatus and comparison with model results," Hear. Res. **140**, 189–201.

Buunen, T. J. F., and Vlaming, M. (**1981**). "Laser-Doppler velocity meter applied to tympanic membrane vibrations in cat," J. Acoust. Soc. Am. **69**, 744–750.

Cooper, N. P., and Rhode, W. S. (**1992**). "Basilar membrane mechanics in the hook region of cat and guinea-pig cocheae: Sharp tuning and nonlinearity in the absence of baseline position shifts," Hear. Res. **63**, 163–190.

Decraemer, W. F. (**2004a**). Personal communication.

Decraemer, W. F. (**2004b**). *Measurement, Visualization and Quantitative Analysis of Complete Three-dimensional Kinematical Data Sets of Human and Cat Middle Ear* (World Scientific, Singapore, in press).

Decraemer, W. F., Khanna, S. M., and Funnell, W. R. J. (**1991**). "Malleus vibration model changes with frequency," Hear. Res. **54**, 305–318.

Guinan, J. J., and Peake, W. T. (**1967**). "Middle-ear characteristics of anesthetized cats," J. Acoust. Soc. Am. **41**, 1237–1261.

Huang, G. T., Rosowski, J. J., Flandermeyer, D. T., Lynch, T. J., and Peake, W. T. (**1997**). "The middle ear of a lion: Comparison of structure and function to domestic cat," J. Acoust. Soc. Am. **101**, 1532–1549.

Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000a**). "A noninvasive method for estimating acoustic admittance at the tympanic membrane," J. Acoust. Soc. Am. **108**, 1128–1146.

Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000b**). "Tests of some common assumptions of ear-canal acoustics in cats," J. Acoust. Soc. Am. **108**, 1147–1161.

Keefe, D. H., Ling, R., and Bulen, J. C. (**1992**). "Method to measure acoustic impedance and reflection coefficient," J. Acoust. Soc. Am. **91**, 470–485.

Lynch, T. J. (**1981**). "Signal processing by the cat middle ear: Admittance and transmission, measurements and models," Ph.D. thesis, Massachusetts Institute of Technology.

Lynch, T. J., Nedzelnitsky, V., and Peake, W. T. (**1982**). "Input impedance of the cochlea in cat," J. Acoust. Soc. Am. **72**, 108–130.

Lynch, T. J., Peake, W. T., and Rosowski, J. J. (**1994**). "Measurements of the acoustic input impedance of cat ears: 10 Hz to 20 kHz," J. Acoust. Soc. Am. **96**, 2184–2209.

Magnan, P., Avan, P., Dancer, A., Smurzynski, J., and Probst, R. (**1997**). "Reverse middle-ear transfer function in the guinea pig measured with cubic difference tones," Hear. Res. **107**, 41–45.

Magnan, P., Dancer, A., Probst, R., Smurzynski, J., and Avan, P. (**1999**). "Intracochlear acoustic pressure measurements: Transfer functions of the middle ear and cochlear mechanics," Audiol. Neuro-Otol. **4**, 123–128.

Matthews, J. W. (**1983**). "Modeling reverse middle-ear transmission of acoustic distortion products," in *Mechanics of Hearing*, edited by E. de Boer and M. A. Viergever (Martinus Nijhoff, The Hague), pp. 11–18.

Merchant, S. N., and Rosowski, J. J. (**2003**). "Auditory physiology," in *Surgery of the Ear* (Decker, Hamilton, Ontario), pp. 59–82.

Møller, A. R. (**1965**). "An experimental study of the acoustic impedance of the middle ear and its transmission properties," Acta Oto-Laryngol. **60**, 129–149.

Nedzelnitsky, V. (**1980**). "Sound pressures in the basal turn of the cat cochlea," J. Acoust. Soc. Am. **68**, 1676–1689.

Neely, S. T., and Gorga, M. P. (**1998**). "Comparison between intensity and pressure as measures of sound level in the ear canal," J. Acoust. Soc. Am. **104**, 2925–2934.

Olson, E. S. (**1998**). "Observing middle and inner ear mechanics with novel intracochlear pressure sensors," J. Acoust. Soc. Am. **103**, 3445–3463.

Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.

Puria, S. (**2004**). "Middle-ear two-port measurements in human cadaveric temporal bones: Comparison with models," in *Middle Ear Mechanics in Research and Otology*, edited by A. Gyo and H. Wada (World Scientific, Singapore).

Puria, S., and Allen, J. (**1998**). "Measurements and model of the cat middle ear: Evidence of tympanic membrane acoustic delay," J. Acoust. Soc. Am. **104**, 3463–3481.

Puria, S., and Allen, J. B. (**1991**). ''A parameteric study of cochlear input impedance,'' J. Acoust. Soc. Am. **89**, 287–309.
Puria, S., and Rosowski, J. J. (**1996**). ''Measurement of reverse transmission in the human middle ear: Preliminary results,'' in *Proceedings of the International Symposium on Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. Lyon, P. Narins, C. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 151–157.
Puria, S., Peake, W. T., and Rosowski, J. J. (**1997**). ''Sound-pressure measurements in the cochlear vestibule of human-cadaver ears,'' J. Acoust. Soc. Am. **101**, 2754–2770.
Rabinowitz, W. M. (**1981**). ''Measurement of the acoustic input immittance of the human ear,'' J. Acoust. Soc. Am. **70**, 1025–1035.
Rosowski, J. J., Peake, W. T., and White, J. (**1984**). ''Cochlear nonlinearities inferred from two-tone distortion products in the ear canal of the alligator lizard,'' Hear. Res. **13**, 141–158.
Rosowski, J. J., Carney, L. H., and Peake, W. T. (**1988**). ''The radiation impedance of the external ear of cat: Measurements and applications,'' J. Acoust. Soc. Am. **84**, 1695–1708.
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 Zweig, G. (**1991**). ''Phenomenological characterization of eardrum transduction,'' J. Acoust. Soc. Am. **90**, 253–262.
Shera, C. A., and Zweig, G. (**1992a**). ''An empirical bound on the compressibility of the cochlea,'' J. Acoust. Soc. Am. **92**, 1382–1388.
Shera, C. A., and Zweig, G. (**1992b**). ''Middle-ear phenomenology: The view from the three windows,'' J. Acoust. Soc. Am. **92**, 1356–1370.
Tonndorf, J., and Tabor, J. R. (**1962**). ''Closure of the cochlear windows: Its effect upon air- and bone-conduction,'' Ann. Otol. Rhinol. Laryngol. **71**, 5–29.
Voss, S. E. (**1998**). ''Effects of tympanic-membrane perforations on middle-ear sound transmission: measurements, mechanisms, and models,'' Ph.D. thesis, Massachusetts Institute of Technology.
Voss, S. E., and Allen, J. B. (**1994**). ''Measurement of acoustic impedance and reflectance in the human ear canal,'' J. Acoust. Soc. Am. **95**, 372–384.
Voss, S. E., and Shera, C. (**2002**). ''Simultaneous measurement of DPOAEs, middle-ear input impedance, and forward/reverse middle-ear transmission in cat,'' Assoc. Res. Otolaryngol. Abs. **25**, 585.
Voss, S. E., Rosowski, J. J., and Peake, W. T. (**1996**). ''Is the pressure difference between the oval and round windows the effective acoustic stimulus for the cochlea?'' J. Acoust. Soc. Am. **100**, 1602–1616.
Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2000**). ''Acoustic responses of the human middle ear,'' Hear. Res. **150**, 43–69.
Wever, E. G., and Lawrence, M. (**1954**). *Physiological Acoustics* (Princeton U.P., Princeton).
Zwicker, E. (**1990**). ''On the influence of acoustical probe impedance on evoked otoacoustic emissions,'' Hear. Res. **47**, 185–190.

Distortion product otoacoustic emission (DPOAE) input/output functions and the influence of the second DPOAE source[a)]

Manfred Mauermann[b)] and Birger Kollmeier[c)]
Medizinische Physik, Universität Oldenburg, D-26111 Oldenburg, Germany

(Received 1 April 2004; revised 12 July 2004; accepted 19 July 2004)

Distortion product otoacoustic emissions (DPOAEs) at $2f_1-f_2$ ($f_2/f_1=1.2$) have two components from different cochlear sources, i.e., a distortion component generated near f_2 and a reflection component from the characteristic site of f_{DP}. The interaction of the two sources may negatively affect the DPOAE input/output (I/O) functions that are used to predict either auditory thresholds or the compression characteristics of the basilar membrane. This study investigates the influence of the reflection component on DPOAE I/O functions in a frequency range for f_2 from 1500 to 4500 Hz in steps of 18 Hz. A time windowing procedure is used to separate the components from the two DPOAE sources. With decreasing stimulus level, the relative contribution of the reflection component increases. I/O functions from the separated distortion component (DCOAE I/O functions) only show smooth changes in shape and slope with frequency, while ''standard'' DPOAE I/O functions show rapid changes between adjacent frequencies, indicating a strong influence from the interference with the second DPOAE source. A reduced variability for adjacent frequencies can be seen as well for prediction of hearing thresholds, when using DCOAE instead of DPOAE I/O functions. [DOI: 10.1121/1.1791719]

PACS numbers: 43.64.Jb, 43.64.Kc [BLM] Pages: 2199–2212

I. INTRODUCTION

Distortion product otoacoustic emissions (DPOAEs) are low-level sinusoids recordable in the occluded ear canal at certain combination frequencies during continuous stimulation with two tones at the frequencies f_1 and f_2 ($f_1<f_2$). They are the result of the nonlinear interaction of the two tones in the cochlea. Various clinical studies have claimed that DPOAEs provide a useful objective audiometric test with high frequency specificity. But in most studies, the reported correlation of audiometric thresholds and DPOAE levels is based only on large averaged databases of many subjects and does not permit individual threshold prediction (e.g., Nelson and Kimberley, 1992; Gorga *et al.*, 1993; Moulin *et al.*, 1994; Suckfüll *et al.*, 1996). However, in the average the DPOAE levels measured at moderate stimulus levels were correlated to hearing threshold. In more recent studies, DPOAE I/O functions were used as a tool to determine the characteristics of BM compression (Buus *et al.*, 2001) or are suggested as tools for the prediction of loudness growth, i.e., to indicate recruitment in subjects with cochlear hearing loss (Neely *et al.*, 2003). Boege and Janssen (2002, see also Gorga *et al.*, 2003) suggested a method to predict thresholds based on individual DPOAE I/O functions, which improves the individual and quantitative prediction of pure tone threshold remarkably. However, there are still large standard errors of the difference between predicted and behavioral threshold. Both the objective prediction of auditory threshold and the prediction of recruitment are of extended interest as an audiometric tool. They may serve as important input parameters for a hearing aid fitting, especially in young children. As one step towards a possible improvement of predictions from DPOAE I/O functions, this paper aims at clarifying to which extent the variability in estimating individual cochlear functions is related to the influence of a second DPOAE on the properties of DPOAE I/O functions.

Based on theoretical and experimental work (Brown *et al.*, 1996; Gaskill and Brown, 1996; Heitmann *et al.*, 1998; Talmadge *et al.*, 1998, 1999; Shera and Guinan, 1999; Mauermann, 1999a,b) it is widely accepted that the DPOAE at $2f_1-f_2$ in human subjects with f_2/f_1 around 1.2 can be interpreted as the vector sum of two components from different cochlear sources: an initial component with its source close to f_2 and a second component from the characteristic site of the distortion product frequency, e.g., $2f_1-f_2$. Several factors indicate that the effects caused by the second source are highly relevant, at least in normal- or near-normal-hearing subjects. The two interfering DPOAE components originate at two different places on the BM and they show different phase characteristics (e.g., Shera and Guinan, 1999). The components interfere at some frequencies constructively when in phase, while they cancel each other out at other frequencies when out of phase. Whenever DPOAEs are measured with a sufficiently high frequency resolution, the DPOAE level shows large quasi-periodic variations (up to 20 dB) across frequency, the so-called DPOAE fine structure (e.g., Mauermann *et al.*, 1997b). Such large interference effects occur only when both vector components are similar in magnitude. In some subjects, the second or ''reflection component'' even exceeded the contribution from the initial distortion source (Talmadge *et al.*, 1999). Multiple reflections on the basilar membrane between the oval window and the

[a)]Parts of this study were presented at the Meeting of the Deutsche Gesellschaft für Audiologie (DGA), 2003 in Würzburg, Germany and at the 27th Midwinter Research Meeting of the Association for Research in Otolaryngology 2004 in Daytona Beach, FL.

[b)]Electronic mail: manfred.mauermann@uni-oldenburg.de
[c)]Electronic mail: birger.kollmeier@uni-oldenburg.de

characteristic site of f_{DP} can add even more significant components and hence influence the observed DPOAE fine structure (Dhar *et al.*, 2002). Furthermore, DPOAE fine-structure patterns show a level-dependent shift in frequency (He and Schmiedt, 1993; Mauermann *et al.*, 1997b). This shift can cause notches in DPOAE I/O functions for some frequencies, especially those near fine-structure minima (He and Schmiedt, 1993). Overall, the slope and shape of DPOAE I/O functions appear to be strongly influenced by the position of the selected frequency within DPOAE fine structure. But the maxima and minima of DPOAE fine structure are not explicitly related to the local cochlear status around the characteristic place of f_2.[1] The level- and frequency-dependent interference effects between the two components probably provide a reason for the variability that can be observed in DPOAE I/O functions and derived predictions (see, e.g., Boege and Janssen, 2002; Gorga *et al.*, 2003).

Nevertheless, none of the studies using DPOAE I/O functions (at a fixed f_2/f_1 ratio) for the prediction of hearing capabilities takes those effects into account (as far as known by the authors). The current study, therefore, investigates to what extent DPOAE I/O functions are affected by the contribution of the second source and if the variability in predictions derived from DPOAE I/O functions can possibly be reduced when the contribution from the second source is considered. Until now, studies on the contribution of the second source have mainly concentrated on measurements at moderate primary levels (e.g., Talmadge, 1999; Mauermann *et al.*, 1999a,b; Kaluri and Shera, 2001) and have not investigated the large range of different levels needed to characterize the influence of the second source on DPOAE I/O functions. There are some indications that the relative influence of the second source changes with stimulus level, e.g., in some subjects the DPOAE fine structure flattens out with increasing level (e.g., Mauermann *et al.*, 1997b). Konrad-Martin *et al.* (2001) investigated the energy of low and high latency components from DPOAE measurements for a fixed f_2 paradigm. They used $L_1 = L_2 + 10$ dB and found the low latency component increasing faster than the high latency component with increases in stimulus level. However, there is no detailed investigation of the changes in the relative contribution of the second source with changes in primary levels when the frequency ratio (f_2/f_1) is fixed. Consequently, we do not know how the relative levels of the components may affect DPOAE I/O functions. To clarify this point, in the first part of this paper, ''standard'' two-source DPOAE I/O functions for frequencies f_2 from 1500 to 4500 Hz (L_2: 20–80 dB in 10-dB steps) are compared with I/O functions of the contribution of only the initial ''distortion component'' (from the f_2 place). We refer to these I/O functions from the ''distortion component only'' as ''distortion component otoacoustic emission'' (DCOAE) I/O functions. To obtain these DCOAE I/O functions, DPOAE components from the two sources must be separated. Recent studies on DPOAE component separation (e.g., Knight and Kemp, 2001; Kalluri and Shera, 2001) suggested two methods to unmix the DPOAE components: time windowing and selective suppression. Both methods showed similar results (Kalluri and Shera, 2001): a smoothed DP-Gram (DPOAE level as function of frequency). However, in these studies, only a fixed set of primary levels was investigated. DPOAE I/O functions were not obtained. Since adequate suppressor levels for the method of selective suppression are not yet known for the whole range of primary levels necessary for this paradigm, we used the time windowing method.

In addition to the theoretical advantage that DCOAE are generated at a single cochlear site (permitting a easier frequency specific interpretation of the I/O functions), the comparison of DPOAE and DCOAE I/O functions in the current study indicates a strong reduction in variability of I/O functions from adjacent frequencies. This could lead to a reduced variability in threshold predictions derived from these I/O functions. Therefore, in a second step, we investigated whether methods for the prediction of pure tone thresholds from DPOAE measurements (Boege and Janssen, 2002; Gorga *et al.*, 2003) could be improved by using DCOAE instead of DPOAE I/O functions as a basis for threshold predictions. Boege and Janssen (2002) and Gorga *et al.* (2003) excluded those DPOAE I/O functions that did not satisfy certain criteria from their analyses. Possibly these functions failed the inclusion criteria because of level and frequency-dependent interference effects from the two sources. Even DPOAE I/O functions that were included may have been affected by contributions of the second DPOAE source. This may have led to an additional variation in the prediction of pure tone thresholds.

Useful prediction of hearing capabilities from DPOAE I/O functions (such as hearing threshold) should be expected to give almost similar results from adjacent frequencies, i.e., threshold estimate should change smoothly with frequency and there should be low variability between adjacent frequencies. No behavioral threshold measurements were performed in this study (except standard audiogram for a rough classification of the subjects). However, as an indication of an improvement of threshold prediction we do expect (1) a reduction in the number of rejected I/O functions and (2) reduced variability of predicted thresholds for adjacent frequencies. To check for a potential gain by the use of DCOAE I/O functions, thresholds estimations and I/O slopes [according to Boege and Janssen (2002)] are computed and compared from both (1) ''standard'' DPOAE I/O functions and from (2) DCOAE I/O functions at adjacent frequencies (f_2 in steps of 18 Hz).

II. EXPERIMENTAL METHODS

A. Subjects

DPOAE data were collected from the right ears of six subjects, aged from 21 to 50 years. Five normal-hearing subjects (AP, BS, IT, JF, and MO) had hearing thresholds better than 15 dB HL at the standard audiometric frequencies from 125 Hz to 8 kHz and one subject (AS_{HL}—for an easier identification of the subject with hearing loss an index HL for ''hearing loss'' is added to this subject's identifier) showed a moderate high-frequency loss (30, 35, and 20 dB HL at 4, 6, and 8 kHz, respectively). All subjects had normal middle-ear function, which was tested by tympanometric measurements

(Interacoustics AZ26). The subjects were seated comfortably in a sound-treated booth during the measurement sessions of 90–120-min duration.

B. Instrumentation and signal processing

An insert microphone, ER10B+ (Etymotic Research), was used to record the DPOAE in the closed ear canal. The microphone output was connected to a low noise amplifier, type SR560 (Stanford Research), and then converted to digital form using a 24-bit ADI-8 Pro AD/DA converter (RME), which was connected optically to a Digi96/8 PAD (RME) digital I/O card in a personal computer (PC). Further on-line analysis was performed using custom-made measurement software (Matlab based—Vers. 6.5, Mathworks) on a PC. All stimuli were generated and attenuated digitally. The output of the DA converters (ADI 8 Pro, 24 bit) was connected to two separate TDT-HB7 headphone buffers (Tucker Davies Technologies). These drove separately, either two ER2 insert-ear phones (Etymotic Research) or two encapsulated speakers of a DT48 headphone (Beyerdynamics) that were coupled with elastic sound tubes to the ER10B. The on-line analysis was used for signal conditioning of the recorded signal including artifact rejection and averaging in the time domain to improve the signal-to-noise ratio. Time frames of $\frac{1}{3}$-second length were used (16 000 samples at a sampling frequency of 48 000 Hz or 14 700 samples at 44 100 Hz). This permitted selection of primary frequencies that are harmonics of 3 Hz for continuous stimulation. Selecting f_2 frequencies at harmonics of 18 Hz and the related f_1 frequencies as harmonics of 15 Hz (both harmonics of 3 Hz) gave primary frequency pairs with an exact frequency ratio f_2/f_1 of 1.2. The two primaries were presented via two separated speakers.

The noise floor of each frame around $2f_1-f_2$ was estimated from the average soundpressure of 15 samples above and below the $2f_1-f_2$ sample in the DPOAE spectrum. The automatic artifact rejection was based on the individual noise floor computed for each frame. The rejection criteria were adapted during the measurement to decrease the averaged noise floor. During each measurement, all recorded frames were kept in memory. The frames to be averaged were selected from all recorded frames by the actual rejection criterion. The averaging stopped automatically if the averaged noise floor was -20 dB SPL or below, and the S/N reached 12 dB. The measurement did not stop before a minimum number of 4 averaged frames, but it always stopped when 128 valid averaged frames or 256 recorded frames were completed. At the beginning of each measurement session, the correct fit of the probe in the ear canal was checked by obtaining the ear canal response to a broadband stimulus. This ''probefit procedure'' was always repeated after 20–40 measurements, to check the stability of the probe fit during a measurement session.

C. Calibration

Level and phase characteristic of the speakers, connected to the ER10B+ probe microphone system, were calibrated using an artificial ear for insert probes (Bruel & Kjaer BK4157). No further individual phase correction during the measurements was performed. The overall level was adjusted slightly for each subject to compensate roughly for the individual ear canal volumes (corrections were less than ± 2 dB). The correction level was obtained by the level difference in the 1–1.5-kHz frequency band of the expected level according to the coupler calibration and the probefit signal recorded individually in each ear canal during the first probefit measurement. The microphone was calibrated as suggested by Siegel (2002), using a Bruel & Kjaer coupler microphone (B&K 4192) as the reference. In the following, we refer to this calibration as coupler calibration.

D. DPOAE measurements

DPOAE at $2f_1-f_2$ ($f_2/f_1=1.2$) were measured with a high frequency resolution of 12 Hz at seven levels L_2 from 20 to 80 dB in 10-dB steps. The corresponding L_1 levels were chosen according to the formula $L_1=0.4L_2+39$ dB as suggested by Kummer *et al.* (1998) to obtain maximum DPOAE levels and an almost logarithmic characteristic of the DPOAE levels as a function of L_2 (Boege and Janssen, 2002). The DPOAE were measured over an f_2 frequency range from 1500 to 4500 Hz, divided into four frequency bands (a) 1.5–2.25 kHz, (b) 2.25–3 kHz, (c) 3–3.75 kHz, and (d) 3.75–4.5 kHz. Within each band, all 301 measurement points (43 frequencies at 7 levels) were randomized to avoid systematic effects in the data with measurement time. Normally one of the frequency bands was measured in a measurement session of about 90- to 120-min duration. The order of the bands was randomized. Finally, the results of the four frequency bands were combined and used for further analysis.

III. DATA ANALYSIS

A. Separation of DPOAE components via time windowing

Earlier studies (e.g., Kaluri and Shera, 2001) have shown that DPOAE at a frequency ratio f_2/f_1 of 1.2 can be treated as the interference of two components with very different phase slope or group delay. This property is used by the method of time windowing to separate the two DPOAE components. Several authors had used time windowing or spectral smoothing in slightly modified form (e.g., Knight and Kemp, 2001; Kaluri and Shera, 2001). The DP-Gram data are transformed to the time or latency domain using an inverse fast Fourier transformation (IFFT). A large phase gradient with frequency (i.e., a long group delay) corresponds to a delayed peak in the resulting ''time domain response.'' We expect a ''time-domain response'' with two peak regions indicating the different group delays or latencies of the two interfering DPOAE components. These two peak regions can be separated by multiplication with an adequate time window.[2] Transforming the low-latency components back to the frequency domain by a fast Fourier transform (FFT) results in a smoothed DP-Gram representing the DPOAE levels from only the initial distortion source near f_2.

The implemented evaluation procedure converts the pressure amplitude and phase spectra from the measured

DP-Grams[3] into their real and imaginary parts for each frequency. The array of complex values was shifted in frequency down to the lowest frequency bin above zero. The complex spectrum was extended to include its negative frequencies[4] using the complex conjugated values of the positive frequencies mirrored at zero.[5] Thus, the IFFT of the complete complex spectrum resulted in a real-time signal similar to an ''impulse response'' referred to as ''time-domain response.''[6] The left panels of Fig. 1 show the absolute values of the resulting real time response obtained by the IFFT for an easier distinction of the two peak regions. In further analysis the original real time signal is used containing both the complete amplitude and phase information and not the absolute values.

To reduce boundary effects, the frequency range of each DP-Gram was selected individually for the IFFT, so that the periodic continuation (due to the discrete signal processing) of the extended DP-Gram spectrum resulted in shapes at the boundaries that were almost typical for the shape of the whole DP-Gram. The chosen DP-frequencies varied for the lower frequency end between 1015 and 1085 Hz and for the upper frequency end between 2962 and 3014 Hz.

The frequency resolution Δf of 12 Hz in the measured DP-Grams corresponds to a resulting time frame T ($T=1/\Delta f$) in the time domain of about 83 ms. All ''time-domain responses'' were shorter than about 60 ms, i.e., the DPOAE peaks died down to the noise floor after this time. Thus, for a T of about 83 ms no aliasing in the time domain was expected. Due to the discrete signal processing, the IFFT of the complete complex spectrum, resulted in a real-time signal that can be regarded as periodically repeated. Slight shifts of the absolute position of the peaks in the time domain response, i.e., of the absolute group delays, can be expected, e.g., from individual variations of the relative phase of the two primaries at the eardrum. Consequently, the starting point of the time window to cut out the low latency region was chosen individually for each subject in a way that the rising edge of the first peak in the time signal was included, i.e., the time window could start at ''negative'' times. The end of the time window was placed at a point at which the magnitude of the time signal had a dip between low and high latency peaks.[7] The same time window determined for the latency response of $L_2=40$ dB SPL was used at all primary levels. This appeared to be an adequate choice for the windowing of the ''time-domain responses'' of the different stimulation levels, too. We used a rectangular time window with hanning shaped ramps (eight samples at each end). The remaining time signal without the low-latency components gave the high-latency part of the time signal.

The low-latency or DCOAE part of the time signal was transformed back to the frequency domain, using a FFT, resulting in a complex spectrum. From this, the DCOAE amplitudes and phases were computed. For further spectral analysis and prediction of threshold, DCOAE that belong to frequencies affected by artifacts of the time windowing procedure were rejected. At the beginning and at the end of the resulting spectrum, a frequency range equal to the relevant spectral width of the low latency time window was eliminated. As relevant spectral width of the time window, we used $1/T_{win}$ with T_{win} equal to the complete duration of the time window, including the ramps.

Figure 1 illustrates the time windowing procedure for one subject (AP) at three levels. The left panels show the magnitudes[8] of the real ''time-domain responses'' (gray line) of the DP-Grams data at different levels. The low-latency or distortion components (DCs) are indicated as thick black lines, the high-latency or reflection components (RCs) as thin dotted black lines, respectively. The comparison of low- and high-latency peaks for different stimulation levels (left panels) shows that at low-stimulation levels the relative contribution of the high-latency components is clearly higher. The right panels show the corresponding frequency domain representations of the low- and high-latency components and the originally measured DP-Gram as magnitude (b,e,h) and phase spectra (c,f,i).

B. Prediction of hearing threshold derived from DPOAE I/O functions

Boege and Janssen (2002) described a method to predict individual auditory thresholds from DPOAE I/O functions. They used the level paradigm $L_1=0.4L_2+39$ dB for maximum DPOAE levels (Kummer *et al.*, 1998) and showed that the DPOAE I/O functions typically have a logarithmic characteristic (Boege and Janssen, 2002). This gives a linear relation between the DPOAE pressure (in μPa) and stimulus level L_2. The ''estimated distortion product threshold level'' L_{EDPT} is finally determined by the level of L_2 at which the linear extrapolated DPOAE pressure equaled 0 μPa. When the threshold estimation is based on the DCOAE I/O function, we refer to this as ''estimated distortion component threshold level'' L_{EDCT}.

Figure 2 illustrates the scheme for estimating L_{EDPT} and L_{EDCT} using the data from subject IT for the frequencies 2466, 2484, and 2502 Hz, respectively. The left panels show the DPOAE I/O function (gray lines and symbols) in comparison to the DCOAE I/O (black lines and symbols). In the right panels the DPOAE and DCOAE levels are converted into pressure and linear functions were fitted to the data points. The gray or black arrows indicate the threshold estimations, L_{EDPT} and L_{EDCT}, respectively. The three rows show the DPOAE and DCOAE I/O functions for three adjacent f_2 frequencies from the same subject (IT) to provide an example of how rapidly the ''standard'' DPOAE I/O functions and associated threshold predictions can vary when the f_2 frequency varies in 18-Hz steps.

Several criteria had to be fulfilled for the I/O function to be used for further threshold estimation. These were a S/N criterion that states that at least three of the measured points need a S/N$\geqslant$6 dB, and three criteria which describe the quality of the linear fit (fitting criteria):

(a) stability index criterion: $r^2\geqslant 0.8$,
(b) standard error criterion: stderr <10 μPa,[9]
(c) slope criterion: $s\geqslant 0.2$.[10]

Boege and Janssen (2002) used DPOAE I/O data at stimulus levels L_2 from 20 to 65 dB in 5-dB steps to fit to a linear equation. In the current study, we measured DPOAE I/O data

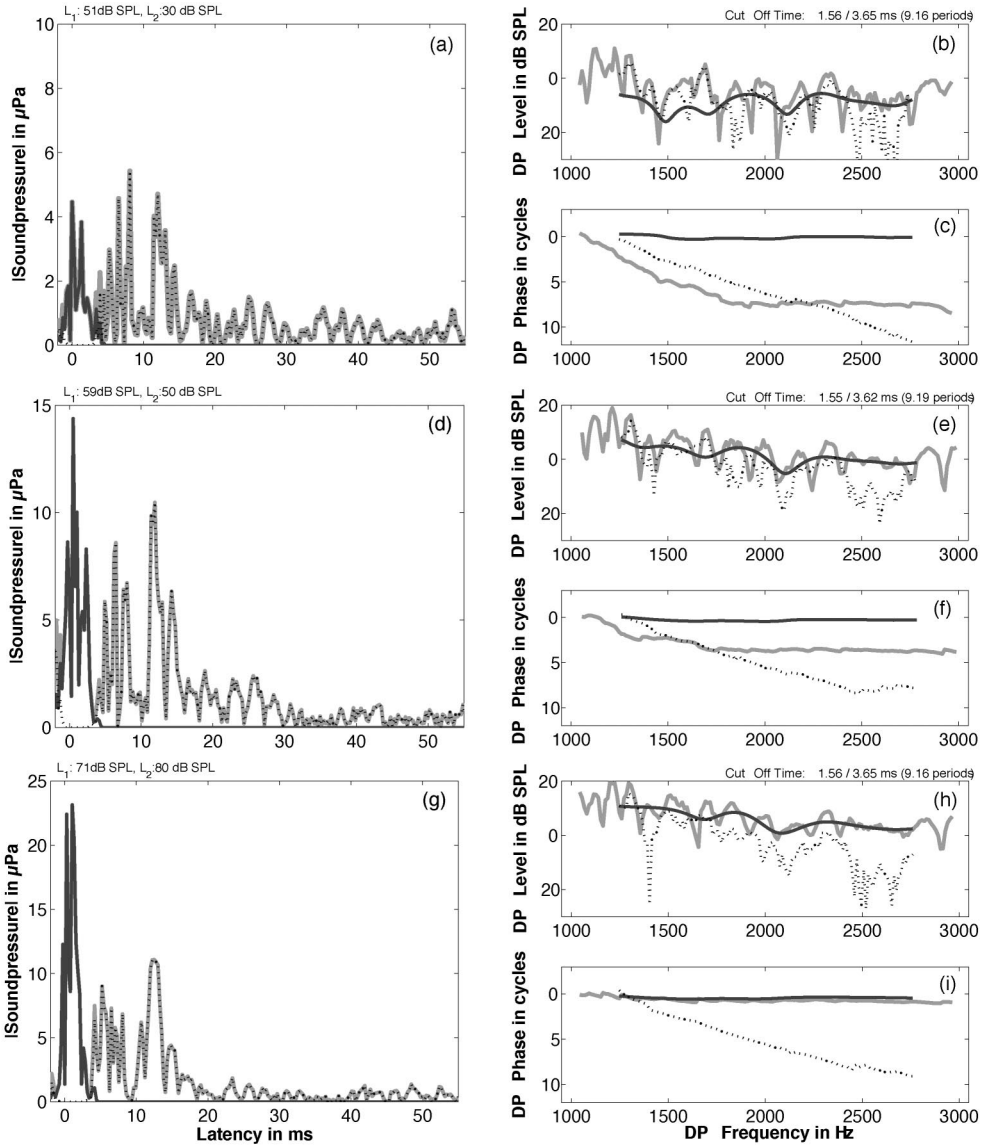


FIG. 1. Illustration of the time windowing procedure for the separation of DPOAE sources. Time domain (a,d,g) and frequency domain representations [(b,e,h) magnitude spectra; (c,f,i) phase spectra] of DPOAE, DCOAE and RCOAE at three stimulus levels L_2 [30 dB SPL (a,b,c); 50 dB SPL (e,f,g); and 80 dB SPL (g,h,i)] from subject AP. Thick gray lines in the right panels indicate the magnitude or respectively phase data of the measured DP-Grams. The frequency range used for the time windowing analysis is given by the frequency range of these lines. In the left panels the gray lines indicate the magnitude of the corresponding real time domain response (partly covered by the thick black lines). The thick black lines in the left panels indicate the low latency components (time domain responses of the DP-Gram data multiplied with the low latency window). In the right panels these lines indicate the corresponding magnitude and phase representation in the frequency domain that gives the DC-Gram data, i.e., the emission level and phase of the separated distortion component. The thin dotted black lines in the left panels and right panels indicate the time domain or corresponding frequency domain representation of the high latency components, respectively. They represent the reflection emissions from the second source at the f_{DP} site of the cochlea. The ''Cut-Off'' time on top of the panels (b), (e), and (h) give the start and end time of the selected low latency window. The number of periods gives the corresponding time duration in periods of the average frequency of the analyzed frequency range (the geometric mean of highest and lowest frequency in the DP-Gram) for comparison with data from Kalluri and Shera (2001).

from 20 to 80 dB in 10-dB steps. For the prediction of auditory thresholds following the rules of Boege and Janssen, we used the data obtained using L_2 levels from 20 up to 70 dB SPL. Although the level paradigm, $L_1 = 0.4L_2 + 39\,\mathrm{dB}$, has only been evaluated extensively for levels up to 65 dB (Kummer *et al.*, 1998) it appears to be reasonable to include the data points from $L_2 = 70\,\mathrm{dB}$ for the threshold estimation, because they fit the assumption of a logarithmic increase of DPOAE level with stimulus level very well. In the current data, this assumption did not hold for L_2 at 80 dB SPL.[11] Therefore, data obtained with L_2 at 80 dB SPL were excluded from the threshold prediction. Only those frequencies that were not affected by artifacts of the time windowing procedure are used for DCOAE I/O functions. This resulted in about 130 comparable DPOAE I/O and DCOAE I/O functions at adjacent frequencies for each subject. Linear fits,

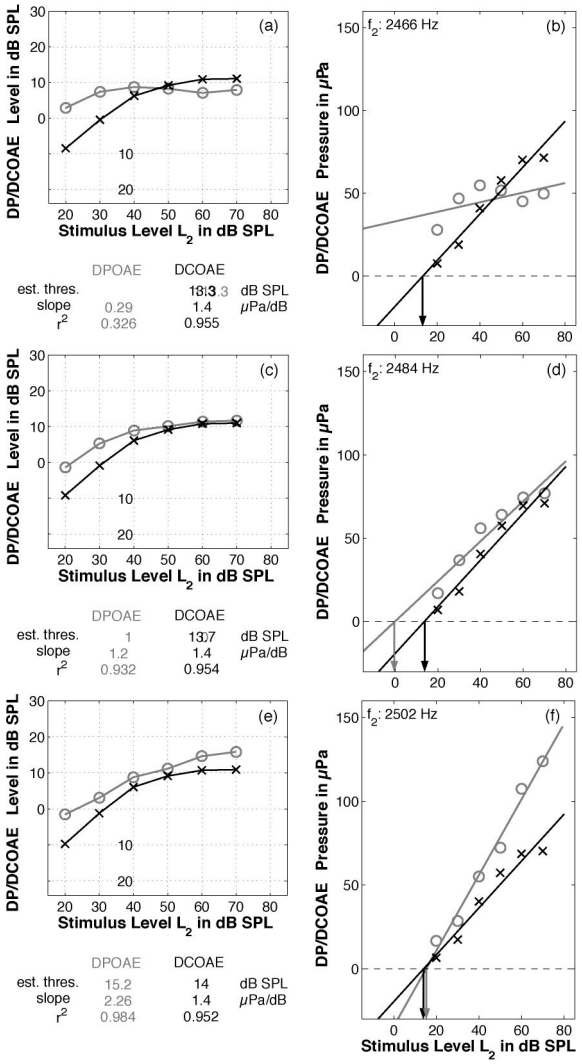


FIG. 2. Illustration of the threshold estimation from DPOAE I/O functions. Left panels show DPOAE (gray lines and symbols) and DCOAE (black lines and symbols) I/O functions from three adjacent frequencies 2466, 2484, and 2502 Hz for subject IT. The right panels show the same data as the corresponding left panels but converted into pressure in μPa. Linear functions are fitted to the data points in the right panels for threshold estimations from DPOAE and DCOAE. The threshold levels L_{EDPT}, L_{EDCT} are the stimulus level corresponding to 0 μPa estimated from the extrapolation of the linear fitted functions (see gray and black arrow in the right panels). The values of stability index r^2 and slope s of the linear fitted functions as well as the estimated thresholds for the tree frequencies from the DPOAE and DCOAE I/O functions are given in small tables below panel (a), (b), and (c), respectively. Notice the strong fluctuations in threshold prediction from the DPOAE I/O functions of the adjacent frequencies while the predictions from DCOAE result in highly comparable values.

including threshold estimations and slopes, were computed for both DPOAE I/O functions and DCOAE I/O functions. To estimate the S/N for the DCOAE data, the noise floor from the measured DPOAE data was used and related to the DCOAE level for each frequency. The estimations of thresholds from DPOAE I/O functions are compared with those from DCOAE I/O functions. We assume that there should not be an extensive variation of threshold prediction from frequency to frequency and thus the procedure that gives the most stable estimate of threshold is the optimal procedure.

IV. RESULTS

A. Contribution of the second source to the DPOAE at different levels

Figure 1 shows the comparison of measured DPOAE (gray line, in the time domain partly covered by the solid black line), the combination of the distortion component (DC or low latency components, dotted black line) and the reflection component (RC or high latency components, solid black line), in the time (a,d,g) and in the frequency domain (g,c,e,f,h,i) for one subject (AP) at three stimulus levels. In a manner analogous to DP-Gram, we refer to the magnitude of the low latency, i.e., distortion component as DC-Gram and for the reflection component as RC-Gram, respectively. The comparison of DC-Grams and RC-Grams permits a direct frequency specific description of the relative contribution of the two DPOAE components. In panel (b) of Fig. 1 we see that the RC-Gram for L_2 at 30 dB (dotted black line) clearly exceeds the contribution of the DC (solid black line) for f_{DP} frequencies from about 1300 to 1700 Hz. Both components are almost equal for frequencies f_{DP} up to 2400 Hz, while the RC decreases for higher frequencies. The frequency regions dominated by the RC are also directly reflected in the phase characteristic of the original DP-Gram [see Figs. 1(c), (f), and (i); solid gray line]. If the magnitude of the reflection emission exceeds the DC [compare black dotted lines and black solid line in Figs. 1(b), (e), and (h), the phase of the measured DP-Gram follows the phase characteristic of the RC, with its strongly rotating phase. There is almost no phase slope or group delay for frequencies dominated by the distortion component [see Figs. 1(b) and (c)]. With increasing stimulus level, the relative contribution of the RC decreases for all frequencies and the phase slope of the DP-Gram data is almost identical with the phase slope of the DC [see Fig. 1(i)]. The remaining subjects show similar results to those shown in Fig. 1 for subject AP. Specifically, for all subjects except IT (not shown here), the data shows that the relative contribution of the second source at lower frequencies (i.e., for f_{DP} up to about 2200 Hz) exceeds the relative contribution of the second source at higher frequencies. For subject IT, the contribution from the second source appears to be almost frequency independent.

In Fig. 3 the relative contribution of DCs and RCs across frequency are summarized for all subjects. To quantify the relative contribution of DCs versus RCs or low- versus high-latency components, respectively, the levels of the root mean square[12] values of the RCs of the ''time-domain responses'' are subtracted from those of the DCs. The RCs enclose the components from all higher latencies, since all high-latency components, even those from multiple reflections, are contributions from the second (or reflection) source and affecting the DPOAE fine structure. Negative values of the difference ΔL indicates a dominance (averaged over all frequencies) of the RCs while positive ΔL indicate a dominance of DCs, i.e., of the initial DPOAE source. These differences ΔL are plotted as functions of stimulus level L_2 (Fig. 3). With decreasing stimulus level L_2 an increase of the contribution from the RCs is found in all subjects. At low levels, the second or reflection source dominates the DPOAE recorded in the ear

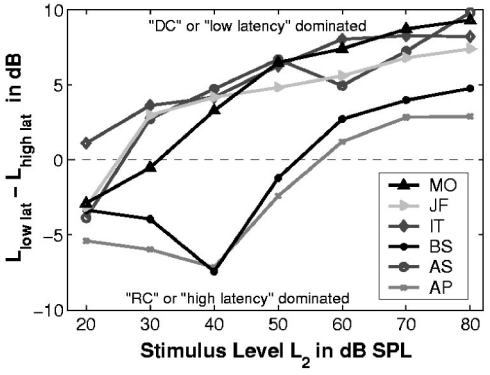


FIG. 3. Differences of the levels of low and high latency components as a function of stimulus level L_2, derived from the time domain responses, are shown for all subjects. The different lines and symbols indicate the results from the different subjects (see legend). Values below zero indicate that the DPOAEs measured in the ear canal are dominated by high latency components while positive values indicate an overall dominance of the low latency components or distortion components generated near f_2. Notice the increased contribution of high latency components with decreasing stimulus level.

canal. In some subjects the RCs dominate the DPOAE recorded in the ear canal even with moderate stimulus levels of 50 dB SPL. Subjects AS and BS, who have strong RCs, show a decrease of ΔL when L_2 increases from 20 to 40 dB SPL. This might be an effect of the poorer signal-to-noise ratio at low stimulation levels. Thus, more noise falls in the longer time windows of the RCs than in the shorter time windows used for the DCs. However, one may speculate that in this level range, the RCs increase faster than the DCs, while the suppression effect of the primary f_1 increases with increasing stimulus level L_1 and reaches a sufficient magnitude to suppress the RCs at levels above 40 dB.

B. Variability of DPOAE I/O functions versus DCOAE I/O functions

Figure 4(a) shows an example from one subject (BS) of DP-Gram data for L_2 levels from 20 to 80 dB SPL. The DP-Gram fine structure is typical of the data from all subjects tested. The DP-Gram shows slight shifts or changes in the fine structure with level. This level dependence might be interpreted in terms of the relative contribution of the two sources changing with level. Figure 4(c) gives the DC-Grams computed from the DP-Grams of Fig. 4(a) by time windowing. The appropriate DPOAE I/O functions [Fig. 4(b)] show a large variability with frequency, including notches and rapid changes of overall shape between adjacent frequencies, while the DCOAE I/O functions [Fig. 4(d)] show a very smooth change with frequency. This is even clearer in Fig. 5 where the DPOAE and DCOAE I/O functions from another subject (BS) are compared over three small frequency bands. DCOAE I/O functions show a drastically reduced variability between adjacent frequencies in comparison to DPOAE I/O functions within the identical frequency range for each subject. Table I gives quantitative indicators for this different variability of DPOAE and DCOAE I/O functions within narrow frequency bands (1872–2232, 2250–2610, 2628–2988, 3006–3366, 3384–3744, and 3762–4122 Hz) for all subjects. Therefore, we computed for each subject the standard

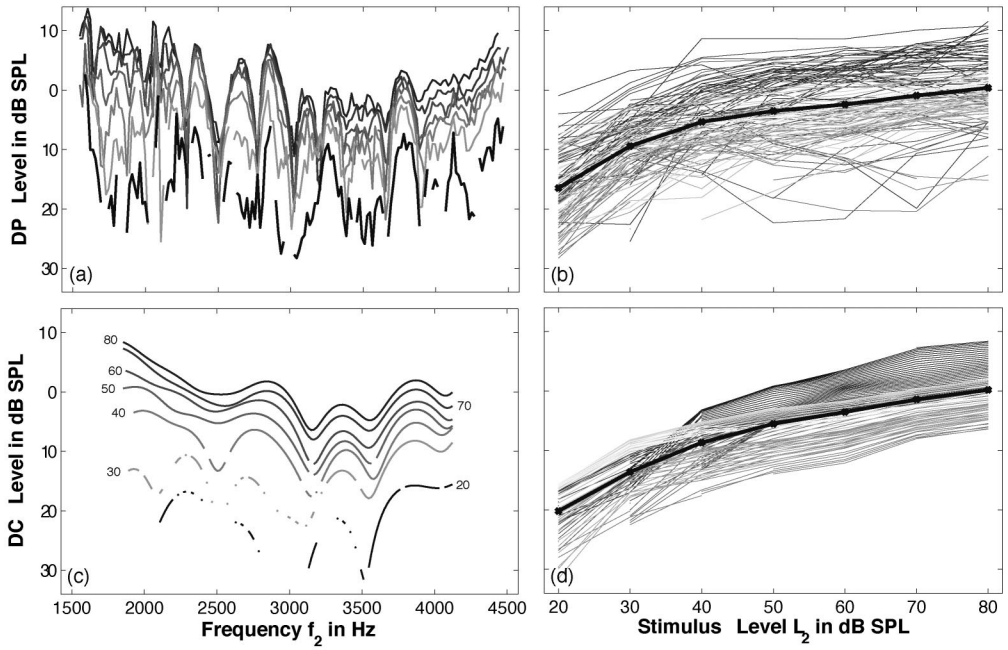


FIG. 4. (a) DP-Grams with a fine frequency resolution (18 Hz) for L_2 levels from 20 to 80 dB SPL (from bottom to top). (c) DC-Grams derived from the data shown in (a) by the method of time windowing. Here the related L_2 levels are indicated as numbers beside the corresponding curves. (b) DPOAE I/O and (d) DCOAE I/O functions corresponding to the DP-Gram data in (a) or DC-Gram data in (b), respectively. All data from subject BS. For a fair comparison the frequency range of the DPOAE I/O functions in (b) is restricted to the frequency range for which also DCOAE I/O functions are available, unaffected by window effects of the time windowing procedure. Only data points that fulfill a S/N criterion of 6 dB are shown here. The frequencies in the DPOAE and DCOAE I/O functions are coded from dark gray for low frequencies to light gray for higher frequencies. Notice the smooth change in shape and slope with frequency of the DCOAE I/O functions (d) and the very rapid changes in slope and shape of the DPOAE functions (b) indicating the bias of the second DPOAE source (see also Fig. 5). The thick black lines in (b) and (d) give the averages over all DPOAE, respectively DCOAE, I/O functions.

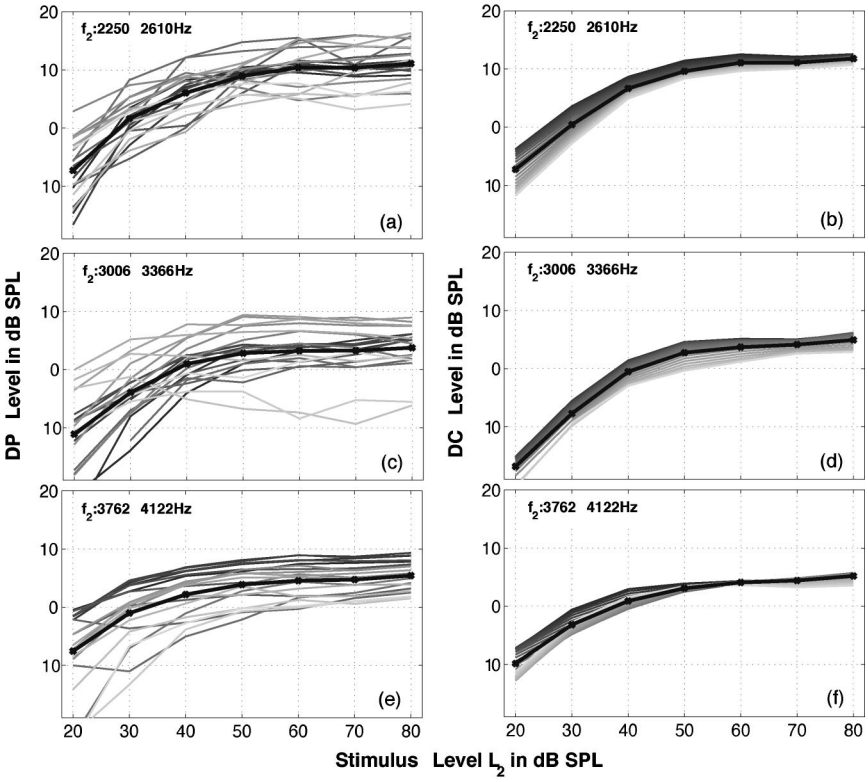


FIG. 5. Left panels (a,b,c) show DPOAE I/O functions from small frequency bands, (a) 2250–2610 Hz, (b) 3006–3366 Hz, and (c) 3762–4122 Hz. Right panels (b,d,f) show the corresponding DCOAE I/O functions. For each panel the coding of gray scales of the frequencies is analogous to the coding in Fig. 4. Notice the wide spread of DPOAE I/O functions in comparison to the clearly reduced variability of DCOAE I/O functions. All data from subject IT.

deviations of the DPOAE levels and respective DCOAE levels within each frequency band for each stimulus level (L_2 from 30 to 80 dB SPL). The mean of these standard deviations across all stimulus levels, as given in Table I, indicates the differences of the DPOAE I/O or DCOAE I/O functions, within each frequency band, i.e., the variability of adjacent I/O functions. Smaller mean standard deviations indicat less variability of the I/O functions. Overall these indicators are substantially smaller for DCOAE than for the DPOAE I/O functions for all subjects and frequency bands. Only the subjects AS and JF show a slightly lower standard deviation of the DPOAE level in two of the six frequency bands. But even in these cases, the use of DCOAE shows a clear improvement in terms of the logarithmic shape of I/O functions and a smooth frequency-dependent trend (not shown here). These comparisons clearly demonstrate that ''standard'' DPOAE I/O functions are corrupted considerably by the contribution from the second DPOAE source and that eliminating the influence of the second source on the I/O function, e.g., with the method employed here, yields much more reliable and plausible I/O functions.

C. Variability of threshold prediction from DPOAE I/O and DCOAE I/O functions

Figure 2 gives an example of the variability of threshold prediction from DPOAE I/O functions (L_{EDPT}) for one subject (IT). For three adjacent frequencies (2466, 2484, and 2503 Hz), the DCOAE I/O functions, and the prediction of hearing threshold L_{EDCT}, are almost the same ($L_{EDCT} \approx 14$ dB SPL, slope$=1.4$). On the other hand, the DPOAE I/O functions vary remarkably. This holds for the derived

TABLE I. Indicators for the variability of adjacent I/O functions within small frequency bands (see first column) for all subjects. The indicators (in dB) for DPOAE I/O functions (white columns) and DCOAE I/O functions (gray columns) are computed as the mean value across stimulus level of the standard deviations of DPOAE/DCOAE levels within each frequency band for each stimulus level (L_2 from 30 to 80 dB SPL). Smaller values indicating less variability of the I/O functions.

	Subject											
	AP		AS_{HL}		BS		IT		JF		MO	
Frequency band in Hz	Mean std. L_{DP}	Mean std. L_{DC}	Mean std. L_{DP}	Mean std. L_{DC}	Mean std. L_{DP}	Mean std. L_{DC}	Mean std. L_{DP}	Mean std. L_{DC}	Mean std. L_{DP}	Mean std. L_{DC}	Mean std. L_{DP}	Mean std. L_{DC}
1872–2232	6.048	0.889	2.817	1.563	3.983	1.531	3.406	0.876	4.025	1.221	5.436	2.317
2250–2610	4.207	1.403	4.660	3.610	6.207	1.558	3.320	0.992	4.415	2.173	4.420	1.568
2628–2988	4.644	1.144	3.290	2.502	4.598	1.172	2.819	1.347	2.376	1.195	2.626	1.669
3006–3366	3.947	1.688	3.710	1.262	4.461	2.320	4.204	1.452	2.279	0.499	3.727	0.949
3384–3744	3.917	0.654	2.175	2.759	3.404	1.999	3.195	1.690	1.776	1.230	2.364	1.206
3762–4122	1.991	0.389	1.860	3.731	2.807	0.963	3.392	0.682	1.645	2.121	2.145	1.059

TABLE II. Comparison of the standard deviations of the estimated threshold levels L_{EDPT} (white columns) and L_{EDCT} (gray columns) in dB SPL, computed across frequency, for the frequencies within six equally spaced, narrow frequency bands (1872–2232, 2250–2610, 2628–2988, 3006–3366, 3384–3744, and 3762–4122 Hz), respectively. Notice the strongly reduced standard deviations of the threshold predictions from DCOAE (L_{EDCT}) in comparison to the standard deviations for prediction from DPOAE (L_{EDPT}).

	Subject											
	AP		AS_{HL}		BS		IT		JF		MO	
Frequency band in Hz	std. L_{EDPT}	std. L_{EDCT}	std. L_{EDPT}	std. L_{EDCT}	std. L_{EDPT}	std. L_{EDCT}	std. L_{EDPT}	std. L_{EDCT}	std. L_{EDPT}	std. L_{EDCT}	std. L_{EDPT}	std. L_{EDCT}
1872–2232	10.90	3.70	2.37	2.01	7.45	4.91	4.48	3.20	—	5.72	8.79	3.53
2250–2610	8.61	2.71	4.09	3.95	6.17	4.38	7.08	2.96	13.29	7.67	6.29	1.83
2628–2988	7.98	2.55	6.67	2.21	5.83	4.91	4.28	2.10	8.64	3.87	4.35	2.79
3006–3366	9.99	2.58	5.63	2.08	16.56	7.04	7.63	2.21	5.91	2.17	7.08	3.18
3384–3744	7.71	2.10	2.58	2.09	8.53	2.75	10.30	6.11	4.27	2.39	5.26	2.28
3762–4122	3.13	0.75	2.44	2.32	8.58	3.42	16.30	5.25	7.12	2.79	5.07	1.86

prediction of threshold as well as the slope of the linear function fitted to the data. This strong fluctuation of threshold predictions from DPOAE I/O functions is quite typical for most subjects. Figure 6 illustrates that the threshold predictions L_{EDCT} from DCOAE I/O functions (thin black lines) change smoothly with frequency for all subjects, while the L_{EDPT} as function of frequency (gray thick lines) results in a strongly fluctuating pattern in most of the subjects. The same holds for the slope (see Fig. 7) of the fitted linear functions, which are also used as indicators of hearing status in several studies (e.g., Kummer *et al.*, 1998). To give quantitative indicators of the differences on variability of threshold estimations from DPOAE and DCOAE I/O functions, Table II lists the standard deviations of the estimated threshold levels L_{EDPT} and L_{EDCT} that are computed for six equally spaced, narrow frequency bands (1872–2232, 2250–2610, 2628–2988, 3006–3366, 3384–3744, and 3762–4122 Hz) for each subject.

The mean standard deviation of threshold predictions for all frequencies f_2 from 1872 to 4122 Hz, which are not rejected due to the fitting criteria, over all subjects and all frequency bands for the predictions from DCOAE I/O functions is about 3.3 dB while the mean standard deviation for the predictions from DPOAE I/O functions is 7.2 dB.

Furthermore, the number of I/O functions, which had to be excluded from further analysis due to the fitting criteria, was clearly higher for DPOAE I/O functions than for DCOAE I/O functions. Overall, 785 DPOAE and 785 DCOAE I/O functions were derived from the same data sets when frequencies affected by windowing effects of the time windowing procedure are eliminated. About 28.8% of the DPOAE I/O functions are rejected due to the fitting criteria, while only about 11.8% of the DCOAE I/O functions have to be rejected (see Table III). The slope criterion (i.e., the fitted slope *s* has to be 0.2 or higher to include the I/O function) does not appear to be critical for the current data set. The ''irregular shaped'' I/O functions would be rejected anyway because of other exclusion criteria, while a lot of ''regular shaped'' I/O functions which would give reasonable threshold predictions [see Figs. 7(c) and (e) and compare with Figs. 6(c) and (e)] would be discarded due to the slope criteria. Without the slope criteria, the number of rejected DCOAE I/O functions even decreases to about 4% (see Table III).

V. DISCUSSION

Since the discovery of otoacoustic emissions by Kemp in 1978, OAEs have been used to determine cochlea proper-

TABLE III. Numbers of DPOAE I/O functions which are rejected due to the several criteria used by Gorga *et al.* (2003). For DPOAE and DCOAE always the identical number of I/O functions at the same frequencies was investigated. No I/O function had been rejected due to the S/N criterion. The white columns show the numbers of rejected DPOAE I/O functions for each criterion (or all criteria). The gray columns show the numbers of rejected DCOAE I/O functions, respectively.

		Rejected I/O functions Criterion									
	No. of I/O	r^2		Std. error		slope *s*		All criteria		All criteria except, slope	
Subject	functions	DPOAE	DCOAE	DPOAE	DCOAE	DPOAE	DCOAE	DPOAE	DCOAE	DPOAE	DCOAE
AP	128	8	0	9	0	7	7	19	0	15	0
AS_{HL}	134	5	0	15	7	0	0	23	0	23	13
BS	127	31	0	1	0	36	27	46	37	31	0
IT	130	33	2	24	11	8	1	54	19	54	19
JF	138	45	0	5	0	49	34	73	34	45	0
MO	133	110	0	0	0	8	0	12	0	11	0
Total no.	790	133	2	141	4	108	62	190	67	144	6
%	100%	16.8%	0.3%	18%	4.4%	13.7%	7.9%	28.8%	11.8%	22.7%	4%

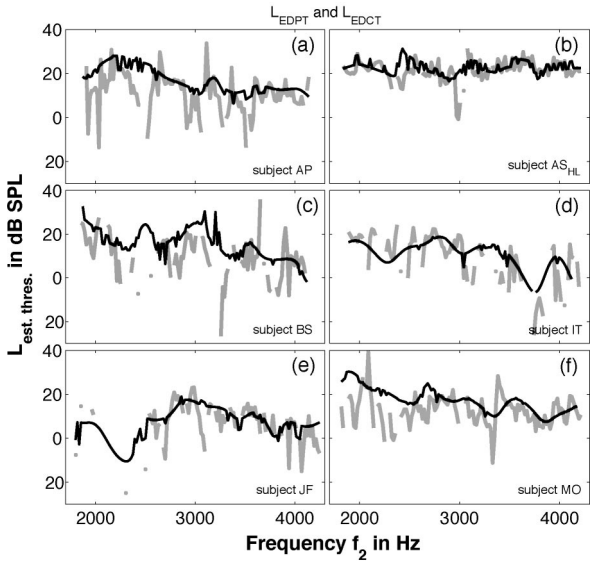


FIG. 6. Predicted thresholds from DPOAE I/O functions (gray lines) and DCOAE I/O functions (black lines) for all six subjects (the subjects in order of the panel indices: AP, AS_{HL}, BS, IT, JF, MO). Predictions from frequencies not fitting the inclusion criteria of stability index r^2 and standard error are left out in the plotted lines. Notice the reduced number of rejections and the smooth characteristics with frequency of the threshold predictions L_{EDCT} from DCOAE I/O functions in comparison to the threshold predictions L_{EDPT} from DPOAE I/O functions.

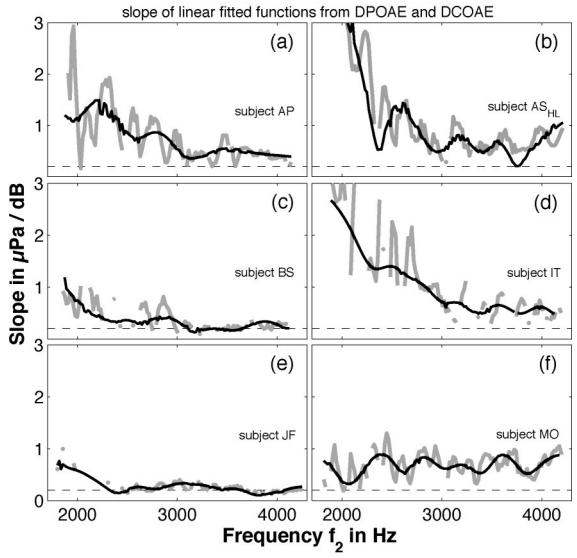


FIG. 7. Slopes of linear fitted functions from DPOAE I/O functions (gray lines) and DCOAE I/O functions (black lines) for all six subjects (subjects in order of the panel indices: AP, AS_{HL}, BS, IT, JF, MO). The slope criterion is indicated as a dashed line, i.e., frequencies at which the slope falls below 0.2 would be rejected due to the slope criterion.

ties, but also to clinically examine cochlear status. Nowadays, OAE are well established in hearing screening procedures, e.g., in neonates, as a qualitative prediction of whether a hearing loss may be present or not. From all kinds of OAE measurements, DPOAE appear to have the highest potential as a tool for a quantitative auditory test, because they are measurable even for hearing losses up to 50 dB. A very promising approach for a quantitative prediction of hearing loss from OAE is the use of DPOAE I/O function to characterize changes in cochlear nonlinearity or for the prediction of thresholds as suggested by Boege and Janssen (2002) and verified by Gorga *et al.* (2003). On average, the predictions of hearing threshold give good results but the variability in the data does not allow for a reliable prediction of auditory thresholds for the individual subject.

A. Influence of the second DPOAE source on DPOAE I/O functions and threshold predictions

Boege and Janssen (2002) discuss the close relationship between the nonlinear I/O characteristic of basilar membrane (BM) movement at the f_2 site from physiological studies and the growth of DPOAE.[13] The possible interferences with a second DPOAE source from a different cochlear site are not taken into account. Likewise, in many studies, DPOAE I/O functions are still seen as reflecting the cochlear status mainly around the f_2 place (e.g., Buus *et al.* 2001; Neely *et al.* 2003). On the other hand, the relevance of the second source on DPOAE at moderate to higher levels has been shown in several studies (Heitmann *et al.*, 1998; Talmadge *et al.*, 1999; Mauermann *et al.*, 1999a,b; Kaluri and Shera, 2001; Knight and Kemp, 2001) and is now widely accepted. The current study shows that the relative contribution from this second source increases with decreasing stimulation level (see Fig. 2) and so leads to increased uncertainty of DPOAE I/O characteristics as indicators of the f_2 site at low levels. These results are in agreement with results from Konrad-Martin *et al.* (2001). From DPOAE measurements with a fixed f_2 paradigm, they also found that the relative contribution from the second source is greater for low-level primaries in normal hearing ears. The strong influence of the second source on DPOAE I/O functions can be seen qualitatively in our data when one compares the large variation between adjacent frequencies for DPOAE I/O functions to the smaller variation in DCOAE I/O functions, for which the second source is excluded (see Figs. 3 and 4). Therefore, avoiding the influence of the second source on I/O functions (i.e., using DCOAE I/O functions instead of DPOAE I/O functions) is highly recommended in order to interpret the I/O functions from DPOAE measurements as reflecting the cochlear status near the characteristic site of f_2. This holds for determining BM compression characteristics from DPOAE I/O functions, e.g., for the use of DPOAE I/O functions as a objective recruitment indicator, as well as for the prediction of hearing thresholds derived from DPOAE I/O functions. Threshold predictions L_{EDCT} from DCOAE I/O functions show a strong reduction of variation between adjacent frequencies in comparison to threshold predictions L_{EDPT} from DPOAE I/O functions (see Fig. 6). The L_{EDCT} follows L_{EDPT} quite closely, similar to a moving average. Thus, the good overall correlation and small mean difference between estimated and pure tone threshold found in other studies using L_{EDPT} as threshold estimator (Boege and Janssen, 2002; Oswald *et al.*, 2002; Gorga *et al.*, 2003) is probably preserved when using L_{EDCT} instead of L_{EDPT} and the standard error of the correlation is expected to decrease due to the reduced variability of threshold predictions for adjacent frequencies. Since no auditory thresholds are measured in the current study, we cannot state whether the use of in-

terpolated DCOAE I/O functions result in a good threshold prediction or not. Maybe in general the changes in nonlinearity that are indicated by changes of distortion product I/O functions are not the only or even not the most important indicator of hearing loss from OAE measurements and, consequently, for threshold prediction. Mauermann *et al.* (1999b) showed, for example, that the reflection site or second source at f_{DP} appears to be even more sensitive to cochlear damages than the distortion site at f_2. However, at the current stage the interpolation of DPOAE I/O functions as suggested by Boege and Janssen (2002) appears to be one of the most promising approaches for threshold prediction from DPOAE measurements. Nevertheless, for reliable individual predictions this method still needs strong improvements. The findings of the current study strongly indicate that a clear improvement of the method can only be obtained by a separation of the DPOAE sources. Thus, under the assumption that in principle DPOAE I/O functions are an adequate tool to predict hearing threshold, further efforts should be invested into a reliable and practicable source separation before relevant improvements can be expected from other approaches like better-fitting algorithms.

In most studies investigating the correlation of hearing threshold and DPOAE measurements used hearing thresholds measured with sinusoids. Therefore, the fine structure of pure tone threshold is another factor that probably has strong influence on the correlation of auditory threshold and threshold estimations from DPOAE measurements. In some subjects, pure tone thresholds show a quasi-periodic fine structure with threshold variations of up to 15 dB (for review see, e.g., Mauermann *et al.*, 2004). The periodicity of DPOAE and pure tone threshold is similar but slightly different (Mauermann *et al.*, 1997a; Talmadge *et al.*, 1998), so that there are no direct correlations between the frequencies of DPOAE maxima and minima and threshold fine structure in humans.[14] The DPOAE fine structure is an interference phenomenon due to two DPOAE sources and multiple reflections within the cochlea, i.e., it does not simply reflect properties of cochlear status near f_2. Furthermore, the fine structure of the hearing threshold itself does not indicate frequency-specific differences of sensitivity, in the sense of reflecting local damages, rather than being the result of multiple reflections of the traveling wave between the characteristic site of the pure tone frequency and the oval window (e.g., Talmadge *et al.*, 1998). Thus, to obtain a meaningful comparison of DPOAE and auditory threshold around the frequency f_2 one has to eliminate both, the second DPOAE source, i.e., the fine structure of DPOAE (and fluctuations in DPOAE I/O functions and derived measures), and the fine structure of the hearing threshold. The fine structure of the hearing threshold can be easily removed by using stimuli with a bandwidth that is broader than the periodicity of threshold fine structure (Long and Tubis, 1988) instead of pure tones. Thus, using noises with a bandwidth of, e.g., $\frac{1}{3}$ oct would still provide a sufficient frequency resolution for auditory threshold measurements while avoiding effects of threshold fine structure. For the reduction of fine structure in DPOAE the second source has to be eliminated.

The removal of the second source, i.e., using DCOAE I/O functions instead of DPOAE I/O functions, shows, furthermore, a clear practical improvement for the method of threshold prediction as suggested by Boege and Janssen (2002) or Gorga *et al.* (2003). Thus, the number of I/O functions rejected for threshold estimation due to the fitting criteria is drastically reduced. This is relevant for clinical applications. Gorga *et al.* (2003) reported that ''18.4% of the total sample of DPOAE I/O functions failed to meet the inclusion criteria associated with the linear regression on DPOAE pressure (μPa) onto DPOAE stimulus level (dB SPL).'' They point out that misses due to the S/N criterion mostly are related to hearing loss, and so at least give qualitative and clinically relevant information. On the other hand, a DPOAE I/O function rejected due to a missed fitting criterion does not yield any information about hearing status. In another clinical evaluation, Oswald *et al.* (2002) reported a rejection of 46% or 26% of the I/O functions, depending on whether modified fitting criteria and procedures were used or not. Thus the reduction from 28.8% to about 11.35% or even 4% rejected I/O functions (without the slope criterion) in the current study contributes substantially to an improvement of the method.

B. Relevance of the second DPOAE source in hearing-impaired ears

In the current study, we investigated five normal-hearing subjects and one subject (AS) with a moderate high-frequency loss in detail. AS_{HL} shows slightly reduced effects due to the second source over the whole frequency range. However, the effect of the second DPOAE source in hearing-impaired subjects is expected to be very individual. An influence of the second source on DPOAE I/O functions can be expected whenever DPOAE fine structure occurs, i.e., if the second source contributes in the same order as the initial distortion source. He and Schmiedt (1996) investigated fine structure in normal- and hearing-impaired subjects. They stated that fine structure always occurs as long as DPOAE can be measured. Mauermann *et al.* (1999b) showed a high sensitivity of DPOAE fine structure in relation to the cochlear status around f_{DP}, i.e., DPOAE fine structure vanishes if f_{DP} falls in a region of hearing loss while the initial source near f_2 still produces a sufficient emission (e.g., for notches in the audiogram). However, the fine structure may even occur if f_{DP} falls in the region of normal hearing, while f_2 already covers a region of moderate hearing loss. This can be seen in subjects with a bandpass hearing-loss or a sloping high-frequency loss (Mauermann *et al.*, 1999b). Especially in those cases, a strong effect from the second source can be expected when using DPOAE I/O functions for the prediction of threshold. Konrad-Martin *et al.* (2002) investigated the relative contribution of the two DPOAE sources in hearing-impaired subjects using a fixed f_2 paradigm, with f_2 at 4 kHz. Looking at the latency domain to quantify the relative contribution of low- and high-latency components, they found the high latency components are even more prominent in sloping hearing losses than in normal ears. In contrast, rising losses showed only a low-latency peak. For certain types of hearing losses, the relative contribution of the second DPOAE source appears to be at least comparable

to normal ears. Overall, a contribution of the second DPOAE source has to be expected from even most hearing-impaired ears. Hence, the technique employed here to eliminate the influence of the second source appears to be very promising for the examination of hearing-impaired patients. However, the examination time for this method may be impractically high (see below).

C. Outlook

Overall, the current studies underline that the contribution from the second DPOAE source is not negligible, especially at low stimulus levels. DPOAE I/O functions are strongly affected by the interference of the two sources, and, therefore, DPOAE I/O functions can hardly be seen as an indicator reflecting the cochlear status of the f_2 site only. In addition to the theoretical advantage of using DCOAE I/O functions instead of DPOAE I/O functions for estimation of behavioral thresholds, relevant practical improvements such as a clearly reduced number of rejected I/O functions usable for prediction were found. Furthermore, strong fluctuations within threshold predictions from adjacent frequencies are almost eliminated for prediction from DCOAE as compared to DPOAE. This might lead to a relevant reduction of the standard errors of threshold predictions. To verify this improved prediction, further studies should be performed that compare the predictions from DCOAE I/O L_{EDCT} and behavioral thresholds. Unfortunately, the separation of the DPOAE components appears to be essential for a reliable use and interpretation of I/O functions from DPOAE measurements. A problem is the enormous measurement effort when using the method of time windowing for separation of the two DPOAE components. A significant reduction of the time effort using this procedure with classical DPOAE measurements is not possible because (1) a high-frequency resolution is needed to avoid aliasing in the time domain and (2) a sufficient frequency range is needed for this procedure to guarantee an adequate time resolution for the separation of low and high-latency components. Furthermore, a lot of frequencies at the edges have to be rejected because of the windowing effects. Thus, the method outlined here is definitely unsuitable for clinical application. A method that might be more practicable for the separation of the DPOAE components, even for clinical use, is the method of selective suppression (e.g., Heitmann *et al.*, 1998). A third tone is presented close to f_{DP} to suppress the second source while leaving the initial distortion component intact. Kaluri and Shera (2001) have shown that these two methods give similar results if an appropriate suppressor level is selected, which suppresses the second source while keeping the initial distortion source almost unaffected. However, the studies on selective suppression known by the authors only investigate a set of fixed moderate primary levels (Heitmann *et al.*, 1998; Siegel *et al.*, 1998; Kaluri and Shera, 2001). However, the optimal suppressor levels may vary when the primary levels are varied over a wide range. Therefore appropriate suppressor levels for the different primary levels have to be found to obtain before DCOAE I/O functions can be obtained using selective suppression. If adequate suppressor levels can be found in future studies, the measurement effort for DCOAE I/O functions would be almost the same as for ''standard'' DPOAE I/O functions.

Another approach to improve the time effort of DPOAE measurements could be the use of chirp stimuli (Mauermann *et al.*, 1998; Long *et al.*, 2004), either to improve the measurement time of DPOAE finestructure for later analysis with a time-windowing procedure or to separate concurrently the two DPOAE components by the usage of chirps with a very rapidly changing frequency (Long *et al.*, 2004). This method has to be demonstrated to give a sufficient S/N for low stimulation levels needed for I/O functions.

Whatever will be the method of choice for the measurement of DPOAE I/O characteristics, this study clearly shows that the two DPOAE components (DCOAE and RCOAE) have to be separated to obtain clear interpretable I/O functions and threshold estimations from DPOAE measurements.

VI. CONCLUSION

(i) The second DPOAE source strongly influences DPOAE I/O functions, especially at low levels.

(ii) The use of DCOAE I/O functions instead of DPOAE I/O functions avoids influences from a second source and so improves the interpretability of I/O functions to reflect the cochlear status at the f_2 site. Furthermore, it reduces the variability of I/O functions between adjacent frequencies.

(iii) Modifying the approach of Boege and Janssen (2002) by estimating behavioral thresholds from DCOAE I/O functions instead of DPOAE I/O functions shows great potential to improve the method. The number of rejected I/O functions and the variability of the predicted thresholds is strongly reduced.

(iv) The separation of the two DPOAE components using a time-windowing procedure is too time consuming for practical applications. A significant reduction of the time effort is not possible due to the required frequency resolution.

(v) In order to use DCOAE I/O functions for studies in clinical routine tests, a less time consuming paradigm for separation of DPOAE components is needed, e.g., adequate suppressor levels for the method of selective suppression have to be found for a sufficient range of primary levels.

ACKNOWLEDGMENTS

This study was supported by Deutsche Forschungsgemeinschaft, DFG Ko 942/11-3. Helpful comments by the two anonymous reviewers and Stefan Uppenkamp on an earlier version of the manuscript are gratefully acknowledged. We would like to thank Glenis R. Long and Monica Wagner for improving the readability of the manuscript.

[1]Although auditory thresholds show a similar quasi-periodic-fine structure as DPOAE fine structure but the periodicity is slightly different and there is no direct correspondence between the position of maxima and minima of both (e.g., Mauermann *et al.*, 1997a; Talmadge *et al.*, 1998).

[2]This is equivalent to a convolution of the complex DP-Gram with the Fourier transform of the time window and results in a smoothing of the

DP-Gram. Therefore some authors use the term ''spectral smoothing'' beside ''time windowing'' (e.g., Kaluri and Shera, 2001).

[3]All measured values for DPOAE amplitude and phase were used for the time windowing analysis, even those with a S/N poorer than 6 dB. However, the DPOAE and DCOAE I/O functions were derived only from data points that fulfill a S/N criterion of 6 dB (see Fig. 4).

[4]Kaluri and Shera (2001) performed the ''spectral smoothing'' as a convolution on a spectrum ''wrapped'' around a cylinder, i.e., taking no negative frequencies into account. This is analogous to the transformation of Knight and Kemp (2001) of only positive frequencies into the time domain. In the current study the DP-Gram data was extended to negative frequencies by the complex conjugated values of the spectral data from the positive frequencies. Thus the IFFT results in a real signal as ''time-domain response'' that is similar to an impulse response. Both ways of the analysis allow a separation of components with different group delay and a reconstruction of the spectral data from the time domain representation.

[5]To avoid any discontinuity of the spectrum boundaries, and hence to minimize artifacts in the IFFT (or, if thinking in terms of spectral smoothing, to guarantee a smooth function to be convolved with the Fourier transform of the time window) for the bin at zero frequency, a real value was used that was calculated from an interpolation between the positive and negative frequency values. This value at zero frequency results in a dc component which is subtracted from the corresponding ''time-response.''

[6]SFOAE and transient-evoked otoacoustic emissions show a variation in group delay with frequency (e.g., Kemp, 1978; Wilson, 1980; Neely *et al.*, 1988). Therefore it is assumed that the high latency components show a varying latency, i.e., a spread of peaks in the latency domain. Therefore other authors (Knight and Kemp, 2001; Kalluri and Shera, 2001) suggest a transformation of the data points before the IFFT to be equidistant on an exponential frequency axis. Their aim is to linearize the underlying curve of phase as a function of frequency for the reflection component. This results in clearer peaks in the time domain (Zweig and Shera, 1995). However, a comparison of both—analysis with linear and exponential frequency axis—showed no practical advantage of the transformation for the further analysis. In most cases it was even easier to distinguish between low and high latency components with the ''spread'' of peaks for the high latency components.

[7]If the definition of such a point appears to be critical, we compared the time signals with data from measurements from a parallel study using a third tone close to $2f_1-f_2$ to suppress the second source. Peaks in the latency domain which were easily affected even from low suppressor levels were seen as high latency components while peaks being almost unaffected were assumed to be low latency components. This allows an almost complete assignment of the peaks belonging either to the high or low to the latency component. Suppression data for comparison is available at a couple of primary levels for all subjects in the current study except for subject AS.

[8]The plots of the ''time-domain response'' were scaled such that they represent the sound pressure at each time per frequency component, e.g., a white spectrum with amplitudes 1 at positive and negative frequencies with zero phases would result in a time domain peak with amplitude 1.

[9]Boege and Janssen computed a linear regression of L_2 onto the measured p_{DP} with the aim to minimize the prediction error for L_{EDPT}. Therefore, they use an inclusion criterion for the standard error to be less than 10 dB (Boege, 2003), whereas Gorga *et al.* (2003) computed a linear regression of p_{DP} onto the independent variable L_2. As inclusion criterion here the standard error has to be less than 10 μPa. The inclusion criterion for the standard error to be less than 10 dB given in Gorga *et al.* (2003, Sec. II E) is a typographical error and should be 10 μPa (Neely, 2003). In the current study we followed the criteria given by Gorga *et al.* (2003).

[10]In Boege and Janssen (2002), Eq. (4), the criterion $s \geqslant 0.1$ is given. However, this appeared to be a typographical error and the actual slope criterion was 0.2 μPa/dB as mentioned in Gorga *et al.* (2003)—see first endnote there.

[11]The DPOAE level at $L_2 = 80$ dB SPL often showed even a decrease in DPOAE level compared to the DPOAE level for $L_2 = 70$ dB SPL.

[12]For both the low and high latency parts of the signal, the rms value is related to the length of the whole time domain response.

[13]For the level paradigm $L_1 = 0.4L_2 + 39$ dB SPL.

[14]While the fine structure of pure tone threshold correlates closely with the periodicity of SOAE (Zweig and Shera, 1995; Shera, 2003) or the fine structure of low-level-transient-evoked emissions, it does not correlate with DPOAE fine structure (Talmadge *et al.*, 1998) to which the prediction of threshold from ''standard'' DPOAE I/O functions may be related.

Boege, P. (**2003**). Personal communication.

Boege, P., and Janssen, T. (**2002**). ''Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O-functions in normal and cochlear hearing loss ears,'' J. Acoust. Soc. Am. **111**, 1810–1818.

Brown, A. M., Harris, F. P., and Beveridge, H. A. (**1996**). ''Two sources of acoustic distortion products from the human cochlea,'' J. Acoust. Soc. Am. **100**, 3260–3267.

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*, Proceedings of the 12th International Symposium on Hearing, Mierlo August 2000, edited by D. J. Breebart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven, pp. 373–381.

Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (**2002**). ''Multiple internal reflections in the cochlea and their effect on DPOAE fine structure,'' J. Acoust. Soc. Am. **112**, 2882–2897.

Gaskill, S. A., and Brown, A. M. (**1996**). ''Suppression of human acoustic distortion product: dual origin of $2f_1-f_2$,'' J. Acoust. Soc. Am. **100**, 3268–3274.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Hoover, B. M. (**2003**). ''Further efforts to predict pure-tone thresholds from distortion product otoacoustic emission input/output functions,'' J. Acoust. Soc. Am. **113**, 3275–3284.

Gorga, M. P., Neely, S. T., Bergman, B., Beauchaine, K. L., Kaminski, J. R., Peters, J., and Jestaedt, W. (**1993**). ''Otoacoustic emissions from normal-hearing and hearing-impaired subjects: Distortion product responses,'' J. Acoust. Soc. Am. **93**, 2050–2060.

Harris, F. P., and Probst, R. (**2002**). ''Otoacoustic Emissions and Audiometric Outcomes,'' in *Otoacoustic Emissions—Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 416–438.

He, N., and Schmiedt, R. A. (**1993**). ''Fine structure of the $2f_1-f_2$ acoustic distortion product: changes with primary levels,'' J. Acoust. Soc. Am. **94**, 2659–2669.

He, N. J., and Schmiedt, R. A. (**1996**). ''Effects of aging on the fine structure of the $2f_1-f_2$ acoustic distortion product,'' J. Acoust. Soc. Am. **99**, 1002–1015.

Heitmann, J., Waldmann, B., Schnitzler, H. U., Plinkert, P. K., and Zenner, H. P. (**1998**). ''Suppression of distortion product otoacoustic emissions (DPOAE) near $2f_1-f_2$ removes DP-gram fine structure—Evidence for a secondary generator,'' J. Acoust. Soc. Am. **103**, 1527–1531.

Kalluri, R., and Shera, C. A. (**2001**). ''Distortion-product source unmixing: a test of the two-mechanism model for DPOAE generation,'' J. Acoust. Soc. Am. **109**, 622–637.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Knight, R. D., and Kemp, D. T. (**2001**). ''Wave and place fixed DPOAE maps of the human ear,'' J. Acoust. Soc. Am. **109**, 1513–1525.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). ''Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears,'' J. Acoust. Soc. Am. **109**, 2862–2879.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., Cyr, E., and Gorga, M. P. (**2002**). ''Sources of DPOAEs revealed by suppression experiments, inverse fast Fourier transforms, and SFOAEs in impaired ears,'' J. Acoust. Soc. Am. **111**, 1800–1809.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). ''The level and growth behavior of the $2f_1-f_2$ distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss,'' J. Acoust. Soc. Am. **103**, 3431–3444.

Long, G. R., and Tubis, A. (**1988**). ''Investigations into the nature of the association between threshold microstructure and otoacoustic emissions,'' Hear. Res. **36**, 125–138.

Long, G. R., Talmadge, C. L., and Lee, J. (**2004**). ''Using sweeping tones to evaluate DPOAE fine structure,'' Assoc. Res. Otolaryngol. Abs. **27**, 34(A).

Mauermann, M., Long, G. R., and Kollmeier, B. (**2004**). ''Fine structure of Hearing-Threshold and Loudness Perception,'' J. Acoust. Soc. Am. (submitted).

Mauermann, M., Uppenkamp, S., and Kollmeier, B. (**1997a**). ''Zusammenhang zwischen unterschiedlichen otoakustischen Emissionen und deren Relation zur Ruhehörschwelle [The relation of different types of otoacoustic emissions and the threshold in quiet],'' in *Fortschritte der Akustik—DAGA 97*, edited by P. Wille (DEGA e.V., Oldenburg), pp. 242–243.

Mauermann, M., Uppenkamp, S., and Kollmeier, B. (**1997b**). ''Periodizität

und Pegelabhängigkeit der spektralen Feinstruktur von Verzerrungsprodukt-Emissionen [Periodicity and dependence on level of the distortion product otoacoustic emission spectral fine-structure],'' Audiol. Akustik **36**, 92–104.

Mauermann, M., Uppenkamp, S., and Kollmeier, B. (**1998**). ''Messung von Verzerrungsprodukt-Emissionen mit Chirp-Signalen [Measurement of distortion product emissions using chirp signals],'' in *Fortschritte der Akustik—DAGA 1998*, edited by A. Sill (DEGA eV, Oldenburg), pp. 316–317.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999a**). ''Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio f_2/f_1,'' J. Acoust. Soc. Am. **106**, 3473–3483.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999b**). ''Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss,'' J. Acoust. Soc. Am. **106**, 3484–3491.

Moulin, A., Bera, J. C., and Collet, L. (**1994**). ''Distortion product otoacoustic emissions and sensorineural hearing loss,'' Audiology **33**, 305–326.

Neely, S. (**2003**). Personal communication.

Neely, S. T., Gorga, M. P., and Dorn, P. A. (**2003**). ''Cochlear compression estimates from measurements of distortion-product otoacoustic emissions,'' J. Acoust. Soc. Am. **114**, 1499–1507.

Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). ''Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli,'' J. Acoust. Soc. Am. **83**, 652–656.

Nelson, D. A., and Kimberley, B. P. (**1992**). ''Distortion product emissions and auditory sensitivity in human ears with normal hearing and cochlear hearing loss,'' J. Speech Hear. Res. **35**, 1142–1159.

Oswald, J. A., Müller, J., and Janssen, T. (**2002**). ''Audiometric threshold estimation in cochlear hearing loss ears by means of weighted extrapolated DPOAE I/O functions,'' Assoc. Res. Otolaryngol. Abs. **21**, #774(A).

Shera, C. A. (**2003**). ''Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves,'' J. Acoust. Soc. Am. **114**, 244–262.

Shera, C. A., and Guinan, Jr., J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.

Siegel, J. H. (**2002**). ''Calibrating Otoacoustic Emission Probes,'' in *Otoacoustic Emissions—Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 416–438.

Siegel, J. H., Dreisbach, L. E., Neely, S. T., and Spear, W. H. (**1998**). ''Vector decomposition of distortion-product otoacoustic emission sources in humans,'' Assoc. Res. Otolaryngol. Abs. **21**, 347(A).

Suckfüll, M., Schneeweiß, S., Dreher, A., and Schorn, K. (**1996**). ''Evaluation of TEOAE and DPOAE measurements for the assessment of auditory thresholds in sensorineural hearing loss,'' Acta Otolaryngol. (Stockh) **116**, 528–533.

Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). ''Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **105**, 275–292.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). ''Modeling otoacoustic emission and hearing threshold fine structure,'' J. Acoust. Soc. Am. **104**, 1517–1543.

Wilson, J. P. (**1980**). ''Evidence for a cochlear origin for acoustic reemissions, threshold fine-structure and tonal tinnitus,'' Hear. Res. **2**, 233–252.

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.

Searching for the optimal stimulus eliciting auditory brainstem responses in humans

Oliver Fobel[a)]
Medizinische Physik, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany

Torsten Dau[b)]
Centre for Applied Hearing Research, Ørsted-DTU, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

(Received 8 August 2003; revised 29 June 2004; accepted 6 July 2004)

This study examines auditory brainstem responses (ABR) elicited by rising frequency chirps. Two chirp stimuli were developed and designed such as to compensate for cochlear travel-time differences across frequency, in order to maximize neural synchrony. One chirp, referred to as the O-chirp, was based on estimates of human basilar membrane (BM) group delays derived from stimulus-frequency otoacoustic emissions (SFOAE) at a sound pressure level of 40 dB [Shera and Guinan, in *Recent Developments in Auditory Mechanics* (2000)]. The other chirp, referred to as the A-chirp, was derived from latency functions fitted to tone-burst-evoked ABR wave-V data over a wide range of stimulus levels and frequencies [Neely *et al.*, J. Acoust. Soc. Am. **83**(2), 652–656 (1988)]. In this case, a set of level-dependent chirps was generated. The chirp-evoked responses, particularly wave-V amplitude and latency, were compared to click responses and to responses obtained with the original chirp as defined in Dau *et al.* [J. Acoust. Soc. Am. **107**(3), 1530–1540 (2000)], referred to here as the M-chirp since it is based on a (linear) cochlea model. The main hypothesis was that, at low and medium stimulation levels, the O- and A-chirps might produce a larger response than the original M-chirp whose parameters were essentially derived from high-level BM data. The main results of the present study are as follows: (i) All chirps evoked a larger wave-V amplitude than the click stimulus indicating that for the chirps a broader range of spectral components contributes effectively to the ABR. (ii) Only small differences were found between the O-chirp and M-chirp responses at low and medium levels. This indicates that SFOAE may not provide a robust estimate of BM group delay, particularly at low frequencies, or that frequency-dependent neural delays exist which are not reflected in the design of these chirps. (iii) The A-chirp produced the largest responses, particularly at low stimulation levels. This chirp might therefore be valuable for clinical applications, particularly in tests where the click stimulus has been used so far.
[DOI: 10.1121/1.1787523] [S0001-4966(94)70510-5]

PACS numbers: 43.64.Qh, 43.64.Ri, 43.64.Bt [WPS] Pages: 2213–2222

I. INTRODUCTION

Transient stimuli like clicks are commonly used in electrophysiological research of the human auditory system to elicit synchronized auditory brainstem responses (ABR). The click is the most common stimulus used in recording the ABR, whether for neurodiagnostic or audiologic purposes. However, in the cochlea, the response to a click is not entirely synchronous: The peak of the response occurs several milliseconds later in low-frequency channels than it does in high-frequency channels (e.g., von Békésy, 1960; Kiang *et al.*, 1965; Kiang, 1975). This is mainly a result of the change of stiffness along the cochlear partition. As a consequence, the phase velocity of the traveling wave depends in a characteristic way upon location, which causes spatial dispersion. It takes more time for the low-frequency region to reach maximal displacement at the apical end of the cochlea. Electrophysiological responses to broadband transients like clicks appear to be largely generated by the synchronized activity of the high-frequency channels on their own. For example, Don and Eggermont (1978) measured human ABR in response to clicks masked by high-pass noise with different cut-off frequencies. This masking technique revealed that the latencies in response to low-frequency stimuli are delayed relative to high frequencies. Don and Eggermont concluded from their data that all frequency regions contribute to the ABR but that the response is dominated by contributions from the 2–3 octaves towards the basal end.

In a later study, Don *et al.* (1994) developed a technique of ''normalizing'' click-evoked ABR using high-pass noise masking. They adjusted for differences in the neural conduction time (wave *I-V* delay) through compression or expansion, and for differences in the cochlear response time through shifts of the derived ''narrowband'' ABR patterns. A summation of the compressed and time shifted responses resulted in the so-called ''stacked ABR.'' Their study demonstrated the effect of temporal delays of cochlear activation on

[a)]Present address: Centre for Applied Hearing Research, Ørsted·DTU, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark; electronic mail: of@oersted.dtu.dk
[b)]Electronic mail: tda@oersted.dtu.dk

the component amplitude of the compound ABR. In a later study, Don *et al.* (1997) showed that this technique is particularly useful for the detection of small intracanalicular acoustic tumors.

A different approach was used by Dau *et al.* (2000) and Wegner and Dau (2002) to study the effects of cochlear dispersion on the formation of ABR. They demonstrated that upward chirps can strongly affect ABR wave V. Their chirp was designed to produce simultaneous displacement maxima along the cochlear partition by compensating for frequency-dependent traveling-time differences. Their equations determining the temporal course of the chirp were derived on the basis of a linear cochlea model (de Boer, 1980) and were calculated to be the inverse of the delay-line characteristic of the human cochlear partition. The fundamental relationship between stimulus frequency and place of maximum displacement was taken from Greenwood (1990). ABR evoked by the broadband chirp showed a larger wave-V amplitude than corresponding click-evoked responses. Dau *et al.* (2000) demonstrated that the ABR is not an electrophysiological event purely evoked by onset or offset of an acoustic stimulus, but that an appropriate temporal organization, determined in part by basilar-membrane (BM) traveling-wave properties, can significantly increase synchrony of neural discharges. Wegner and Dau (2002) demonstrated that such a chirp is also very useful for retrieving frequency-specific information, particularly at low frequencies.

The model of de Boer (1980) is based upon the experimental observations of von Békésy (1960). Von Békésy's measurements were performed with the aid of a microscope to detect and measure the movements of cochlear structures. The movements had to be much larger than under the influence of ''natural'' sounds. In fact, von Békésy used very high sound pressure levels (SPL), of the order of 120 to 140 dB. In later studies, cochlear vibration patterns were measured with more sensitive techniques and under more natural circumstances. These studies of cochlear mechanics have established that the response of the BM to tones at characteristic frequency (CF) is generally nonlinear and compressive while it responds essentially linearly to tones with a frequency well below CF (e.g., Rhode, 1971; Sellick *et al.*, 1982; Robles *et al.*, 1986; Ruggero *et al.*, 1997; Rhode and Recio, 2000). At low stimulus levels, if plotted as isointensity curves as a function of stimulus frequency, the BM response patterns are sharply frequency tuned around the CF. Recent psychophysical studies suggest that the BM is also the primary source of the nonlinearities observed in masking experiments (e.g., Oxenham and Plack, 1998; Plack and Oxenham, 1998) and that cochlear nonlinearities have a significant influence on a wide range of basic auditory processes, such as frequency selectivity (e.g., Moore *et al.*, 1999), temporal integration (Oxenham *et al.*, 1997), and loudness growth (e.g., Yates *et al.*, 1990; Moore and Glasberg, 1997). Damage to the cochlea, and in particular the outer hair cells, results in a reduction in sensitivity and a loss of compression at CF (Ruggero and Rich, 1991; Ruggero *et al.*, 1995). In such a situation, the pattern of BM vibration is similar to the ''insensitive'' response, originally found by von Békésy.

Recently, Shera and Guinan (2000, 2003) introduced a method to measure BM group delay in the human auditory system. They measured stimulus-frequency otoacoustic emissions (SFOAE) for a stimulus level of 40 dB SPL, and derived the SFOAE group delay, τ_{SFOAE}, from the slope of the SFOAE-phase versus frequency function. According to the theory of coherent reflection filtering (Shera and Zweig, 1993; Zweig and Shera, 1995) the BM group delay τ_{BM} is given by $\tau_{\mathrm{BM}}=0.5\times\tau_{\mathrm{SFOAE}}$. It is assumed that, at low stimulus levels, the BM transfer functions can be described by minimum-phase-shift filters (Zweig, 1976). Shera *et al.* (2002) therefore directly related τ_{BM} to cochlear-filter bandwidth. They compared the results to psychophysical estimates of auditory filter bandwidth obtained using notched-noise forward masking experiments (Shera *et al.*, 2002; Oxenham and Shera, 2003). The estimates of frequency selectivity obtained in the different approaches were in good agreement. These estimates of cochlear frequency selectivity at low levels, obtained with the method of SFOAEs, will most likely differ from those values assumed in de Boer's (1980) model. Thus, a corresponding chirp stimulus that would be designed to compensate for frequency-dependent traveling-time differences, can be expected to differ in its waveform from the chirp developed by Dau *et al.* (2000).

Another approach to compensate for delays across frequency would be to base the chirp parameters on wave-V latency values obtained in tone-burst-evoked ABR data. Gorga *et al.* (1988) measured tone-burst-evoked ABRs over a wide range of stimulus levels and frequencies. Their wave-V-latency data were described by Neely *et al.* (1988) by the following power-law relation:

$$\tau_b = a + bc^{-i}f^{-d}, \tag{1}$$

where i represents tone-burst intensity, f indicates tone-burst frequency, and a, b, c, and d are constants [cf. their Eq. (1)]. Neely *et al.* (1988) assumed that the first term, parameter a, represents the frequency and level-independent neural component of the latency while the second term in Eq. (1) reflects the mechanical component of the latency due to the propagation in the cochlea thus representing BM group delay. By comparing the ABR data from Gorga *et al.* (1988) with tone-burst otoacoustic emission(OAE) data from Norton and Neely (1987)—who used a subgroup of the subjects from Gorga *et al.* (1988)—Neely *et al.* (1988) found a much larger inter- and intraindividual variability in the OAE data than in the ABR data which might suggest that BM group delay can be better estimated with ABR than with OAE.

The current study deals with the development and test of chirp stimuli in an attempt to find an ideal stimulus eliciting ABR in humans. One chirp is generated based on the SFOAE data by Shera and Guinan (2000). The second chirp is generated on the basis of ABR wave-V latency data by Neely *et al.* (1988). Corresponding chirp-evoked responses are compared at various stimulus levels with results obtained with the original model-based chirp by Dau *et al.* (2000), and with conventional click data. The underlying hypothesis is that, at low stimulation levels, the new chirps might produce a better synchronization than the original chirp since the latter one was derived on the basis of high-level BM data. The

results of the present study could be valuable both for a better understanding of basic mechanisms underlying the generation of ABR and for clinical applications using chirp-evoked ABR as an objective indicator of hearing threshold.

II. CHIRP STIMULI

A. OAE-based chirp stimulus (O-chirp)

The first chirp stimulus is based on the experimental SFOAE data of Shera and Guinan (2000). They did experiments for stimulus frequencies in the range from 0.5 to 10 kHz in humans, at a level of 40 dB SPL. Emission group delays, τ_{SFOAE}, were calculated and related to BM group delays, $\tau_{\mathrm{BM}}=0.5\times\tau_{\mathrm{SFOAE}}$, as a function of CF. The data can be roughly approximated by the following function (Shera, personal communication):

$$\tau_{\mathrm{BM,O}}(f):=t(f)=c(f/[\mathrm{Hz}])^{\alpha}, \tag{2}$$

with the constants $c=0.15$ s and $\alpha=-0.5$.[1] $\tau_{\mathrm{BM,O}}$ can also be interpreted as reflecting the propagation time $t(f)$ needed to arrive at the place of resonance for frequency f. In order to compensate for dispersion across frequency, we introduced the variable transformation $t\rightarrow t_0-t$, with $t_0=\tau_{\mathrm{BM,O}}(50\ \mathrm{Hz})$ and calculated the following inverse function $f_{\mathrm{O}}(t)=t^{-1}(f)$:[2]

$$f_{\mathrm{O}}(t)=\left(\frac{c}{t_0-t}\right)^2. \tag{3}$$

This function, reflecting the change of the instantaneous frequency with time was then integrated over time to derive the instantaneous phase $\varphi_{\mathrm{O}}(t)$ of the resulting chirp

$$\varphi_{\mathrm{O}}(t)=2\pi\int_0^t f_{\mathrm{O}}(t')dt' \tag{4}$$

$$=2\pi c^2\left(\frac{1}{t_0-t}-\frac{1}{t_0}\right). \tag{5}$$

The chirp stimulus is then given by

$$s_{\mathrm{O}}(t)=A_{\mathrm{O}}(t)\sin[\varphi_{\mathrm{O}}(t)-\varphi_0], \tag{6}$$

whereby φ_0 determines the starting phase of the chirp. The amplitude factor $A_{\mathrm{O}}(t)$ was chosen as

$$A_{\mathrm{O}}(t)=\sqrt{\frac{df_{\mathrm{O}}(t)}{dt}}=\sqrt{\frac{2c^2}{(t_0-t)^3}}, \tag{7}$$

in order to produce a stimulus with a flat magnitude spectrum. Since the stimulus $s_{\mathrm{O}}(t)$ is based on OAE data, it is referred to as the "O-chirp" throughout the present paper.

B. ABR-based chirp stimulus (A-chirp)

The second chirp stimulus developed in this study is based on the tone-burst-evoked ABR data by Gorga *et al.* (1988). They used tone bursts at ten frequencies (0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, and 8 kHz) and nine intensities (20 to 100 dB SPL in 10-dB steps). Their data were described by Neely *et al.* (1988) by a power law [cf. Eq. (1)], relating latency to stimulus frequency, assuming that the total wave-V latency represents the sum of mechanical and neural components. Neely *et al.* (1988) assumed the neural component to be independent of frequency and level. The second term in their Eq. (1) was assumed to represent the component of ABR latency due to the mechanical propagation within the cochlea. Hence they estimated the BM group delay by

$$\tau_{\mathrm{BM,A}}(i,f)=bc^{-i}f^{-d}, \tag{8}$$

where i represents the tone-burst intensity (in dB SPL divided by 100), f represents tone-burst frequency (divided by 1 kHz), and b, c, and d are constants with the values $b=12.9$ ms, $c=5.0$, and $d=0.413$, according to the data fit from Neely *et al.* (1988). Whether this term actually represents only the mechanical component of the observed ABR latency, as assumed by Neely *et al.* (1988), or inherently also some frequency and/or level-dependent neural/synaptic contribution, has not been resolved yet. In any case, Eq. (8) represents the frequency and level-dependent part of wave-V latency. From this, the inverse function and the function for the instantaneous phase $\varphi_{\mathrm{A}}(i,t)$ can be calculated as follows:[3]

$$\varphi_{\mathrm{A}}(i,t)=\frac{2\pi(bc^{-i})^{1/d}}{\frac{1}{d}-1}\left[\frac{1}{(t_0(i)-t)^{1/d-1}}-\frac{1}{t_0(i)^{1/d-1}}\right]. \tag{9}$$

The chirp stimulus is then given by

$$s_{\mathrm{A}}(i,t)=A_{\mathrm{A}}(i,t)\sin[\varphi_{\mathrm{A}}(i,t)-\varphi_0], \tag{10}$$

whereby φ_0 defines the starting phase of the chirp. The amplitude factor $A_{\mathrm{A}}(i,t)$ was chosen as

$$A_{\mathrm{A}}(i,t)=\sqrt{\frac{(bc^{-i})^{1/d}}{d[t_0(i)-t]^{1/d+1}}} \tag{11}$$

in order to produce a stimulus with flat magnitude spectrum. Throughout the current study, the stimulus $s_{\mathrm{A}}(i,t)$ will be referred to as the "A-chirp," since it is based on ABR data.

C. Comparison of the different chirp stimuli

The O- and A-chirps are compared to the original chirp defined in Dau *et al.* (2000) that was based on de Boer's (1980) linear cochlea model. For direct comparison, the realization of the chirp with a flat magnitude spectrum is used that was also developed in Dau *et al.* (2000) and denoted as the "flat-spectrum chirp." Since this chirp is based on a model, it is referred to as the "M-chirp" in the following.

Within de Boer's (1980) model, the propagation time, $\tau_{\mathrm{BM,M}}(f)$, needed to arrive at the place of resonance for the frequency f, is approximately given as

$$\tau_{\mathrm{BM,M}}(f)\propto(f+165.4\ \ \mathrm{Hz})^{-1.1}, \tag{12}$$

which clearly differs from Eqs. (2) and (8) representing the corresponding functions for the two other chirps. The left panel of Fig. 1 illustrates the calculated BM group delays on the basis of Eqs. (2), (8), and (12). The group delays derived from the SFOAE experiments (solid curve) are about 2–5 ms larger than those predicted on the basis of de Boer's model (dashed curve). The shaded dots represent the original BM group delay estimates of the SFOAE data by Shera and Guinan (2000). The shift of the SFOAE-based estimates to-

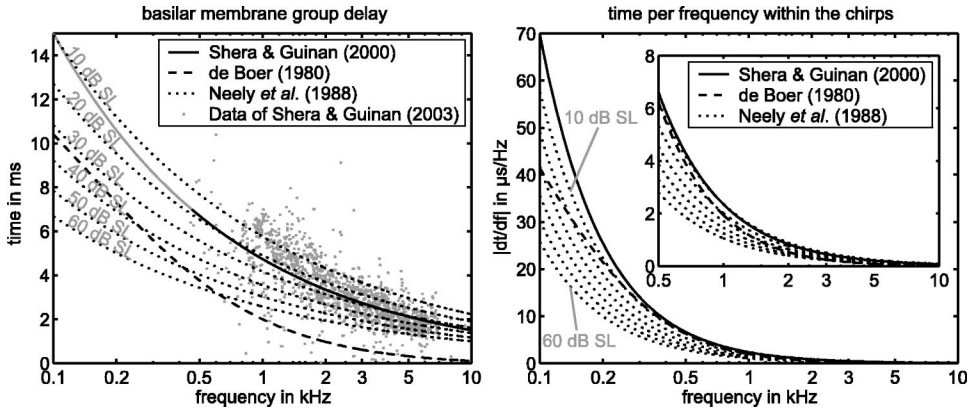


FIG. 1. Left: BM group delay as a function of frequency. The shaded dots represent the original BM group delay estimates of Shera and Guinan (2003), derived from SFOAE data. The black solid line represents the data fit of Shera and Guinan (2000). This function is interpolated toward the lower frequencies (gray solid line). The dashed line represents the group delay on the basis of the linear cochlea model by de Boer (1980). The dotted lines indicate the group delays predicted by Neely *et al.* (1988) for stimulus intensities from 10 dB SL (top dotted curve) up to 60 dB SL (bottom dotted curve). Right: Time-per-frequency change for the different chirps. This was directly calculated from the data in the left panel. The inset is a replot of the data for the frequency range from 0.5 to 10 kHz, using a rescaled ordinate.

ward higher values is reasonable, at least qualitatively, since frequency selectivity is increased at lower levels and this should be accompanied by a larger group delay at all frequencies. The group delay estimates predicted by the equation of Neely *et al.* (1988) depend on frequency and intensity (dotted curves). At medium levels, the delay is between about 1 ms (at 10 kHz) and about 8 ms (at 0.1 kHz).

For the generation of the chirps, *not the absolute* values but the *change* of group delay with frequency is important. Thus, the derivative of the group-delay versus frequency function is shown in the right panel of Fig. 1. The main differences between the SFOAE-based curve and the model-based prediction occur at low frequencies, but differences are also present at frequencies above 500 Hz. This means that the instantaneous frequency of the corresponding O-chirp will vary more slowly than that of the M-chirp, particularly at low frequencies. The corresponding ABR-based group-delay functions clearly differ from the other ones. There is a large variation with stimulus level. For example, at higher levels, the resulting A-chirp will vary much faster in instantaneous frequency over time than in the cases of the O- and M-chirp, resulting in much shorter chirp durations. This is shown in Fig. 2 and described in more detail further below.

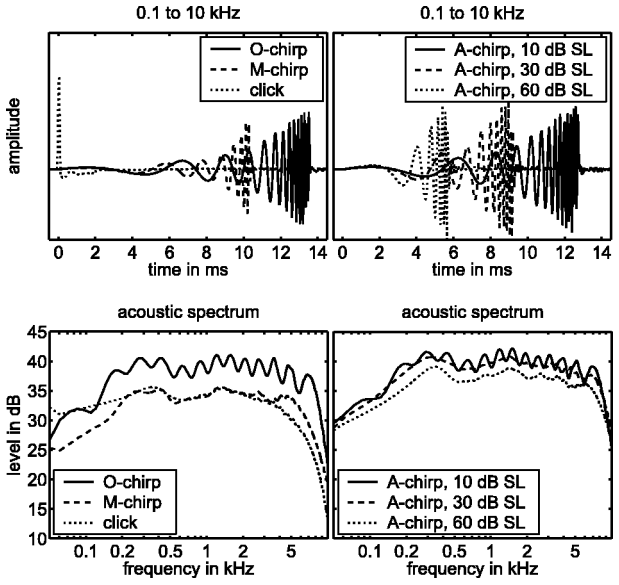


FIG. 2. Temporal waveforms (upper panel) and corresponding acoustic spectra (lower panel) of the broadband stimuli (0.1–10 kHz). The left panels show the O-chirp, the M-chirp, and the click stimulus, which are indicated as solid, dashed, and dotted functions, respectively. The right panels show the corresponding functions for the level dependent A-chirps generated for 10, 30, and 60 dB SL. Different levels were indicated by different line styles. For better comparison, all waveforms are shown for a level of 100 dB peSPL.

III. METHOD

A. Subjects

Nine normal-hearing subjects (two female and seven male) with no history of hearing problems and audiometric thresholds of 15 dB HL or better participated in the experiments. All subjects were between 28 and 38 years of age, and either volunteered or were paid for the participation in the experiments.

B. Apparatus

The experiments were carried out with a PC-based computer system which controlled stimulus presentation and recording of evoked potentials. A digital signal processing (DSP) card (Ariel DSP32C) converted the digitally generated stimulus (16 bit, 25 kHz sampling rate) to an analogous waveform. The output of the DSP card was connected to a digitally controlled audiometric amplifier, which presented the stimulus through an insert earphone (Etymotic Research ER-2) to the subject.

Electroencephalic activity was recorded from the scalp via silver/silver chloride electrodes, attached to the vertex (positive) and the ipsilateral mastoid (negative). The forehead served as the site for the ground electrode. Interelectrode impedance was maintained below 5 kΩ. Responses were amplified (80 dB) and filtered (30–3000 Hz) with a commercially available evoked potential amplifier (TDT DB4/HS4). The amplified signal was digitized by the

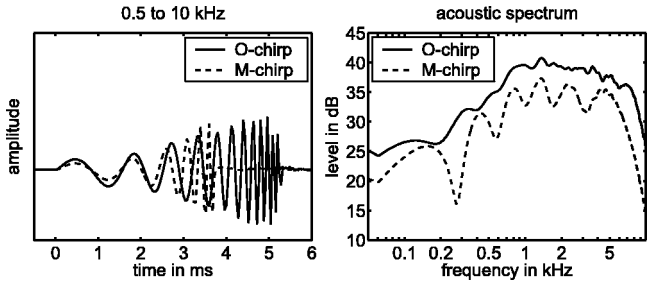


FIG. 3. Temporal waveforms (left panel) and corresponding acoustic spectra (right panel) of the (0.5–10-kHz) M-chirp (dashed lines) and the (0.5–10-kHz) O-chirp (solid lines).

DSP card (16 bit, 25 kHz sampling rate), which also performed artifact rejection and signal averaging. Responses were recorded for 37 ms following the stimulus onset. Offline filtering (digital low-pass, 1600 Hz, order 4) was done to suppress noise.

C. Stimuli and procedure

Chirps as described in Sec. II were used as stimuli. The nominal edge frequencies of the chirps were 0.1 and 10 kHz resulting in durations of 13.52 ms for the O-chirp and 10.32 ms for the M-chirp. The durations for the A-chirp varied between 5.72 ms for a sensation level (SL) of 60 dB and 12.72 ms for 10 dB SL. To compare results with standard ABR measurements, an 80-μs click stimulus was generated. The upper left panel of Fig. 2 shows the acoustic waveforms of the O-chirp, the M-chirp, and the click stimulus. The waveforms of the A-chirp for stimulation levels of 10, 30, and 60 dB SL are depicted in the upper right panel. The corresponding acoustic spectra are given in the lower panels. They were obtained by coupling the ER-2 insert earphone to a Brüel and Kjær ear simulator (type 4157) with a 1/2-inch condenser microphone (type 4134), a 2669 preamplifier, and a 2610 measuring amplifier. The spectra were derived from fast Fourier transforms (FFTs) of 100-trial time-domain averages of the stimulus over an analysis frame of 2048 samples using a sampling rate of 25 kHz. The waveforms were not windowed prior to FFT.

Since Shera and Guinan (2000) collected only very few data points below 0.5 kHz, Eq. (2) may represent only a poor description of the real behavior in this frequency region. Therefore, additional O- and M-chirp stimuli were generated with nominal edge frequencies of 0.5 and 10 kHz. The corresponding durations were 5.24 ms for the O-chirp and 3.68 ms for the M-chirp. Figure 3 shows the acoustic waveforms (left panel) and spectra (right panel) of these stimuli.

For all stimuli, the presentation level was varied between 10 and 60 dB SL in 10-dB steps. To determine the sensation level for the click, M-chirp and O-chirp, the absolute hearing thresholds were measured individually with an adaptive three-interval three-alternative forced-choice (one-up, two-down) procedure. At threshold (0 dB SL) the mean peak-equivalent sound pressure level (peSPL) was 47.2 dB for the click, 40.5 dB for the 0.1–10-kHz M-chirp, and 37.3 dB for the 0.1–10-kHz O-chirp.

The time course of the A-chirp varies with level. Thus, to be able to calculate the time course of, e.g., the 60-dB SL A-chirp one needs to know the peSPL value corresponding to 60 dB SL. This requires the knowledge of the corresponding hearing threshold. To determine the hearing threshold one needs to have the time course of the corresponding chirp. To get around this problem, a hearing threshold of 40 dB peSPL was assumed, comparable to the one for the M-chirp. This resulted in presentation levels between 50 and 100 dB peSPL, corresponding to the sensation levels from 10 to 60 dB. To verify this approach, hearing thresholds for the 10-, 30- and 60-dB SL realizations of the A-chirp were measured individually with the same procedure as described above. This resulted in different hearing thresholds for the different chirps, with a mean value of 37.5 dB peSPL, ranging from 36.1 to 38.9 dB peSPL. The presentation levels for the A-chirp were therefore overestimated by less then 3.9 dB.[4]

The subject lay on a couch in an electrically shielded, soundproof room, and electrodes were attached. The subject was instructed to keep movement at a minimum, and to sleep if possible. The lights were turned off at the beginning of the session. Each session lasted between one and two hours, depending on the subject's ability to remain still. The ear of stimulation was chosen randomly, i.e., for each subject one ear was chosen and then maintained. The acoustic signals were delivered at a repetition rate of 20 Hz for all stimulus conditions. A temporal jitter of ± 2 ms was introduced to minimize response superimposition from preceding stimuli. Thus the time interval between the onsets of two successive stimuli varied randomly and equally distributed between 48 and 52 ms. Each trial consisted of 3000 averages. For each stimulus condition, two independent trials were stored in separate buffers. These are illustrated as superimposed wave forms in the figures to show response replicability.

D. Statistical analysis

Wave-V peak-to-peak amplitude was analyzed in all stimulus conditions. The amplitude was measured from the peak to the largest negativity following it. For each condition, wave-V amplitude was averaged across subjects. A Wilcoxon matched-pairs signed-rank test ($\alpha = 0.05$) was performed to test whether the response amplitude differed significantly for two comparison stimuli.

IV. RESULTS

Figure 4 shows mean ABR, averaged across all nine subjects, obtained with the OAE-based (0.1-10-kHz) O-chirp (upper left panel), the original model-based M-chirp (upper right panel), the ABR-based A-chirp (lower left panel), and the click (lower right panel). Results for different stimulus levels are indicated on separate axes along the ordinate, and labeled with the corresponding sensation level (dB SL). Wave-V peaks are marked by small black triangles. Wave V is the only peak that can be observed in *all* stimulus conditions. For the O-chirp, no earlier waves are present, even at the highest stimulation levels. In contrast, for the M-chirp,

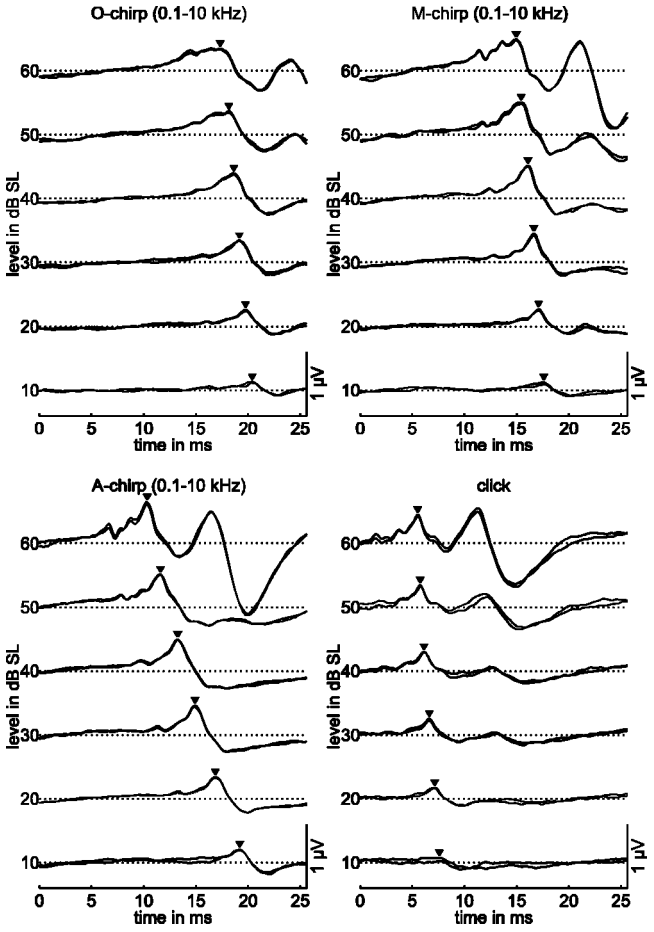


FIG. 4. ABR waveforms obtained with the 0.1–10-kHz O-chirp (upper left panel), M-chirp (upper right panel), A-chirp (lower left panel), and click (lower right panel), averaged across all nine subjects. The stimulation level varied from 10 to 60 dB SL, as indicated. At each level, two independently averaged waveforms are superimposed to show response replicability. The black triangles indicate wave-V peaks.

A-chirp, and the click, waves I and III become visible at the highest levels. Interestingly, for the A-chirp, wave I is visible even down to a level of 20 dB SL.

Figure 5 (left panel) summarizes the quantitative values for the mean wave-V amplitude as a function of the stimulation level. The click-evoked wave-V amplitude, represented by the filled squares, is always smaller than that obtained with any of the broadband chirps, represented by the other filled symbols. For example, the M-chirp (filled circles), leads to amplitude values that are more than twice the values for the click at most stimulus levels. This agrees well with the results found in Dau *et al.* (2000). At the lowest stimulation level, the A-chirp (filled downward triangles) evoked an amplitude that is about three times as large as that for the click. The amplitude-level function for the A-chirp looks like shifted by about 0.44 μV with respect to the click curve. For the A- and M-chirp, statistical analysis revealed significantly larger amplitudes than for the click at *all* stimulus levels while for the (broadband) O-chirp the difference was significant only for 10 and 40 dB SL.

Now consider the results for the O-chirp (filled upward triangles) in comparison to the original M-chirp (filled circles), having in mind that the O-chirp was based on 40 dB SPL otoacoustic emission data while the M-chirp was derived from a (linear) cochlea model based on high-level BM data. At levels of 40 dB SL and above, wave-V amplitude is smaller for the O-chirp than for the M-chirp, while at the lower levels, wave-V amplitude is about the same for the two stimuli. Statistical analysis of the amplitude data revealed significant differences between the O- and M-chirp only for levels of 50 and 60 dB SL, where wave-V amplitude for the M-chirp is higher. The results for the smaller chirp bandwidth, ranging from 0.5–10 kHz, are indicated by the corresponding open symbols. Results for the O- and the M-chirp are given by triangles and circles, respectively. The response waveforms are not shown explicitly for these two conditions. Statistical analysis revealed a significant difference between the two chirps only for a level of 20 dB SL, where the O-chirp evoked a higher wave-V amplitude than the M-chirp.

Next consider the results for the A-chirp (filled downward triangles) in comparison to the original one (M-chirp). The A-chirp resulted in a larger wave-V amplitude than the M-chirp (and any other stimulus tested here) at nearly all stimulation levels. However, statistical analysis revealed significant differences between A- and M-chirp only for low stimulation levels (10 and 20 dB SL).

Comparison of the results for the A-chirp and the O-chirp shows that the A-chirp elicited a higher wave-V amplitude than the O-chirp at all stimulation levels used here. In this case, statistical analysis results in significant differences for low and high stimulation levels (10, 20, 50, and 60 dB SL).

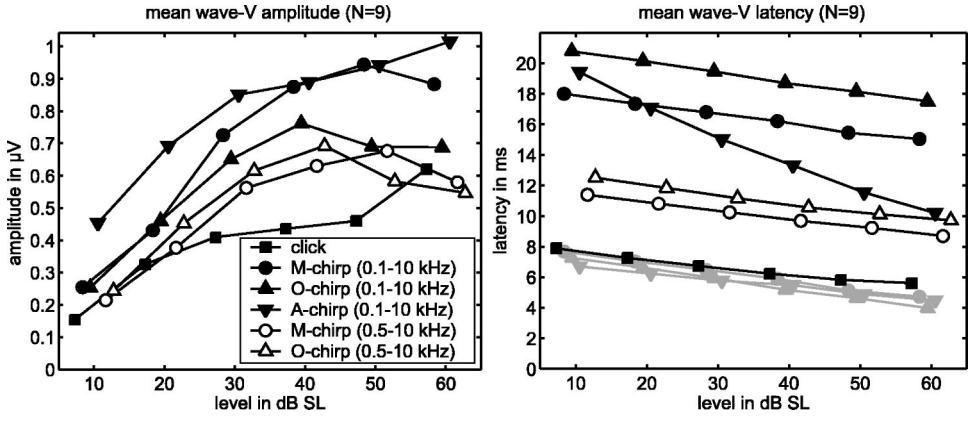


FIG. 5. Average ABR wave-V amplitude (left panel) and latency (right panel), as a function of stimulation level. Different symbols indicate different stimulus conditions. ■: click, ●: 0.1–10-kHz M-chirp, ▲: 0.1–10-kHz O-chirp, ▼: 0.1–10-kHz A-chirp, ○: 0.5–10-kHz M-chirp, and △: 0.5–10-kHz O-chirp. The shaded symbols in the right panel indicate the offset latencies for the corresponding stimuli. For better visibility, the symbols are slightly shifted along the abscissa.

The right panel of Fig. 5 shows the mean wave-V latency behavior obtained with the different stimuli. Except for the A-chirp, all functions are roughly in parallel to each other but shifted relative to each other by some amount. For these functions, the latency decreases by about 2–3 ms for a 50-dB level change (from 60 to 10 dB SL), which is consistent with literature data (e.g., Hoth and Lenarz, 1994). The main differences between the functions correspond to the differences in the respective stimulus durations, as is illustrated by the shaded functions in the same panel of the figure. They indicate the latency values for the three broadband chirps relative to stimulus *offset* instead of stimulus onset. The very similar values in this view are consistent with the idea behind the chirp paradigm that, ideally, the displacement maxima on the BM should occur in *all* channels at the same time, and thus, the latencies for the chirp and the click should be similar if expressed relative to stimulus offset. Thus, since the duration of the A-chirp changes strongly with level, this must be directly reflected in a relatively steep function if expressed relative to stimulus onset.

V. DISCUSSION

Previous studies have shown that an upward chirp can evoke a significantly larger wave-V amplitude than the conventional click (Dau *et al.*, 2000; Wegner and Dau, 2002). The equations defining the upward chirp in these earlier studies were calculated to be the inverse of the delay-line characteristic of the cochlear partition on the basis of de Boer's (1980) cochlea model. However, since this model does not take BM nonlinearity (compression) into account which is associated with level-dependent frequency selectivity, the model probably underestimates real BM group delays at low and medium stimulus intensities. The intention of the present study was to design and test chirp stimuli that might potentially cause an even larger neural synchronization than the original chirp, at least at lower stimulus intensities. Two different strategies for the generation of the stimuli were used: One stimulus, the O-chirp, was calculated from stimulus-frequency-emission group-delay data by Shera and Guinan (2000), recorded at a stimulation level of 40 dB SPL in humans. The other stimulus, the A-chirp, was based on measured ABR wave-V latency values obtained with tone pulses at various frequencies and levels. The corresponding responses evoked by these chirps were compared with results obtained with the original M-chirp and the click.

A. Usefulness of the OAE-based chirp stimulation for ABR

The data of the present study showed only small differences between O- and M-chirp stimulation. For the frequency region above 500 Hz, where reliable SFOAE data were available, the two 0.5–10-kHz chirps produced about the same wave-V amplitude (open symbols in Fig. 5). No significant advantage was found for the O-chirp at the lower levels (except for 20 dB SL), and no advantage was observed for the M-chirp at higher levels. For the broadband conditions (0.1–10 kHz; closed circles and upward triangles in Fig. 5), the M-chirp produced a larger potential amplitude than the corresponding O-chirp at the two highest levels.

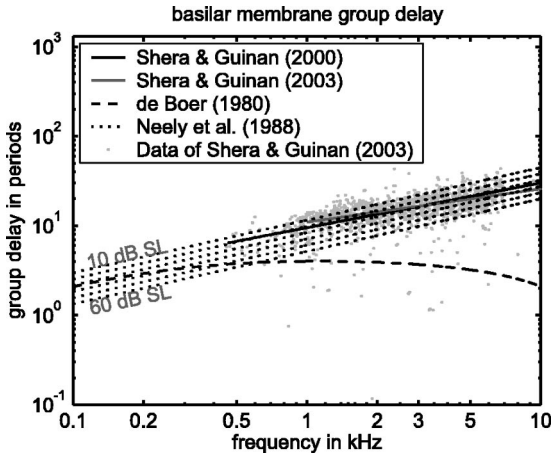


FIG. 6. BM group delay as a function of frequency. The group delay is indicated in dimensionless form in units of periods of the stimulus frequency. The shaded dots represent the BM group delay estimates of Shera and Guinan (2003), derived from SFOAE data. The gray solid line (starting at 1 kHz) represents their recent data fit, while the black solid line represents the fit function from 2000. The latter one was used in the present study as the basis for the generation of the O-chirp. The dashed line indicates the group delay based on the cochlea model by de Boer (1980). The dotted lines show the group delays predicted from Neely *et al.* (1988) for stimulus intensities ranging from 10 dB SL (top dotted curve) up to 60 dB SL (bottom dotted curve).

This indicates that, for the O-chirp, our assumed extrapolated group-delay-versus-frequency function at the very low frequencies (0.1–0.5 kHz) probably does not match the real system very well. At these very low frequencies, the function underlying the M-chirp appears to represent the better choice.

What might be the reason(s) for the similar results at low intensities obtained with these two stimuli? The hypothesis underlying the generation of the O-chirp was that, according to the theory of reflection filtering (Shera and Zweig, 1993; Zweig and Shera, 1995), reflection-source emission group delay is determined by the group delay of the BM mechanical transfer function at its peak. Thus, otoacoustic emissions may be used to provide a noninvasive measure of BM group delay, at least at low levels. However, it is possible that the estimate of the BM group delay by the SFOAE-based group delay is not very reliable. First, the emission data show large scatter and vary considerably with frequency (see Fig. 1), even though the trend lines of Shera and Guinan (2003) were quite robust within and across subjects. This large variability seems consistent with results from the study by Neely *et al.* (1988) where also large variations of OAE data were found within and across subjects. The scatter in the OAE data does not seem to arise from measurement noise—the measurements are quite reproducible in each subject (Shera and Guinan, 2003). Instead, the scatter may come from intrinsic variations in emission phase that are correlated with variations in emission amplitude across frequency (Shera and Guinan, 2003). Second, the ratio $\tau_{\mathrm{SFOAE}}/\tau_{\mathrm{BM}}$ (derived from their animal experiments) was found to be somewhat less than the predicted value of 2 such that there might be a mismatch in the predicted BM group delay above 1 kHz. Third, further following the recent findings of Shera and Guinan (2003), there is a breakdown in the proportion-

ality between emission and mechanical BM group delay in the apical part of the cochlea. Thus, for frequencies below 1 kHz, it is uncertain if the SFOAE data can be used to get a reliable estimate of BM group delay.

The O-chirp of the present study was based on a formula whose parameters differ somewhat from those suggested by Shera and Guinan on the basis of their most recent results (Shera and Guinan, 2003). The group-delay versus frequency function representing their fit to these recent data is indicated as the gray line starting at 1 kHz in Fig. 6. The group delay in this figure is indicated in dimensionless form in units of periods of the stimulus frequency, and expressed as BM group delay [assuming $\tau_{BM}(f) = 0.5 \times \tau_{SFOAE}(f)$]. In addition, the raw data as well as the other curves from Fig. 1 are replotted here for comparison. The different scale was used in order to make the data easily comparable with the figures in Shera and Guinan (2003). It is possible, in principle, that a chirp based on the estimate, in combination with an appropriate extension toward the lower frequencies (<1 kHz), would lead to larger evoked response amplitudes.

Indeed, the finding that the OAE based chirp does not produce a larger response amplitude than the other chirps does not necessarily imply that SFOAE group delay is a poor estimator of BM group delay. The contributions of the low-frequency components of the stimulus to the ABR are mainly responsible for the advantage of the chirp over the click (Dau *et al.*, 2000; Wegner and Dau, 2002). In fact, it is still possible that the estimate of BM group delay using SFOAE is reasonable at medium and high frequencies while it may be problematic at frequencies below about 1 kHz. Another aspect is that it is problematic to compare the results of the intensity-independent O-chirp with the results of the intensity-dependent A-chirp except at levels near the one used in the OAE measurements (40 dB SPL). It is possible, at least in principle, that had the SFOAE delay estimates been available for a similar range of intensities, no significant differences between A- and O-chirps would have been found at any intensity. At 30 and 40 dB SL, the differences were not significant (see Sec. IV). Finally, in more general terms, it is important to note, that a stimulus that causes a maximum amount of synchronicity at *brainstem* level (where wave V is generated) does not necessarily imply that the same stimulus also causes a maximum synchronized activity at BM level, as has been implicitly assumed in our previous studies (Dau *et al.*, 2000; Wegner and Dau, 2002) and also indirectly in the study by Neely *et al.* (1988). Wave-V latency always represents the sum of a mechanical and a neural delay, and the neural delay also might be frequency- and level-dependent. It is not well known what exactly needs to be synchronized to maximize ABR amplitude. It is possible that maximal ABR occur when first-spike latencies are equalized across CF. However, first-spike latencies are not necessarily corresponding to group delays. Thus, the O-chirp might not be optimal even if SFOAEs would provide a good measure of BM group delay.

B. Capabilities of the ABR-based chirp-stimulation paradigm

The recordings obtained with the A-chirps showed the largest response amplitude. Indeed, wave-V amplitudes for the A-chirp were found to be higher than for any of the other stimuli of the present study at most stimulation levels. This was true even though the chirp was designed on the basis of average data from a completely different set of subjects (Gorga *et al.*, 1988) and using different equipment for the recordings as in the present study. One might argue that it is not surprising to obtain good ABR results with a stimulus that was developed on the basis of ABR data. On the other hand, for a complex and nonlinear system like the auditory system, it is not clear in advance that a composite stimulus like the chirp, that sweeps through the frequencies at a rate determined by the latency values obtained in separate recordings with (transient) tone pulses would necessarily lead to such large responses. At medium and high stimulation levels, the excitation on the BM in response to tone pulses is certainly not frequency specific, due to effects of spread of excitation associated with cochlear nonlinearities. However, at low levels, the quasilinear approach that is implicitly assumed in the chirp-generation paradigm might be appropriate. The results of the present study clearly suggest that the A-chirp represents a very effective stimulus. The advantage is particularly large at low stimulation levels where the response amplitude is about three times as large as that obtained with click stimulation (even though the peak equivalent sound pressure level of the click was about 10 dB *higher* than that for the A-chirp if compared at the same SL). The finding that the A- chirp also produces the clearest wave I suggests that it also very effectively stimulates the earliest neural processing station, the auditory nerve. However, since the derivation of the chirp was based on ABR wave-V latency data, it remains not possible to clearly separate between mechanical and neural/synaptic delays. In fact, the results of the present paper do *not* depend on whether the assumptions of Neely *et al.* (1988) about the relative contributions of mechanical and neural delay are true or not. It is not possible to finally verify or falsify their assumptions on the basis of our experimental data. However, whatever the exact contributions of the different components to the overall delay are, the A-chirp may in any case be interesting and valuable for clinical application, e.g., as an objective indicator of hearing threshold. It might be particularly useful in all applications where the traditional click stimulus has been used so far.

VI. SUMMARY AND CONCLUSIONS

- Two chirp stimuli were developed such as to compensate for travel-time differences across frequency. One stimulus, the O-chirp, was based on BM group-delay estimates (at 40 dB SPL) obtained with SFOAE (Shera and Guinan, 2000). The other one, the A-chirp, was based on functions fitted to tone-pulse-evoked ABR wave-V latencies at various stimulation levels (Gorga *et al.*, 1988; Neely *et al.*, 1988). ABR obtained with these chirps were compared to

click responses and responses to the original chirp stimulus, the M-chirp, which is based on a linear cochlea model (Dau *et al.*, 2000).

- All chirps caused a larger wave-V response amplitude than the click. This is mainly caused by activity from the entire frequency range contributing to the chirp response while in the case of the click the lower frequencies do not contribute effectively to the response.
- No significant differences between the response amplitudes obtained with the O-chirp and the M-chirp were found, not even at low stimulation levels where an advantage of the O-chirp was expected. One possible explanation might be that SFOAE group delays do not allow a reliable estimate of BM group delays, particularly at low frequencies (<1 kHz). Another explanation might be that level- and frequency-dependent neural delays are involved in ABR (wave-V) latency which are not reflected in the design of these two chirps.
- The A-chirp caused the largest responses and is particularly effective at very low levels where wave-V amplitude is about three times as large as for the click. This level-dependent chirp intrinsically includes both mechanical and neural delays since it was derived from wave-V latency data. The A-chirp might be very useful for clinical applications, e.g., in connection with objective tests of hearing threshold. Specifically, this chirp might be valuable in all applications where the standard click stimulus has been used so far.

ACKNOWLEDGMENTS

The authors would like to thank the Associate Editor and two anonymous reviewers for their constructive and helpful comments and suggestions. The authors thank Chris Shera for very interesting discussions and for providing us with his data. The authors also thank our colleagues from the research group ''Medizinische Physik'' at the University of Oldenburg for fruitful discussions. This study was supported by the Deutsche Forschungsgemeinschaft (DFG), Research Program ''Neural basis of cognitive functions'' (SFB 517). Finally, the authors are grateful for the support from the Technical University of Denmark (DTU) in the final phase of this project.

[1]These constants describe the fit to the data of Shera and Guinan (2000). In the meantime, the authors collected more data points and provided different values for the two parameters: $c=0.43$ s and $\alpha=-0.63$ (Shera and Guinan 2003).

[2]The time point t_0 can be chosen somewhat arbitrarily. The only constraint is that the denominators of the fractions inside the brackets in Eq. (5) are not allowed to be zero. Dau *et al.* (2000) used $t_0=t(0\text{ Hz})$. Since this value is not defined within Eq. (2), we used $t_0=\tau_{\mathrm{BM,O}}(50\text{ Hz})$.

[3]$t_0(i)$ can again be chosen somewhat arbitrarily, but in this case it is a function of the stimulus intensity i. We chose $t_0(i)=\tau_{\mathrm{BM,A}}(i,50\text{ Hz})$.

[4]A more accurate method would have been to use these thresholds to calculate new realizations of the 10- to 60-dB SL A-chirps and to measure their thresholds again. However, this would have resulted in a time-consuming iterative process.

von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York).

de Boer, E. (**1980**). ''Auditory physics. Physical principles in hearing theory I,'' Phys. Rep. **62**, 87–174.

Dau, T., Wagner, O., Mellert, V., and Kollmeier, B. (**2000**). ''Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion,'' J. Acoust. Soc. Am. **107**(3), 1530–1540.

Don, M., and Eggermont, J. J. (**1978**). ''Analysis of the click-evoked brain stem potentials in man using high-pass noise masking,'' J. Acoust. Soc. Am. **63**(4), 1084–1092.

Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1994**). ''Auditory brainstem response (ABR) peak amplitude variability reflects individual differences in cochlear response times,'' J. Acoust. Soc. Am. **96**(6), 3476–3491.

Don, M., Masuda, A., Nelson, R., and Brackmann, D. (**1997**). ''Successful detection of small acoustic tumors using the stacked derived-band auditory brain stem response amplitude,'' Am. J. Otol. **18**(5), 608–621.

Gorga, M. P., Kaminski, J. R., Beauchaine, K. A., and Jestead, W. (**1988**). ''Auditory brain-stem responses to tone bursts in normal hearing subjects,'' J. Speech Hear. Res. **31**(1), 87–97.

Greenwood, D. D. (**1990**). ''A cochlear frequency position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**(6), 2592–2605.

Hoth, S. and Lenarz, T. (**1994**). *Elektrische Reaktionsaudiometrie* (Springer, Berlin).

Kiang, N. Y. S. (**1975**). ''Stimulus representation in the discharge patterns of auditory neurons,'' in *The Nervous System, Vol. 3: Human Communication and Its Disorders*, edited by D.B. Tower (Raven, New York), pp. 81–95.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* M.I.T. Research Monographs 35 (MIT Press, Cambridge).

Moore, B. C., and Glasberg, B. R. (**1997**). ''A model of loudness perception applied to cochlear hearing loss,'' Aud. Neurosci. **3**, 289–311.

Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). ''Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism,'' J. Acoust. Soc. Am. **106**(5), 2761–2778.

Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). ''Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli,'' J. Acoust. Soc. Am. **83**(2), 652–656.

Norton, S. J., and Neely, S. T. (**1987**). ''Tone-burst-evoked oto-acoustic emissions in normal hearing subjects,'' J. Acoust. Soc. Am. **81**(6), 1860–1872.

Oxenham, A. J., and Plack, C. J. (**1998**). ''Suppression and the upward spread of masking,'' J. Acoust. Soc. Am. **104**(6), 3500–3510.

Oxenham, A. J., and Shera, C. A. (**2003**). ''Estimates of Human Cochlear Tuning at Low Levels Using Forward and Simultaneous Masking,'' J. Assoc. Res. Otolaryngol. **4**(4) in press.

Oxenham, A. J., Moore, B. C. J., and Vickers, D. A. (**1997**). ''Short-term temporal integration: evidence for the influence of peripheral compression,'' J. Acoust. Soc. Am. **101**(6), 3676–3687.

Plack, C. J., and Oxenham, A. J. (**1998**). ''Basilar-membrane nonlinearity and the growth of forward masking,'' J. Acoust. Soc. Am. **103**(3), 1598–1608.

Rhode, W. S. (**1971**). ''Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique,'' J. Acoust. Soc. Am. **49**(4), 1218–1231.

Rhode, W. S., and Recio, A. (**2000**). ''Study of the mechanical motions in the basal region of the chinchilla cochlea,'' J. Acoust. Soc. Am. **107**(6), 3317–3332.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). ''Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases,'' J. Acoust. Soc. Am. **80**(5), 1364–1374.

Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**(4), 1057–1067.

Ruggero, M. A., Rich, N. C., Robles, L., and Recio, A. (**1995**). ''The effects of acoustic overstimulation, other cochlear injury and death on basilar membrane responses to sound,'' in *Effects of Noise on Hearing: Vth International Symposium*, edited by R.J. Salvi, A. Axelsson, D. Henderson, and R. Hamernik (Thieme Medical, Stockholm).

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**(4), 2151–2163.

Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). ''Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique,'' J. Acoust. Soc. Am. **72**(1), 131–141.

Shera, C. A., and Guinan, J. J. (**2000**). ''Frequency dependence of stimulus-

frequency-emission phase: Implications for cochlear mechanics,'' in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 381–387.

Shera, C. A., and Guinan, J. J. (**2003**). ''Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning,'' J. Acoust. Soc. Am. **113**(5), 2762–2772.

Shera, C. A., and Zweig, G. (**1993**), ''Order from the chaos: Resolving the paradox of periodicity in evoked otoacoustic emission,'' in *Biophysics of Hair-Cell Sensory Systems*, edited by H. Duifhuis, J.W. Horst, P. van Dijk, and S.M. van Netten (World Scientific, Singapore), pp. 54–63.

Shera, C. A., 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**(5), 3318–3323.

Wegner, O., and Dau, T. (**2002**). ''Frequency specificity of chirp-evoked auditory brainstem responses,'' J. Acoust. Soc. Am. **111**(3), 1318–1329.

Yates, G. K., Winter, I. M., and Robertson, D. (**1990**). ''BM nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range,'' Hear. Res. **45**(3), 203–220.

Zweig, G. (**1976**). ''Basilar membrane motion,'' Cold Spring Harbor Symp. Quant. Biol. **40**, 619–633.

Zweig, G., and Shera, C. A. (**1995**). ''The origin of periodicity in the spectrum of otoacoustic emissions,'' J. Acoust. Soc. Am. **98**(4), 2018–2047.

Using statistical decision theory to predict speech intelligibility. III. Effect of audibility on speech recognition sensitivity

Hannes Müsch[a)]
Sound ID, 3430 West Bayshore Road, Palo Alto, California 94303

Søren Buus
Communications and Digital Signal Processing Center, Department of Electrical and Computer Engineering (440 DA) and Institute for Hearing, Speech, and Language, Northeastern University, Boston, Massachusetts 02115

(Received 6 October 2003; revised 15 June 2004; accepted 22 July 2004)

The speech recognition sensitivity (SRS) model [H. Müsch and S. Buus, J. Acoust. Soc. Am. **109**, 2896–2909 (2001)] is a macroscopic model for predicting speech intelligibility. The present study proposes a modification to the relation between the SRS model's audibility-noise variance and the signal-excitation to noise-excitation ratio (SNR_E) in the auditory periphery. The modified relation is derived from data obtained in nine studies that measured normal-hearing listeners' consonant-recognition performance at several levels of speech-spectrum shaped noise. When the audibility-noise variance is directly proportional to the relative power of the noise excitation in the auditory periphery, the SRS model yields good predictions of the data. Four of the nine studies also reported consonant-recognition performance in various filtering conditions. Good predictions of these data were achieved with SRS model parameters that were consistent with the model parameters fitting the speech-in-noise data and with the model parameters used in the original SRS papers. [DOI: 10.1121/1.1791716]

PACS numbers: 43.66.Ba, 43.71.An, 43.71.Es [DDO] Pages: 2223–2233

I. INTRODUCTION

Recently, Müsch and Buus (2001a,b) proposed the speech recognition sensitivity (SRS) model as a tool for predicting the intelligibility of a variety of speech materials degraded by filtering and noise. The SRS is a macroscopic model of speech intelligibility that uses statistical decision theory to predict average performance in a speech-recognition task from the audibility spectrum of the speech signal, the linguistic entropy (van Rooij and Plomp, 1991) of the speech, and the number of response alternatives in the listening test. The goal of the SRS model is similar to that of the articulation index (AI) (French and Steinberg, 1947; Beranek, 1947; Fletcher and Galt, 1950; Kryter, 1962a; ANSI, 1969), the SII (ANSI, 1997), and the STI (Steeneken and Houtgast, 1980, 1999). However, the predictions of the AI-based models (i.e., the AI, SII, and STI) differ from the predictions of the SRS model because of different underlying assumptions.

Müsch and Buus (2001a,b) showed that the SRS model accounts for several sets of data, including some that are problematic to predict with AI-based models. In particular, intelligibility in conditions in which the speech signal comprised several spectrally distinct bands of speech (such as in band-stop filtered speech) is consistently underestimated by the AI (e.g., Kryter, 1962b), a shortcoming that ultimately led to the exclusion of such filter configurations from the scope of the standardized AI (ANSI, 1969) and SII (ANSI, 1997) and recently prompted an extension of the STI aimed at resolving this shortcoming (Steeneken and Houtgast, 1999). The SRS model successfully predicted intelligibility in these conditions (Müsch and Buus, 2001a,b). The SRS model also succeeded in predicting a finding by Hirsh *et al.* (1954) that has been recognized as a challenge to the assumptions underlying articulation theory (Licklider, 1959). Hirsh *et al.* (1954) reported that two listening conditions that yield identical nonsense-syllable recognition scores can produce word-recognition scores that differ from one another. The AI cannot explain such phenomenona,[1] but they are inherent to the SRS model structure and were successfully predicted by it (Müsch and Buus, 2001a). An additional advantage of the SRS model over AI-based models is that it incorporates effects of test structure such as the number of response alternatives available to the listener and the linguistic entropy of the test material into the prediction. AI-based models must incorporate these factors empirically.

Because Müsch and Buus's (2001a,b) main objective was to verify the SRS model's ability to account for synergetic and redundant interactions of intelligibility contributions from spectrally separate bands of speech, comparatively little attention was given to modeling the effect of SNR on intelligibility. The SNR was related to the relevant SRS model parameter through a piecewise-linear function [see Eq. (6) in Müsch and Buus, 2001b] because this approach worked well for the SII and there was insufficient data available to support any alternatives. Indeed, this simple relation agreed reasonably well with many, but not all data. One instance where it did not agree with expectations was the prediction of Kryter's (1962a) set of transformation curves between the AI and the percent-correct score. Although good predictions were possible by adjusting the SRS model parameters directly, the implied SNRs were too low.

a)Electronic mail: hmuesch@soundid.com

Consequently, Müsch and Buus (2001a, p. 2904) observed that the relation through which the SNR entered the SRS model would likely need revision.

The purpose of this investigation is to revise the function that relates the SNR to the relevant SRS model parameter by fitting the SRS model to several data sets from the literature. Data sets were selected to be strongly affected by signal audibility while having only a small impact on model parameters that are unrelated to signal audibility. This selection reduces the chance of model parameters unrelated to signal audibility unduly affecting the fits through accidental tradeoffs. These data sets and the SRS model predictions of them are presented and discussed in Secs. III–VI. Additional predictions of data that did not meet the selection criteria for the former predictions are presented in Sec. VII. These predictions rely on the model parameters derived in the first part and serve as a test of their general applicability.

Before presenting the details of these predictions, we briefly review the SRS model in the next section.

II. REVIEW OF THE SRS MODEL

A detailed explanation of the SRS model and its underlying assumptions is given in Müsch and Buus (2001a). Briefly, the SRS model assumes that listeners correlate an auditory representation of the received signal with memorized templates of ''ideal'' representations of speech signals. All signal templates are assumed to be mutually orthogonal. The ideal signal exactly matches one of the templates, and the correlation between this template and the ideal signal is unity. The correlation between the ideal signal and all other templates is zero. If the listener were to receive only ideal signals, performance would be perfect. In practice, however, only noise-degraded versions of the signal are available and the correlations between the noise-degraded signal and the templates become random variables. These random variables can be transformed into decision variables with equal variances. The means are zero for the nontarget templates and larger than zero for the target template. The mean value of the transformed correlation between the auditory representation of the signal and the target template is called the congruence index, Δ. After correlating the representation of the received signal with all templates, the detector selects the template with the highest transformed correlation as the received signal. Performance depends on the relation between the variance of the transformed random variables, σ^2, represented by noises in the model, and the congruence index, Δ, and is calculated as a sensitivity index, d', which is a descriptor of intelligibility:

$$d' = \frac{\Delta}{\sqrt{\sigma^2}}. \tag{1}$$

The overall variance, σ^2, in Eq. (1) results from three noise sources that are thought to be uncorrelated [Eq. (2)]:

$$\sigma^2 = \sigma_p^2 + \sigma_N^2 + \sigma_{cog}^2. \tag{2}$$

One noise source is related to the speech-production process. Its variance is denoted by σ_p^2. It arises because any utterance produced by a speaker only approximates the ideal signal. The difference between the ideal signal (the template) and the signal actually produced is a production-related noise. The production-noise components in the various frequency bands are assumed to be correlated across frequency [see Eq. (4) of Müsch and Buus, 2001a].

The second noise source is related to the audibility of the speech signal. Its variance is denoted by σ_N^2. When parts of the dynamic range of speech are inaudible because they are below absolute or masked threshold, a fraction of the cues relevant for speech decoding are unavailable and intelligibility suffers. The loss of speech cues is expressed in the SRS model through an increase in the variance of the audibility noise, σ_N^2. An increased audibility-noise variance lowers the d', which worsens the predicted intelligibility. The relation between speech audibility, expressed as a frequency-dependent ratio of speech excitation to noise excitation in the auditory periphery, $\mathrm{SNR_E}$, and the audibility-noise variance, $\sigma_N^2 = f^2(\mathrm{SNR_E})$, is the focus of this article.

The third noise source models the effect of linguistic entropy on intelligibility. Its variance is denoted by σ_{cog}^2. Spoken messages with a high degree of sequential redundancy (e.g., sentences) require less audibility than messages with little redundancy (e.g., isolated words) to be recognized with equal accuracy (Miller *et al.*, 1951). To account for this effect, the model assumes an additional variance whose magnitude depends on the linguistic entropy of the speech material. This additional variance, σ_{cog}^2, is small when the message has a high degree of redundancy and large when the redundancy in the message is small.

The sensitivity index, d', of Eq. (1) is a descriptor of intelligibility that is similar to the AI, SII, and STI in the sense that it is monotonically related to intelligibility and must be transformed to yield a predicted performance (percent correct recognition). For AI-based models these transformations are typically derived empirically for every data set. In the SRS model the relation between d' and percent correct recognition, P_c, is fixed and depends only on the number of response alternatives, M, that are available to the listener. It is given in Eq. (3), where P_c is the percentage of test items identified correctly, $f_{SN}(\cdot, d', 1)$ is the Gaussian probability density function with mean d' and unit variance, and $F_N(\cdot, 0, 1)$ is the cumulative Gaussian probability function with zero mean and unit variance

$$P_c = 100 \int_{x=-\infty}^{\infty} f_{SN}(x, d', 1) \cdot F_N^{M-1}(x, 0, 1)\, \mathrm{d}x. \tag{3}$$

A graphical representation of Eq. (3) is available in Müsch and Buus's (2001a) Fig. 1.

III. METHOD

The objective of this investigation is to define the relationship between signal audibility and the variance of the audibility noise, σ_N^2. Signal audibility is expressed as a frequency-dependent signal-to-noise ratio, $\mathrm{SNR_E}$, in the auditory periphery. At any critical-band rate, the $\mathrm{SNR_E}$ is the difference in dB between the excitation level generated by the speech energy at that critical-band rate and the excitation level generated by the combination of masking sounds and

an internal noise floor that accounts for the absolute threshold of hearing (cf. Zwicker and Feldtkeller, 1999). Thus, the relation whose form is to be established in this investigation is of the form $\sigma_N^2 = f^2(\mathrm{SNR_E})$. Because this function cannot be observed directly, it must be deduced by applying the entire SRS model with several guesses of $f^2(\mathrm{SNR_E})$ until the form that best predicts observed intelligibility scores is found. This process will be greatly helped if data sets are selected for which predictions depend primarily on $f^2(\mathrm{SNR_E})$ and the impact of other model parameters on the predictions is small. As will be shown, this is the case for intelligibility measurements in speech-spectrum-shaped noise.

If the spectrum of the masking noise matches the spectrum of the speech, the long-term acoustic speech-to-noise ratio (SNR) is the same at all frequencies. When, in addition, speech and masker spectra are relatively smooth—as they are when averaged across a phonetically balanced set of speech stimuli produced by many talkers—and the noise levels high enough to exceed the threshold of audibility at all frequencies, the signal-to-noise ratio in the auditory periphery, $\mathrm{SNR_E}$, will also be frequency independent and equal to the acoustic SNR. Assuming for the moment that these assumptions are met, the SRS model can be simplified so that the predicted d' depends on f^2 (SNR) and only one other parameter.[2] Specifically, in the fundamental formula of the SRS model,[3]

$$d' = \frac{\Sigma_i \Delta(i)}{\sqrt{\sigma_{cog}^2 + \Sigma_i \Sigma_j (r(i,j)\sigma_p(i)\sigma_p(j)) + \Sigma_i \sigma_N^2(i)}}, \quad (4)$$

the numerator, which is the sum of all band congruence indices in the pass band, can be replaced by a constant. This is permitted because the bands that contribute to intelligibility are always the same, irrespective of the SNR. Using the same congruence-index density function as Müsch and Buus (2001a,b), the numerator is fixed at 19, which results from adding the congruence indexes of all speech bands, reflecting that speech is audible over its entire spectrum. The denominator can also be simplified by combining terms that do not change with SNR. The model structure shows that the cognitive-noise variance, σ_{cog}^2, is independent of $\mathrm{SNR_E}$. The powers of the production noises in the bands, σ_p^2, and their correlations $r(i,j)$ are also independent of $\mathrm{SNR_E}$. Therefore, the two left-hand terms under the square root can be combined into one parameter, X, which represents the sum of the cognitive and production noise powers. Because the $\mathrm{SNR_E}$ is the same in all bands, the audibility noises in the individual bands can be replaced by an overall audibility noise, which depends on $\mathrm{SNR_E}$ (and, because of the assumptions made, SNR) in the same way as the band audibility noises.

With these observations, the formula determining the speech sensitivity index, d', simplifies to

$$d'(\mathrm{SNR}) = \frac{19}{\sqrt{X + f^2(\mathrm{SNR})}}. \quad (5)$$

In this simplified relation, d' is independent of the shape of the congruence-index density function, the correlation between the production noises across frequency, and the relative sizes of production and cognitive noises. The predicted d' depends only on the SNR and one other parameter, X. Consequently, data collected under conditions that allowed the simplifications leading to Eq. (5) can be predicted by adjusting $f^2(\mathrm{SNR})$ and only one additional free parameter. They are therefore particularly suited for determining the shape of $f^2(\mathrm{SNR_E})$.

To take advantage of these simplifications only studies that tested intelligibility of broadband speech in the presence of spectrum-matched noise were admitted into the study. As will be detailed later, the goodness of the spectral match varied across the studies that were selected, but in each case was deemed sufficient to justify the simplifications of Eq. (5).

To complete the prediction, the measured intelligibility must be related to the calculated speech recognition sensitivity index, d'. This is accomplished by applying Eq. (3), which requires that the number of response alternatives, M, be known. Therefore, one additional requirement for inclusion in this study was that the response-set size could be easily defined. This means that either a closed response set was used or that the listeners were familiar with the stimulus set so that it could be assumed that they limited their responses to include only valid stimuli. Finally, to facilitate comparison across data sets, cognitive-noise variances were required to be of similar magnitude.

A literature search revealed that the largest number of data sets meeting these criteria were studies that measured the recognition of individual consonants in the context of nonsense syllables for normal-hearing listeners who were tested in their native language using only auditory cues (i.e., no visual information).

In summary, the relation between $\mathrm{SNR_E}$ and audibility-noise variance, $\sigma_N^2 = f^2(\mathrm{SNR_E})$, is estimated by using Eqs. (3) and (5) to predict consonant recognition in speech-spectrum-shaped noise as a function of SNR. The data used for the predictions were obtained from the literature and are described in the next section.

IV. DESCRIPTION OF DATA

A literature search for studies with data that conform to the criteria outlined in the previous section revealed the following studies: Boothroyd and Nittrouer (1988), Steeneken (1992), Grant and Walden (1996), Rankovic and Levy (1997), Hornsby and Ricketts (2001), Horwitz *et al.* (2002), and Devore *et al.* (2002, 2003). In addition, we were allowed access to one unpublished study by Ben Hornsby.

The relevant data from all studies are summarized in Fig. 1, which shows the percentages of correctly recognized consonants as a function of SNR. Several aspects of the experimental procedure are summarized in Table I. Although the data were obtained under similar test conditions, sizable differences in listeners' performances are apparent. Two studies collected data at −15 dB SNR and both report chance performance at that SNR. However, for higher SNRs, where performance exceeds chance, there is a large spread of performances for the same SNR. For example, at 9 dB SNR Grant and Walden (1996) obtained perfect intelligibility,

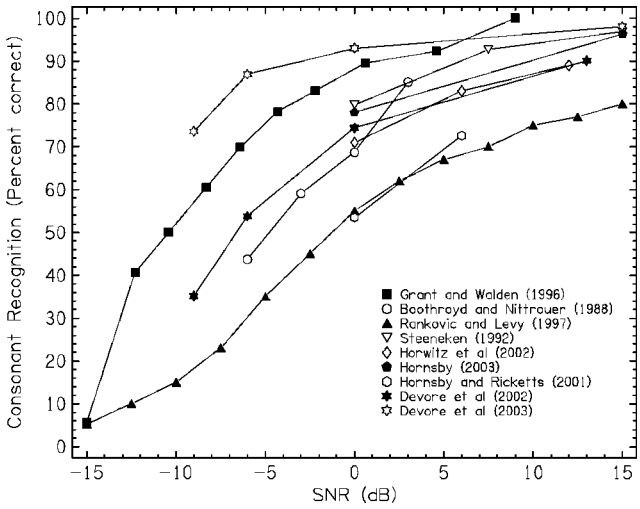


FIG. 1. Percent correct recognition of consonants in a nonsense-syllable context as a function of SNR. The noise was shaped to match the spectrum of the speech (see Table I for details). Results are compiled from nine studies. Data obtained in quiet are plotted at 15 dB SNR, reflecting the assumption that at 15 dB SNR or higher all speech cues relevant for intelligibility are available to the listener. This assumption is identical to that made in the SII (ANSI, 1997).

whereas Rankovic and Levy (1997) obtained only about 74% correct. At −9 dB SNR the best performance in any study was about 75% correct (Devore *et al.*, 2003) and the worst was only about 15% correct (Rankovic and Levy, 1997). Despite the large spread, the functions have similar slopes, as might be expected from studies that used speech material of similar linguistic complexity.

Factors contributing to the wide range of performance levels at the same SNR are differences in response-set size, different definitions of SNR, different absolute listening levels, and differences in the clarity with which the syllables were produced. These factors will be discussed next.

Some of the performance differences apparent in Fig. 1 can be attributed to differences in the number of response alternatives available to the listener. Recognition performance increases when the number of response alternatives decreases (Miller *et al.*, 1951; Green and Birdsall, 1958; Müsch and Buus, 2001b). Green and Birdsall (1958) and Müsch and Buus (2001b) demonstrated that Eq. (3) explains the effect of response-set size on performance and it will be used here to account for the different response-set sizes. For this transformation, it was assumed that the size of the response set equaled the size of the stimulus set. This is definitely correct for those studies in which listeners marked their responses on a list of all possible responses, which they had in front of them (Grant and Walden, 1996; Rankovic and Levy, 1997; Hornsby and Ricketts, 2001; Horwitz *et al.*, 2002; Hornsby, 2003). In other studies (Boothroyd and Nittrouer, 1988; Steeneken, 1992), subjects were allowed to write down what they heard, i.e., the response set was formally open. However, because the number of different stimuli was relatively small and listeners were thoroughly familiar with the test set, they were likely to be aware of all possible response alternatives and to limit their responses to members of the (known) stimulus set. By doing so, they effectively adopted a closed-response-set paradigm with the response set matching the stimulus set.

The data in Fig. 1 are replotted in Fig. 2, but performance in Fig. 2 is expressed as speech recognition sensitivity, d'. The transformation to d' accounts for expected effects of response-set size and makes clear that the relatively

TABLE I. Number of consonants and vowels, talkers, productions per talker per token, and listeners for the studies used in this article. Also listed are the speech level and details about the speech-spectrum-shaped noise. With the exception of the entry for Steeneken (1992), the "number of productions per token per talker" refers to the entire syllable.

	No. of								
Study	Initial consonants (CV or CVC)	Final consonants (VC or CVC)	Center consonants (VCV)	Vowels in context	Talkers[a]	Productions per token per talker	Listeners	Speech level	Noise spectrum
Boothroyd and Nittrauer (1988)	10	10	···	10	1 m	1	32	comfortable	"equal masking on speech"
Grant and Walden (1996)	···	···	18	1	1 f	10	8	approx. 70 dB SPL	talker-matched
Rankovic and Levy (1997)	19	19	···	3	2 m/2 f	1	8	60 dB SPL	approx. speech spectrum[d]
Steeneken (1992)	17	11	···	15	4 m	>150[g]	4	70 dB SPL	average of two talkers[e]
Hornsby and Ricketts (2001)	22	21	···	2	1 m	1	9	65 dB SPL[c]	average speech spectrum[f]
Horwitz *et al.* (2002)	22	21	···	3	1 m/1 f	1	6	77 dB SPL[b]	talker-matched
Devore *et al.* (2002)	9	9	···	1	1 m	3	6	65 dB SPL	talker-matched
Devore *et al.* (2003)	9	9	···	1	1 m	3	3	66 dB SPL	talker-matched
Hornsby (2003)	22	···	···	1	1 m/1 f	1	8	65 dB SPL[c]	average speech spectrum[f]

[a]m=male, f=female.
[b]In 2cc coupler.
[c]In Zwislocki coupler.
[d]Spectrum level flat until 1 kHz, −6 dB/oct between 1 and 4 kHz, and −20 dB/oct between 4 and 8 kHz.
[e]Averaged spectra of talker and one other talker of same sex.
[f]Long-term speech spectrum of Byrne *et al.* (1994).
[g]Estimated productions *per consonant*; each talker recorded 50 lists of 51 CVCs, every initial consonant occurred three times in each list; some vowels and final consonants were used more than three times per list.

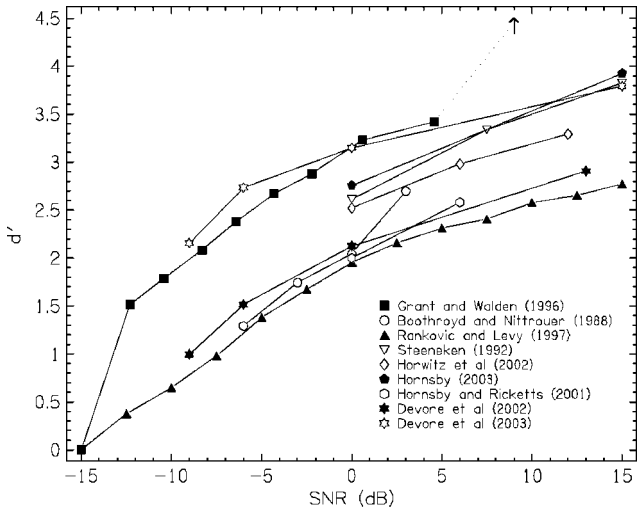


FIG. 2. The data of Fig. 1 expressed as a sensitivity index, d'. The percent-correct scores of Fig. 1 were transformed into d' by way of Eq. (3) in the text. This representation accounts for performance differences that result from differences in the number of response alternatives available to the listeners. The arrow for the datum at 9 dB SNR in Grant and Walden's (1996) data indicates that the corresponding d' is infinity.

small response sets used in the studies by Devore *et al.* (2002) and by Boothroyd and Nittrouer (1988) are likely to be responsible for their scores being superior to those obtained by Rankovic and Levy (1997) and by Hornsby and Ricketts (2001). Likewise, differences between Grant and Walden's (1996) data and Devore *et al.*'s (2003) data are largely eliminated when differences in the size of the response set are accounted for. On the other hand, correcting for the response-set size only slightly affects the relative positions of the data obtained by Steeneken (1992), Horwitz *et al.* (2002), Hornsby (2003), and the two groups stated above. Altogether, there remains a considerable spread of performances, indicating that differences in response-set sizes are only partially responsible for the wide range of scores seen in Fig. 1.

One possible reason for some studies achieving higher scores than others at the same reported SNR may be differences in the definition of SNR. Most studies define the SNR as the average RMS level of the tokens relative to the level of the masking noise. However, two studies define the SNR as the level of a portion in the carrier phrase relative to the level of the noise (Boothroyd and Nittrouer, 1988; Devore *et al.*, 2003). Studies that defined the speech level as the average RMS level of the tokens are subdivided into two groups: in one group (Rankovic and Levy, 1997; Hornsby and Ricketts, 2001; Horwitz *et al.*, 2002; Hornsby, 2003) tokens are simply concatenated without pauses and their overall RMS level is measured. In this approach, the natural level variations among different syllables are maintained, which is also true for studies that base their speech-level estimates on the level of the carrier phrase. In the other group (Grant and Walden, 1996; Devore *et al.*, 2002) the individual tokens were RMS equalized. This equalization eliminates the natural level differences between syllables. Boothroyd and Nittrouer (1988) differed from all other studies by measuring the "peak" level rather than the RMS level. Their paper does not contain a definition of "peak" level. However, the CUNY nonsense syllable test (NST, Resnick *et al.*, 1975) uses the same carrier phrase as Boothroyd and Nittrouer (1988) and appears to use a similar calibration. An analysis of the original recordings of the CUNY NST indicates that Resnick *et al.*'s (1975) method of determining speech level, which presumably is equivalent to the method used by Boothroyd and Nittrouer (1988), defines SNRs as approximately 5 dB higher than SNRs determined by measuring the RMS level of all concatenated tokens. In Figs. 1 and 2, Boothroyd and Nittrouer's (1988) data are plotted at the SNRs at which they were reported, but SNRs were assumed to be 5 dB lower in the model predictions discussed in the following sections to bring the SNR definition more in line with that used in the other studies.

Absolute listening level is another variable that affects intelligibility, even if an effort is made to maintain a constant SNR (e.g., Hornsby and Ricketts, 2001). Because the studies report different listening levels (Table I) and differences in the methods used to measure them, absolute presentation level may also have influenced the data.

Differences in the clarity of speech production among talkers (van Wijngaarden and Houtgast, 2004) and differences in experimental design are other possible reasons for performance differences in listening tasks that otherwise appear to be of equal difficulty.

The foregoing discussion suggests that there are numerous factors that can affect performance. Key aspects of the experimental design of the studies used in this paper are summarized in Table I. Inspection of this table did not reveal any pattern that explains the spread of the data.

In the following predictions, we attempt to account for effects related to response-set size, talking style, and general "difficulty" of the task with the appropriate SRS-model parameters. In light of the fact that the definitions of SNR are somewhat inconsistent across studies, the remaining differences are addressed by allowing a parameter "offsetting" the SNR (see Sec. V) to vary among the studies.

V. RESULTS

As was pointed out in Sec. III, a heuristic approach was needed to find the function $f^2(\mathrm{SNR_E})$ that optimized the fit of Eq. (5) to the data. Experimentation indicated that a relation of the general form $f^2(\mathrm{SNR}_E)=10^{A\mathrm{SNR}_E+B}$ would provide a good fit to the data. This function implies that the logarithm of the audibility-noise variance is linearly related to the $\mathrm{SNR_E}$ in dB:

$$10\cdot\log_{10}(\sigma_N^2)=A\cdot\mathrm{SNR_E}+B. \tag{6}$$

To determine suitable values for A and B, the following fitting procedure was applied. For each data point (i.e., measured intelligibility score), the effective σ_N^2 was found by inverting Eqs. (3) and (5). To invert Eq. (5), a value for X, which represents the sum of production- and cognitive-noise variances, must be assumed. Each data set was allowed a separate value of X because different tests, which were recorded by different talkers, will have different production-noise variances, and cognitive-noise variances may also have

TABLE II. Best-fitting model parameters for prediction of consonant recognition in noise. The second column from the right shows the correlation between SNR and $\log_{10}(\sigma_N^2)$. The high negative correlation indicates that a linear relation between SNR and $\log_{10}(\sigma_N^2)$, such as Eq. (7), is very appropriate for modeling the data. The column on the right indicates the number of data points that went into the calculation of the correlation coefficient.

Study	X	B	r^2	N
Boothroyd and Nittrauer (1988)	25.2	12.1	−0.991	4
Grant and Walden (1996)	27.9	9.1	−0.998	7
Rankovic and Levy (1997)	45.4	18.2	−0.985	12
Steeneken (1992)	23.0	15.5	−0.976	3
Hornsby and Ricketts (2001)	42.2	16.8	−1.000	2
Horwitz *et al.* (2002)	31.5	14.8	−0.992	3
Devore *et al.* (2002)	40.9	15.6	−0.998	4
Devore *et al.* (2003)	24.9	8.8	−0.993	4
Hornsby (2003)	22.7	14.0	−1.000	2

varied among the tests. However, for all data points within a data set X had to be the same. The slope parameter A in Eq. (6) was forced to be the same for all data sets. Ideally, the parameter B, which is an offset to the SNR, would also be the same for all studies. However, in recognition of the fact that differences in SNR definition were speculated to be one reason for the large performance differences among the studies (see Sec. IV), B was initially allowed to assume separate values for each data set. Like X, it was held constant within a given data set.

With these constraints, the best-fitting slope parameter A was -0.85. The goodness of fit decreased only slightly when the slope was forced to be -1. A slope of -1 has great theoretical appeal because it implies that the power of the audibility noise in the SRS model is proportional to the noise-to-signal power ratio in the auditory periphery. Because the fit with a slope of -1 is almost as good as the fit with the optimal slope, the slope was fixed at -1 from here on. With the slope fixed, Eq. (6) can be simplified and the audibility noise of the SRS model is given by Eq. (7):

$$10 \cdot \log_{10}(\sigma_N^2) = -\mathrm{SNR_E} + B, \tag{7a}$$

$$\sigma_N^2 = b \cdot \frac{n}{s}, \tag{7b}$$

where $\mathrm{SNR_E}$ is the ratio, in dB, of the signal excitation power, s, to the noise excitation power, n, in the auditory periphery, and B and b are two related fitting constants [$B = 10 \cdot \log_{10}(b)$]. The best-fitting values of X and B are summarized in Table II. Figure 3 shows the logarithm of the audibility-noise variance derived from the data (symbols) and the best-fitting relation between SNR and the logarithm of the audibility-noise variance (lines). The corresponding RMS difference between the measured performance and the predicted performance averaged across all studies is 2.9 rau (rationalized arcsine units; Studebaker, 1985).

Whereas the use of different talkers implies that X should vary from study to study, the value for B should remain constant provided the same definition of SNR is used. Although the data sets at hand employed different SNR definitions, we proceeded to fit these data with the parameter B constrained to be the same for all studies. This was done to provide an estimate of B that is general enough to be a useful guideline for fitting data in the future. Using Eq. (7) and allowing X to vary from data set to data set, the best-fitting B was 12.8 dB. As expected, the goodness of fit decreased, but was still acceptable at 9.8 rau.

Equation (7) calculates the variance of the sum of all audibility noises (i.e., across the entire frequency range of the model). When the $\mathrm{SNR_E}$ varies across frequency, audibility-noise variances for the individual model bands must be determined. The audibility-noise variance, σ_{Ni}^2, in a band ranging from the lower band edge, f_i, to the higher band edge, f_j, is calculated as

$$10 \cdot \log_{10}(\sigma_{Ni}^2) = -\mathrm{SNR_E} + B + 10 \cdot \log_{10}\left(\frac{z_j - z_i}{19\ \mathrm{Barks}}\right), \tag{8a}$$

$$\sigma_N^2 = b \cdot \frac{n}{s} \cdot \left(\frac{z_j - z_i}{19\ \mathrm{Barks}}\right), \tag{8b}$$

where z_i and z_j are the critical band rates (expressed in Barks) at f_i and f_j, respectively. The values for B and b are as in Eq. (7).

Three data points had to be excluded from the fits just discussed, because they represent extreme performances that would have made fits meaningless. One is the perfect score (100% correct) at 9 dB SNR in Grant and Walden's (1996) data. According to the SRS model, perfect performance is impossible to achieve because it requires that all performance-limiting noises in the model are zero, resulting in an infinite d'. Note, however, that the model can predict performance arbitrarily close to 100% correct. The data points at -15 dB SNR were also excluded from the fit. At this SNR chance performance was reported, implying a d' of zero and an audibility noise with infinite variance.

VI. INTERIM DISCUSSION

A linear relation between $\mathrm{SNR_E}$ and $10 \log_{10}(\sigma_N^2)$, such as Eq. (7), is very appropriate for modeling the data. This is evidenced by the high negative correlation between the two variables (see Table II). For every data set, the correlation is -0.976 or better. No systematic variations from a straight-line model are evident, except perhaps in the data of Rankovic and Levy (1997). There are two possible explanations for this. First, Rankovic and Levy (1997) only reported the ranges of their data, but not the group means. We treated the centers of these ranges as means and Fig. 1 as well as the predictions in Fig. 3 are based on these interpolated values. The systematic deviation from a straight line in Fig. 3 for the Rankovic and Levy (1997) data may simply reflect that the range is skewed around the true unknown mean. Second, even though all studies used "speech-spectrum-shaped" noise, there are differences in the accuracy with which the noise spectra match the speech spectra. Whenever there is a mismatch, the performance versus SNR function will extend over a wider range of SNRs and assume a different shape than when the noise and speech spectra match perfectly. Rankovic and Levy (1997) appear not to have shaped their noise to match the spectrum of their speech material (see Table I), but instead used a generic speech spectrum. The mismatch between the spectra that was likely to result may

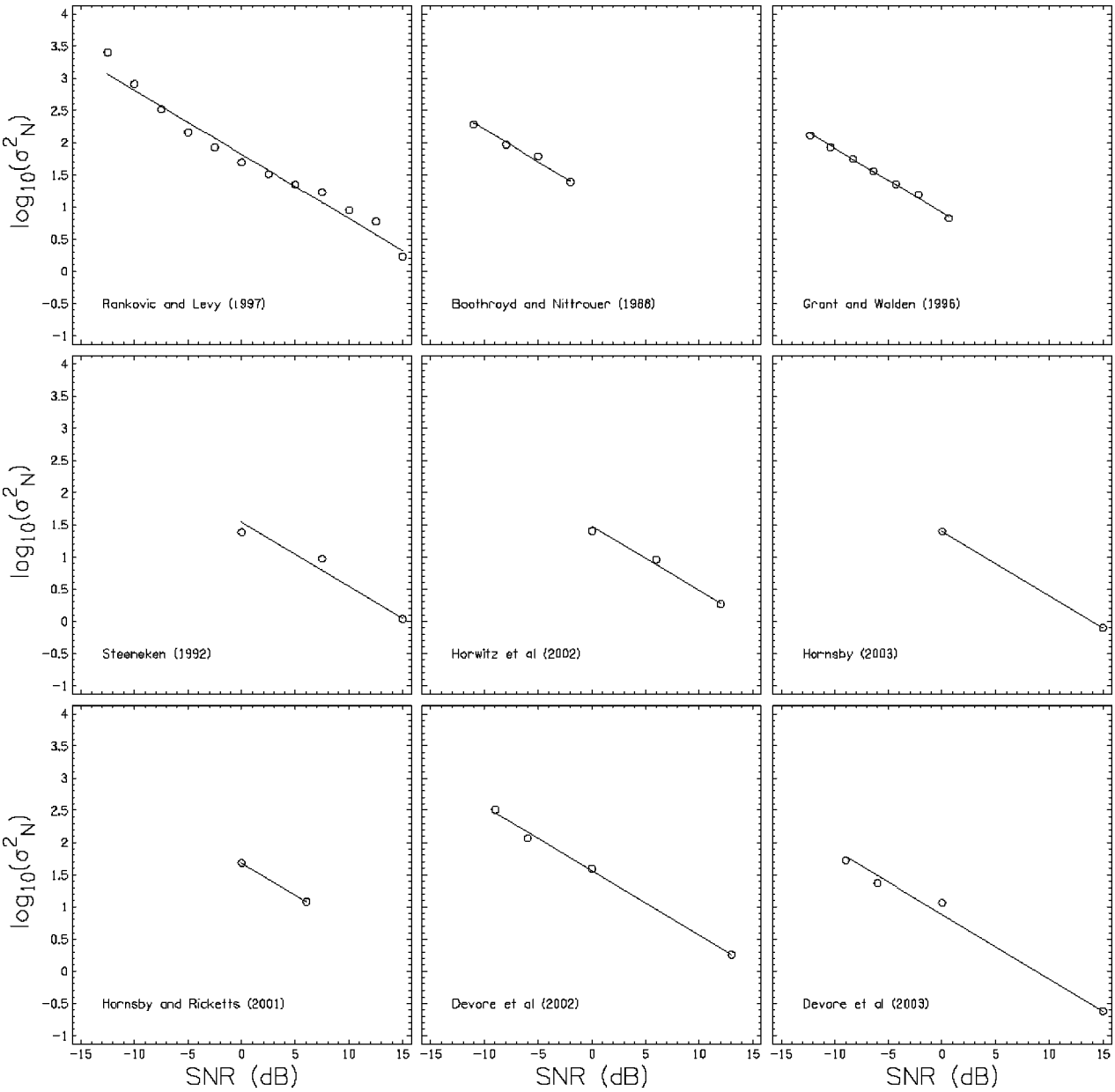


FIG. 3. Relation between SNR and the logarithm of the audibility–noise variance. The lines show the SRS model assumption [Eq. 7(a)] and the symbols show the relation between SNR and the logarithm of the audibility–noise variance that predicts the measured performances.

be one cause for the ''curvature'' in their transformed data (see Fig. 3). Despite these issues, Eq. (7) nevertheless provides a good approximation to their data.

In general, it is likely that there was some degree of mismatch between the speech and masker spectra in all studies except those that used talker-matched noise (cf. Table I). Therefore, the assumptions that allowed the simplifications of Eq. (5) were not met completely. This could be part of the reason for our failure to achieve a good fit when the parameter B was forced to be the same in all studies. Accounting for the spectral mismatch would have been possible only for those studies where detailed information of the spectra was available. But, perhaps more importantly, it would have required specifying a frequency-dependent congruence index density function and a frequency-dependent SNR_E, which would have increased the number of free parameters drastically. We felt that ignoring the residual SNR_E variation across frequency to keep the number of free parameters small was justified for the purpose of the current investigation.

The studies by Hornsby (2003) and Hornsby and Ricketts (2001) provide only two data points each. Although these fits may seem trivial, there was no guarantee that these data could be fit. First, the best-fitting slope might not conform to that of -1 found to fit the data set as a whole. Second, X is a variance and is therefore constrained to be non-negative. The fact that the data could be fit and that the best-fitting parameters are similar to those fitting the other studies makes these data sets interesting despite the small number of data points.

VII. ADDITIONAL PREDICTIONS

In the previous section, predictions were restricted to broadband conditions because these allowed the simplifica-

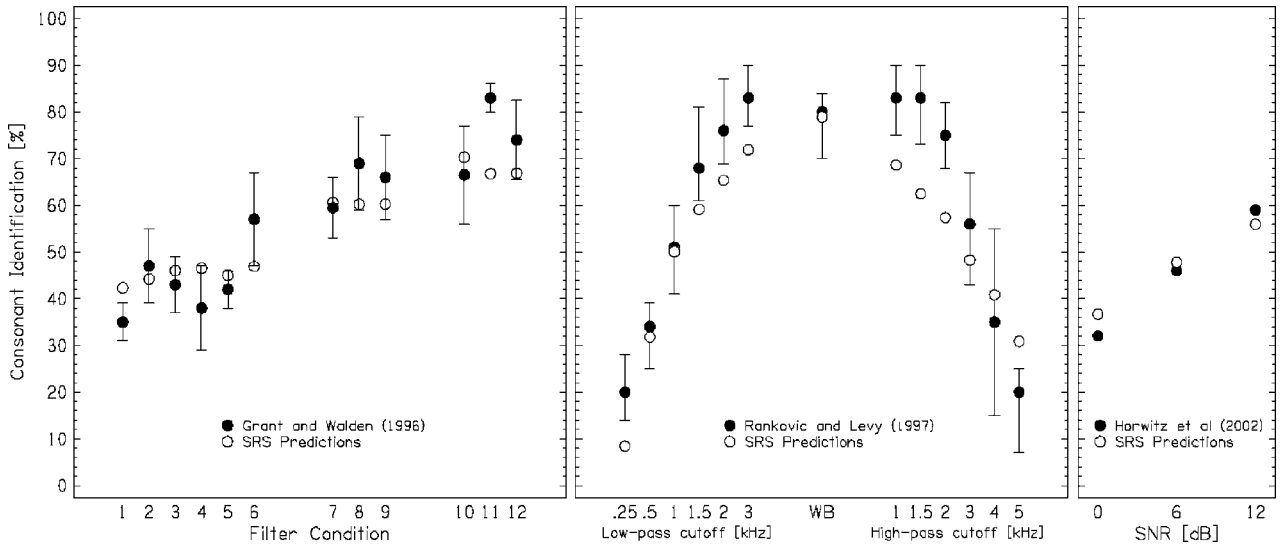


FIG. 4. Measured (filled circles) and predicted (open circles) consonant-recognition performance of band-limited speech. See text for details.

tions that reduced Eq. (4) to Eq. (5). However, four studies (Steeneken, 1992; Grant and Walden, 1996; Rankovic and Levy, 1997; Horwitz *et al.*, 2002) also measured consonant recognition of band-limited speech. The SRS model should be able to predict these data as well. Moreover, it should do so with model parameters that agree with those found in Sec. V. To verify this expectation, we predict the band-limited portions of the aforementioned data sets in this section.

A. Method

The fits in Sec. V specified the SRS model parameters for every data set almost completely. The only degree of freedom left is the relative contribution of production- and cognitive-noise variances to their already-established sum, X. The relative sizes of production- and cognitive-noise variances were inconsequential for predictions of broadband conditions where all bands are audible. This was one reason why Eq. (4) could be reduced to Eq. (5). However, the relative sizes of production- and cognitive-noise variances do affect predictions when speech bands are eliminated through high-pass, low-pass, band-pass, or band-stop filtering. When a signal is filtered, the production-noise power decreases together with the bandwidth, but the power of the cognitive noise remains constant. To predict how much the combined noise X decreases as a result of filtering, the relative contributions of cognitive noise and production noise must be known.

The production-noise variance is controlled by a single parameter, the inherent d'_p, which describes the talker's speech-production ability. Because d'_p is thought to characterize the talker, it does not change with signal modifications, such as filtering or adding background noise. Consequently, only one such value can be selected for each data set. However, different values can be used for different data sets because they were recorded by different talkers. The production noise $\sigma_p^2(i)$ in the frequency band i is related to the inherent d'_p through the congruence index in that band, Δ_i:

$$\sigma_p^2(i) = \left(\frac{\Delta_i}{d'_p}\right)^2. \tag{9}$$

The overall production-noise power is derived by combining the band-production noises of all audible bands, taking into account that the band-production noises are partially correlated.[4]

One value of d'_p was selected for each study to minimize the RMS difference between the observed and predicted consonant-recognition performances. Because X, σ_{cog}^2, and d'_p are interrelated and because X was already fixed in Sec. VI, the selection of d'_p also determined a corresponding value of the cognitive-noise variance, σ_{cog}^2. In the fits, the range of d'_p was constrained to yield positive values of σ_{cog}^2. Of course, the same result could have been obtained by adjusting the variance of the cognitive noise and letting the inherent d'_p float. Unless otherwise stated, the values of X and B are those of Table II. All predictions were carried out with a discrete approximation of the SRS model with a resolution of 1/10 Bark, using the same congruence-index density function that was used by Müsch and Buus (2001a,b). It is proportional to the importance function of the SII.

B. Results

1. Grant and Walden (1996)

Grant and Walden's (1996) consonant-recognition scores of band-pass filtered speech in quiet, along with the SRS-model predictions, are shown in the left panel of Fig. 4. The filled circles show the measured scores in percent correct (ordinate) for each of 12 filtering conditions (abscissa). The error bars indicate ± one standard deviation. These values are replotted from Grant and Walden's (1996) Fig. 6. The open circles are the best-fitting SRS-model predictions. The edge frequencies of these filters are described in Fig. 2 of Grant and Walden (1996). Briefly, conditions 1–6 were narrow filters that were expected to yield an AI of 0.1. The

filters of conditions 7–9 were wider to yield an AI of 0.2, and the filters of conditions 10–12 were wider still to yield an AI of 0.3.

The best-fitting inherent d'_p is 2.1, which implies a cognitive-noise variance, σ^2_{cog}, of 0.59. The resulting RMS error of the prediction is 7.7 rau.

2. Rankovic and Levy (1997)

Rankovic and Levy (1997) measured consonant recognition in quiet for six low-pass and six high-pass filter conditions in addition to the broadband conditions predicted earlier. The filled circles in the center panel of Fig. 4 show the measured scores in percent correct (ordinate) as read from Rankovic and Levy's (1997) Fig. 6. The error bars indicate the range between the highest and lowest measured scores, as read from their Table I. The open circles represent the best-fitting SRS-model predictions. The six data points on the left of the panel are for low-pass conditions and the six data points on the right are for high-pass conditions. The cutoff frequencies are indicated on the abscissa. The datum in the middle is the broadband condition at 15 dB SNR. This datum was also included in Fig. 1. The best-fitting inherent d'_p is 1.63, which implies a cognitive-noise variance, σ^2_{cog}, of 0.96 and yields an RMS error of 11.9 rau.

The data indicate that broadband speech at an SNR of +15 dB yields a lower mean performance than speech in quiet that is high-pass filtered at 1 or 1.5 kHz or low-pass filtered at 3 kHz. This contradicts the widely held assumption that speech bands at an SNR of 15 dB or higher provide their maximum intelligibility potential (cf. ANSI, 1997). Recently, Studebaker and Sherbecoe (2003) showed that this assumption may be incorrect and intelligibility contributions increase beyond 15 dB SNR. In agreement with this finding, the audibility-noise variance continues to decrease [see Eq. (7)] and performance continues to increase as the SNR exceeds 15 dB, but the predicted performance increase is very small, resulting in predictions in which all filter conditions yield lower intelligibility scores than the broadband condition at 15 dB SNR.

3. Horwitz et al. (2002)

The previous studies applied various amounts of band limiting to speech in quiet. In contrast, Horwitz *et al.* (2002) used only one filter condition (low-pass filtering at 1.78 kHz), but presented the filtered speech at three different SNRs (0, 6, and 12 dB SNR). The data and model predictions are shown in the right panel of Fig. 4. The filled circles show the measured performances. The SRS-model predictions are illustrated by open circles.

The best-fitting inherent d'_p of 2.34, which implies a cognitive-noise variance of 9.22, results in an RMS error of 3.3 rau. This cognitive-noise variance appears to be unrealistically high. Fortunately, good fits, although suboptimal, can be obtained with more realistic values of the cognitive-noise variance. For example, an inherent d'_p of 1.99, which implies $\sigma^2_{cog}=1.00$—the value normally assumed for nonsense syllables—results in an RMS error of 6.7 rau when the parameter B is increased by 1.3 dB (from 14.8 to 16.1 dB). This RMS error is calculated across both the filtered and broadband conditions to account for the fact that changing B also affects the predictions for the broadband condition.

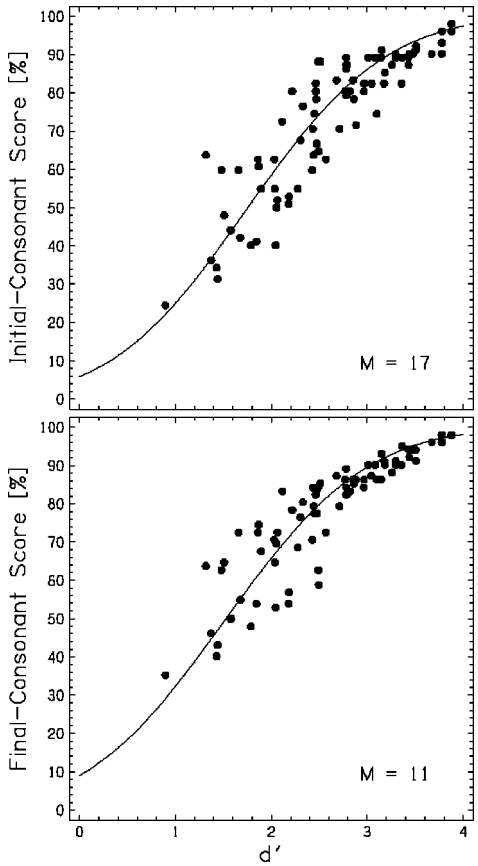


FIG. 5. Prediction of initial (top) and final (bottom) consonant recognition in 78 different listening conditions. Data are from Steeneken (1992). The symbols show the measured performance plotted at the calculated d'. The lines are the model predictions, and the scatter around the lines shows the prediction error. See text for details.

4. Steeneken (1992)

Steeneken's (1992) study includes 26 filter conditions tested at three SNRs each, providing data for a total of 78 different listening conditions.[5] The filter conditions varied in shape, ranging from simple high- and low-pass conditions to conditions with multiple, spectrally disconnected passbands. Passbands as well as stopbands were always at least one octave wide. The SNRs were 15, 7.5, and 0 dB.

The results of the SRS-model fits are shown in Fig. 5. There are two sets of results because initial and final consonants of the nonsense CVCs were scored separately and had different response-set sizes. The symbols in the upper panel show the measured initial-consonant scores as a function of the recognition sensitivity, d'. The symbols in the lower panel show the same relation for the final-consonant scores. Because the recognition sensitivity, d', does not depend on whether initial or final consonants are scored, every data point in the upper panel has a corresponding data point at the same d' in the lower panel.[6] The solid lines are graphical representations of Eq. (3) for $M=17$ (top) and $M=11$ (bottom), reflecting different numbers of initial and final consonants to be recognized. These transformations are specified

by the model structure and cannot be adjusted to match the data. As for the previous data sets, the model was fit by adjusting the inherent d'_p. The best-fitting value is $d'_p = 2.44$, which implies a cognitive-noise variance of 2.48. The resulting RMS error is 8.27 rau.

VIII. DISCUSSION AND CONCLUSION

In the predictions of Müsch and Buus (2001a,b) the variance of the cognitive noise for nonsense syllables was assumed to be 1.00. The variances of the cognitive noises that yield the best predictions in Sec. VII are 0.59, 0.96, and 2.48 for the studies of Grant and Walden (1996), Rankovic and Levy (1997), and Steeneken (1992), respectively. For the data of Horwitz *et al.* (2002), good predictions could be achieved with a cognitive-noise variance of 1.00, although the best-fitting variance was 9.22. Overall, these values are in good agreement with the assumptions of our earlier work.

Likewise, the predictions of Müsch and Buus (2001a,b) assumed that the inherent d'_p of a typical talker is 3.5. The best-fitting inherent d'_p's of this study are 1.6, 2.1, 2.3 (2.0), and 2.4, where the value in parentheses indicates the d'_p for the suboptimal fit to Horwitz *et al.*'s (2002) data. These values are somewhat lower than the d'_p used in our previous work, but are largely consistent with it.

If predictions are to be made for a single talker or a small group of talkers (as in the case of the data modeled here), the production-noise variance must be determined experimentally by applying the SRS model to intelligibility data collected with the respective talker(s). Only when the talker is representative of the entire population of talkers will it be possible to use a generic production-noise variance. The sum of our modeling so far indicates that typical production noises yield inherent d'_p values between 1.5 and 3.5. The ''true'' range may be somewhat smaller as some of the variability may have resulted from accidental tradeoffs. At this point we feel that we do not have enough experience with the model to suggest a d'_p to be used when predicting intelligibility of a generic talker.

The data modeled here exhibit a wide range of performance levels for quite similar test conditions. Accounting for these data with a universal set of parameters will be hard for any intelligibility-prediction paradigm as long as the causes for these differences are not identified. One known cause for performance differences is the number of response alternatives. The SRS model accounts for this effect and reduces the spread of performances somewhat (compare Figs. 1 and 2). However, a large spread remains.

In many instances we had to make assumptions about certain aspects of the data because not all the information needed for the predictions was reported. We preferred this to limiting ourselves to a very small set of data tailored to suit the input requirements of the SRS model. To be useful, the SRS model must predict the full spectrum of data that is being collected in practice. With the predictions presented here, we believe the SRS model has come one step closer to this goal.

IX. SUMMARY

Consonant-recognition performance in speech-spectrum-shaped noise, as reported in nine studies, was used to derive the function relating the SRS model's audibility-noise variance to the signal-to-noise excitation ratio in the auditory periphery, $\mathrm{SNR_E}$. It was found that a linear function was well suited to describe the relation between the $\mathrm{SNR_E}$ (in dB) and ten times the logarithm of the audibility-noise variance, $10 \cdot \log_{10}(\sigma_N^2)$. The best-fitting slope was very close to -1.0, which indicates that the audibility noise in the SRS model is almost directly proportional to the noise-to-signal excitation-power ratio in the auditory periphery. This finding suggests that Eq. (8) replace the relation between SNR and audibility noise proposed by Müsch and Buus (2001a,b).

Four of the nine studies also measured intelligibility under filter conditions. These conditions were also predicted. Only one free parameter per study was available for these fits because the model was constrained in the earlier fits of the broadband conditions. Despite these constraints, the SRS model predicted performance under filtering adequately, and did so with parameters that were internally consistent and consistent with the parameters used in Müsch and Buus (2001a,b).

ACKNOWLEDGMENTS

We thank Sasha Devore and Ben Hornsby for allowing access to their data and Mary Florentine for in-depth discussions. We also thank the four reviewers for comments that helped improve the presentation of the material. This work was funded by NIH Grant No. R01DC00187.

[1]The SII provides for speech material-dependent importance functions, which, in principle, can predict different word scores for listening conditions that yield identical nonsense-syllable scores. However, a detailed analysis shows that differences in the importance function do not account for the effects observed by Hirsh *et al.* (1954).

[2]This notation makes use of the assumption that the signal-to-noise ratio in the auditory periphery, $\mathrm{SNR_E}$, equals the acoustic speech-to-noise ratio ($\mathrm{SNR_E} = \mathrm{SNR}$). Because we assume this equivalence for the data discussed in this paper, $\mathrm{SNR_E}$ and SNR are often used interchangeably.

[3]This equation is identical to Eq. (7) of Müsch and Buus (2001a). See their Fig. 2 for a flow diagram illustrating this equation in a more intuitive way.

[4]The correlation between bands i and j, r_{ij}, depends on their spectral separation. The correlation is high when the bands are spectrally close and small when the bands are spectrally distant. It is quantified by $r_{i,j} = e^{-0.0766(z_i - z_j)^2}$ where z_i and z_j are the center frequencies, expressed in Bark, of bands i and j, respectively. The constant 0.0766, corresponding to an equivalent rectangular width of 3.2 Bark, was found to be optimal for describing four sets of data in Müsch and Buus (2001a,b) and is used throughout this article [see also the central term in Eq. (4)].

[5]Steeneken's (1992) study also contains a smaller data set with female talkers only. We did not attempt to predict this smaller set.

[6]There is some evidence that initial consonants might be recognized more accurately than final consonants and that difference might be attributed to a better production of the initial consonants (Redford and Diehl, 1999). If this was indeed the case, then it could be argued that different production-noise variances should be used to predict performance with the two types of consonants. However, no such trend was seen in Steeneken's (1992) data. When the difference in response-set size was accounted for, the difference between initial and final consonants was nonsignificant ($p = 0.27$ in two-tailed t test). Accordingly, we did not adjust model parameters to model initial- and final-consonant recognition. This is also in keeping with the ultimate goal of the SRS model, which is to provide a macroscopic model of speech intelligibility.

ANSI (**1969**). ''Methods for the calculation of the articulation index,'' ANSI S3.5-1969 (American National Standards Institute, New York).
ANSI (**1997**). ''Methods for calculation of the speech intelligibility index,'' ANSI S3.5-1997 (American National Standards Institute, New York).
Beranek, L. L. (**1947**). ''The design of speech communication systems,'' Proc. Inst. Radio Eng. **35**, 880–890.
Boothroyd, A., and Nittrouer, S. (**1988**). ''Mathematical Treatment of context effects in phoneme and word recognition,'' J. Acoust. Soc. Am. **84**, 101–114.
Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Kotby, M. N., Nasser, H. A., Wafaa, A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G., Westerman, S., and Ludvigsen, C. (**1994**). ''An international comparison of long-term average speech spectra,'' J. Acoust. Soc. Am. **96**, 2108–2120.
Devore, S., and Shinn-Cunningham, B. G. (**2003**). ''Perceptual consequences of including reverberation in spatial auditory displays,'' in Proceedings of the International Conference on Auditory Displays, Boston, MA, 6–9 July, 2003, pp. 75–78.
Devore, S., Shinn-Cunningham, B. G., Durlach, N. I., and Colburn, H. S. (**2002**). ''The influence of reverberation on spatial release of masking in consonant identification,'' J. Acoust. Soc. Am. **111**, 2422.
Fletcher, H., and Galt, R. H. (**1950**). ''The perception of speech and its relation to telephony'' J. Acoust. Soc. Am. **22**, 89–151.
French, N. R., and Steinberg, J. C. (**1947**). ''Factors governing the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **19**, 90–119.
Grant, K., and Walden, B. E. (**1996**). ''Evaluating the articulation index for auditory-visual consonant recognition,'' J. Acoust. Soc. Am. **100**, 2415–2424.
Green, D. M., and Birdsall, T. G. (**1958**). ''The effect of vocabulary size on articulation score,'' Technical Memorandum No. 81 and Technical Note AFCRC-TR-57-58, University of Michigan, Electronic Defense Group.
Hirsh, I. J., Reynolds, E. G., and Joseph, M. (**1954**). ''Intelligibility of different speech materials,'' J. Acoust. Soc. Am. **26**, 530–538.
Hornsby, B. W. Y. (**2003**). Personal communication.
Hornsby, B. W. Y., and Ricketts, T. A. (**2001**). ''The effects of compression ratio, signal-to-noise ratio, and level on speech recognition in normal-hearing listeners,'' J. Acoust. Soc. Am. **109**, 2964–2973.
Horwitz, A. R., Dubno, J. R., and Ahlstrom, J. B. (**2002**). ''Recognition of low-pass-filtered consonants in noise with normal and impaired high-frequency hearing,'' J. Acoust. Soc. Am. **111**, 409–416.
Kryter, K. D. (**1962a**). ''Methods for the calculation and use of the articulation index,'' J. Acoust. Soc. Am. **34**, 1689–1697.
Kryter, K. D. (**1962b**). ''Validation of the articulation index,'' J. Acoust. Soc. Am. **34**, 1698–1702.
Licklider, J. C. R. (**1959**), ''Three auditory theories,'' in *Psychology: A Study of A Science*, Vol. 1, edited by H. Koch (McGraw–Hill, New York), pp. 41–144.
Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). ''The intelligibility of speech as a function of the context of the test material,'' J. Exp. Psychol. **41**, 329–335.
Müsch, H., and Buus, S. (**2001a**). ''Using statistical decision theory to predict speech intelligibility. I. Model structure,'' J. Acoust. Soc. Am. **109**, 2896–2909.
Müsch, H., and Buus, S. (**2001b**). ''Using statistical decision theory to predict speech intelligibility. II. Measurement and prediction of consonant-discrimination performance,'' J. Acoust. Soc. Am. **109**, 2910–2920.
Rankovic, C. M., and Levy, R. M. (**1997**). ''Estimating articulation scores,'' J. Acoust. Soc. Am. **102**, 3754–3761.
Redford, M. A., and Diehl, R. L. (**1999**). ''The relative perceptual distinctiveness of initial and final consonants in CVC syllables,'' J. Acoust. Soc. Am. **106**, 1555–1565.
Resnick, S. B., Dubno, J. R., Hoffnung, S., and Levitt, H. (**1975**). ''Phoneme errors on a nonsense syllable test,'' J. Acoust. Soc. Am. **58** (Suppl. 1), S114.
Steeneken, H. J. M. (**1992**). ''On measuring and predicting speech intelligibility,'' doctoral dissertation, Soesterberg, The Netherlands.
Steeneken, H. J. M., and Houtgast, T. (**1980**). ''A physical method for measuring speech-transmission quality,'' J. Acoust. Soc. Am. **69**, 318–326.
Steeneken, H. J. M., and Houtgast, T. (**1999**). ''Mutual dependence of octave-band weights in predicting speech intelligibility,'' Speech Commun. **28**, 109–123.
Studebaker, G. A. (**1985**). ''A 'rationalized' arcsine transformation,'' J. Speech Hear. Res. **28**, 455–462.
Studebaker G. A., and Sherbecoe R. L., (**2002**). ''Intensity-importance functions for bandlimited monosyllabic words,'' J. Acoust. Soc. Am. **111**, 1422–1436.
van Rooij, J. C. G. M., and Plomp, R. (**1991**). ''The effect of linguistic entropy on speech perception in noise in young and elderly listeners,'' J. Acoust. Soc. Am. **90**, 2985–2991.
van Wijngaarden, S., and Houtgast, T. (**2004**). ''Effect of talker and speaking style on the Speech Transmission Index,'' J. Acoust. Soc. Am. **115**, 38–41.
Zwicker, E., and Feldtkeller, R. (**1999**). *The Ear as a Communication Receiver* (Acoustical Society America, New York).

Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights[a)]

Joshua M. Alexander[b)] and Robert A. Lutfi
Department of Communicative Disorders and Waisman Center, University of Wisconsin, Madison, Wisconsin 53706

(Received 23 October 2003; revised 23 June 2004; accepted 28 June 2004)

Informational masking (IM) refers to elevations in signal threshold caused by masker uncertainty. The purpose of this study was to investigate two factors expected to influence IM in hearing-impaired listeners. Masked thresholds for a 2000-Hz signal in the presence of simultaneous multitone maskers were measured in 16 normal-hearing (NH) and 9 hearing-impaired (HI) listeners. The maskers were 70 dB SPL average total power and were comprised of fixed-frequency components between 522 and 8346 Hz that were separated from each other by at least $\frac{1}{3}$ oct and from the signal by at least $\frac{2}{3}$ octs. Masker uncertainty was manipulated by randomly presenting each masker component with probability $p=0.1,0.2,...,0.9$, or 1.0 across different trial blocks. Energetic masking was estimated as the amount of masking for $p=1.0$, where masker uncertainty was minimum. IM was estimated as the amount of masking in excess of energetic masking. Decision weights were estimated by a regression of the listener's yes/no responses against the presence or absence of the signal and masker components. The decision weights and sensation levels (SLs) of the stimulus components were incorporated as factors in a model that predicts individual differences in IM based on the level variance (in dB) at the output of independent auditory filters [Lutfi, J. Acoust. Soc. Am. **94**, 748–758 (1993)]. The results showed much individual variability in IM for the NH listeners (over 40 dB), but little IM for most HI listeners. When masker components were presented to a group of NH listeners at SLs similar to the HI listeners, IM was also similar to the HI listeners. IM was also similar for both groups when the level per masker component was 10 dB SL. These results suggest that reduced masker SLs for HI listeners decrease IM by effectively reducing masker variance. Weighting efficiencies, computed by comparing each listener's pattern of weights to that of an ideal analytic listener, were a good predictor of individual differences in IM among the NH listeners. For the HI listeners weighting efficiency and IM were unrelated because of the large variation in masker SLs among individual listeners, the small variance in IM, and perhaps because broadened auditory filters in some listeners increased the covariance in auditory filter outputs. [DOI: 10.1121/1.1784437]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Sr [GK] Pages: 2234–2247

I. INTRODUCTION

Informational masking (IM) is the term often used to refer to elevations in signal threshold that cannot be attributed to the energy-based masking at the filters of the auditory periphery. One factor known to produce considerable amounts of IM is uncertainty regarding the spectral properties of the masker. According to this view, IM can be quantified by the difference in masked threshold for maskers whose properties vary unpredictably from trial to trial and maskers with unvarying properties (cf. Watson *et al.*, 1976). Since signal detection for the latter type of masker is limited primarily by energetic masking, IM is equivalent to the residual amount of masking after accounting for the contribution of energetic masking:

$$\mathrm{IM}=\mathrm{TM}-\mathrm{EM}, \tag{1}$$

where TM is the total masking and EM is the energetic masking, all in dB (cf. Lutfi, 1990).

For multitone maskers with random frequencies, pure-tone detection thresholds can be elevated by 50 dB for normal-hearing (NH) adult listeners (Neff and Green, 1987; Neff and Dethlefs, 1995; Oh and Lutfi, 1998) or greater for preschool children (Allen and Wightman, 1995; Oh *et al.*, 2001; Lutfi *et al.*, 2003b; Wightman *et al.*, 2003). Compared to these groups, relatively little is known about the factors that influence IM in hearing-impaired (HI) listeners. To the authors' knowledge, only three other studies (Kidd *et al.*, 2001; Doherty and Lutfi, 1999; Micheyl *et al.*, 2000) have investigated IM for a pure-tone signal in HI listeners.

Kidd *et al.* (2001) adapted the levels of random-frequency multitone maskers in a two-interval, forced-choice procedure to obtain the signal-to-masker ratio at masked threshold in NH and HI listeners. The signal was a sequence of eight contiguous 60-ms tone bursts at 750 or 1000 Hz and was fixed at 20 dB above the signal threshold in quiet. The maskers consisted of two sets of tone bursts, one set below and one set above the signal frequency, and were played synchronously with the eight signal bursts. For the

[a)]Portions of this research were presented at the 25th and 26th Midwinter Research Meetings of the Association for Research in Otolaryngology: St. Petersburg Beach, FL, January 2002 [Assoc. Res. Otolaryngol. Abs.: 180]; Daytona Beach, FL, February 2003 [Assoc. Res. Otolaryngol. Abs.: 115].

[b)]Electronic mail: jmalexal@wisc.edu

''multiple-bursts different'' (MBD) maskers, the frequencies of the masker bursts varied from burst to burst within each of the two trial intervals. For the ''multiple-burst same'' (MBS) maskers, the masker bursts were kept at constant frequencies within each interval, like the signal, but were varied across the two intervals in each trial. Kidd *et al.* hypothesized that the signal would perceptually segregate from the MBD maskers since their spectral-temporal properties were different from the signal, thereby lowering detection thresholds. It was found that the signal-to-masker ratio increased for both types of maskers as hearing loss for the signal increased and that this factor accounted for most of the variance in the masked thresholds. The signal-to-masker ratios for the MBS maskers were high even for the NH listeners and increased only slightly as a function of hearing loss. The signal-to-masker ratios for the MBD maskers were much less than they were for the MBS maskers for both groups, but steadily increased with increasing hearing loss so that the difference between the two masker types was minimal for the listeners with the greatest amounts of hearing loss. Kidd *et al.* speculated that the relative increase in signal-to-masker ratios for the MBD maskers with increasing amounts of hearing loss for the signal signified that ''hearing loss adversely affects the ability to *listen analytically*'' (p. 118).

To listen analytically a listener must attend to a specific target feature of the stimulus (signal) and ignore all other nontarget features (maskers). Analytic listening can be operationally described by decision weights that denote the emphasis or attention that each feature is given in a listener's decision. This provides a quantitative method for describing and comparing different listening strategies. For a specified stimulus parameter, frequency in this case, decision weights are analogous to an attentional filter; spectral regions given greater weight will tend to have greater influence on a listener's decision than those given less weight. Decision weights are often estimated by correlation or multivariate regression coefficients that describe the relationship between the listener's responses and the properties of the stimulus ensemble, assuming a particular decision model (cf. Berg, 1989; Lutfi, 1995; Richards and Zhu, 1994).

In order to describe the association between the masked threshold for a random-frequency masker and decision weights, we adopt a theoretical framework in which the listener is assumed to make decisions about the presence or absence of the signal by forming a weighted sum of the output of independent auditory filters (Lutfi, 1993). Using sensation level, SL (dB above threshold in quiet), as the filter output, the decision variable, D, is

$$D=\sum_{i=1}^{n} w_i \mathrm{SL}_i, \tag{2}$$

where i corresponds to individual signal and masker tones and where w_i represents the decision weights. In this equation, weights and SLs are both linearly related to D so that if a masker component is half the SL of another masker component but is given twice as much weight, then both components will have an equal effect on the magnitude and the variance of D. For a yes/no task it is assumed that the listener responds ''yes'' if and only if D exceeds some constant internal criterion, C, and responds ''no'' otherwise. Models of this type have proven quite successful in predicting the amount of IM obtained in conditions like those of the present study (Lutfi, 1992, 1993; Oh and Lutfi, 1998; Lutfi *et al.*, 2003b; Wright and Saberi, 1999; Richards *et al.*, 2002; Tang and Richards, 2002).

One such model is the component relative entropy (CoRE) model (Lutfi, 1993). Oh and Lutfi (1998) develop specific predictions from this model for the case in which the masker is a combination of tones with random frequencies. The prediction is that IM will be proportional to the standard deviation σ in level (dB) at the output of a typical auditory filter,

$$\mathrm{IM}\cong d'\sqrt{n}\sigma, \tag{3}$$

where d' is the index of sensitivity and n is a free parameter representing the number of independent auditory filters assumed to be involved in a listener's decision. In this study, the contribution of independent auditory filters (given by n) is obtained directly from w_i, rather than treated as a free parameter. The standard deviation of filter outputs, moreover, is derived from the SLs of the individual masker components so that the predicted amount of IM is then given by

$$\mathrm{IM}\cong d'\sqrt{\sum w_i^2 \mathrm{SL}_i^2}\,\sigma. \tag{4}$$

From Eq. (4) it is apparent that increasing the weights and/or the SLs of masker components also increases the standard deviation of the decision variable, which increases the amount of IM.

The ability to listen ''analytically'' in this context can be evaluated by comparing a listener's decision weights to those of an ideal observer—an observer that would give a weight of one to the signal and zero weight to all masker components. Berg (1990) describes a weighting efficiency measure that represents the degree to which a listener's weights approaches the ideal weights. In a task in which listeners discriminated the difference in level of a designated signal tone in the presence of level-varying nonsignal tones, Doherty and Lutfi (1999) found that weighting efficiency for listeners with mild to moderate, sloping hearing loss tended to increase as the frequency of the signal and the amount of hearing loss increased. If maskers in a region of hearing loss receive greater weight, as these results indicate, then the amount of IM is expected to be larger.

The influence of SL on IM is inferred from a study by Micheyl *et al.* (2000). Micheyl and colleagues found that under certain circumstances HI listeners had lower masked thresholds for random-frequency multitone maskers than NH listeners. They measured masked threshold for a 1000-Hz signal in listeners with normal hearing and in listeners with symmetric or asymmetric hearing losses (i.e., same or different amounts of hearing loss for the two ears). The signal was gated on and off with four-tone maskers whose frequencies were pseudo-randomly selected from six possible fixed frequencies. For listeners with symmetric hearing loss, the masker levels were set to 40 dB SL. For listeners with asymmetric hearing loss, the masker levels were set to 40 dB SL in the better ear and adjusted for equal loudness in the ear

with greater hearing loss. Micheyl *et al.* found that there was not a statistically significant difference in the masked thresholds between the listeners with normal hearing, the listeners with symmetric hearing loss, and the better-hearing ear of those with asymmetric hearing loss. However, in the asymmetric hearing loss group, masked thresholds for the ears with a greater degree of hearing loss were significantly *less* than they were for the better-hearing ears and were significantly less than they were for the NH listeners. Micheyl *et al.* suggested that these differences might be due to the loudness balancing procedure used to adjust the masker levels for the listeners with asymmetrical hearing loss. Because of loudness recruitment, the SLs of the maskers were lower and averaged only 20 dB SL compared to 40 dB SL in the better-hearing ears. These outcomes support the hypothesis that IM is dependent on masker SLs.

In summary, the literature suggests two different interpretations for the difference in susceptibility of NH and HI listeners to IM. Kidd *et al.* (2001) suggest that HI listeners might have a specific deficit when analytically listening for a signal amongst a group of maskers with similar spectral-temporal characteristics. In terms of decision weights, this suggests that HI listeners might have less efficient weighting strategies when detecting a signal tone in a group of randomly occurring masker tones. The results of Micheyl *et al.* (2000), however, suggest that HI listeners might have lower amounts of IM because the SLs of the masker components will be less in regions of hearing loss, effectively reducing masker uncertainty. We test these interpretations in the following experiments.

II. EXPERIMENT I: EQUAL-SPL MASKER COMPONENTS

A. Method

1. Listeners

Sixteen NH listeners (3 males and 13 females) between the ages of 19 and 24 years, including the first author (NH14), and nine HI listeners (1 male and 8 females) between the ages of 23 and 79 years (median age of 35 years) completed the study.[1] Hearing losses were assumed to be primarily sensorineural since none of the HI listeners had an air-bone gap (i.e., a difference between air conduction and bone conduction thresholds) greater than 10 dB between 500 and 4000 Hz in the test ear. For all HI listeners, the extent of sensorineural hearing loss was about the same for both ears, except for listener HI8 who had a mild sensorineural hearing loss in the nontest ear and a moderate hearing loss in the test ear. Each listener used the right ear for the experiments, except for HI8 who used the left ear.

2. Stimuli

The signal and masker were computer generated with a 40-kHz sampling rate using the MATLAB® programming language (The MathWorks, Inc.) and a 16-bit sound card (Sound Blaster™ audio card; Creative Technology, Ltd.) before being passed through a programmable attenuator (Tucker-Davis Technologies, PA4). All stimuli were presented monaurally through a Sennheiser HD 520 (II) earphone.[2] The signal was a 2000-Hz tone that occurred with a probability of $p=0.5$ on the single interval trial and was gated on and off synchronously with the maskers with 10-ms $\cos^2$ onset/offset ramps for a total duration of 300 ms. On every trial, the phase of the signal and of each masker component was randomly selected from a rectangular distribution ($0-2\pi$ rad).

The maskers were designed to simplify assumptions about the processing occurring at the auditory periphery. In order to limit the amount of energetic masking, the frequencies of the masker components were remote from the signal frequency and only a small number of components were used. Furthermore, the masker components were separated in frequency from one another so that they would occur in largely independent auditory filters. The masker components were ten randomly occurring, fixed-frequency tones separated by 1/3 oct in the range from 522 to 8346 Hz (522, 657, 828, 1043, 1314, 3312, 4173, 5258, 6624, 8346), excluding the 2/3-oct region on either side of the signal so as to further limit the amount of energetic masking (cf. Neff and Green, 1987; Neff and Callaghan, 1988). Uncertainty or variability in the masker ensemble was created by independently varying the nominal probability of occurrence, $p=0.1,0.2,...,0.9$, or 1.0, of each masker component.[3] Since the masker components were either on or off, the analytical approximation for the standard deviation term in Eq. (4) is given by the standard deviation of the binomial distribution, $\sigma_b=\sqrt{p(1-p)}$.

The level of each masker component was adjusted so that the expected total power of the masker ensemble within the block of trials was always 70 dB SPL:

$$I_i=70-10\log_{10}(10p), \tag{5}$$

where I_i is the intensity level of each masker component in dB SPL and p is the probability of occurrence for each masker component for a given trial. Within a block of trials, both I_i and p were kept constant. Listeners HI1, HI6, and HI7 could not detect the signal above chance with the maskers at these levels. For these three listeners the masker levels were lowered by 10 dB so that average total power of the maskers was 60 dB SPL. Since the number of components in the masker complex and the actual total power varied from trial to trial, neither could be used as a reliable cue for signal detection.

3. Procedure

Listeners ran the experiments individually while seated in a double-walled, sound-attenuated chamber. A single-interval, yes/no procedure with visual feedback was used to estimate masked threshold for the signal in the presence of the multitone maskers. With this procedure, decision weights can be estimated by correlating a listener's yes/no responses with the presence/absence of the tones in the single interval (see the next section). The signal levels were adapted using a two-down, one-up decision rule which estimates the 70.7% point on the psychometric function (Levitt, 1971). The signal level was limited so that the combined total power of the signal and maskers was no greater than 90 dB SPL. When listeners reached this level and responded incorrectly, the

TABLE I. For the test ears of the individual HI listeners, quiet thresholds for the masker components and the 2000-Hz signal in dB SPL are provided along with their age. The mean quiet thresholds and standard deviations for the HI and NH listeners are also listed.

		Frequency (Hz)										
Listener	Age	522	657	828	1043	1314	2000	3312	4173	5258	6624	8346
HI1	21	49.2	50.9	44.5	36.6	31.6	39.3	19.5	20.8	27.2	31.3	35.2
HI2	23	38.2	39.8	39.9	38.0	33.2	41.7	38.3	33.8	31.9	46.2	61.0
HI3	35	18.4	19.6	18.8	9.6	17.7	4.7	10.5	24.8	42.7	62.5	65.6
HI4	46	31.9	44.0	55.4	55.8	50.8	47.2	39.0	33.1	38.5	40.6	41.4
HI5	41	34.1	37.5	38.8	49.6	55.1	50.8	42.1	43.2	47.6	70.8	74.1
HI6	28	39.9	41.8	44.1	50.8	67.0	70.8	60.8	76.5	77.5	76.5	87.0
HI7	76	31.8	35.7	34.4	34.2	34.2	42.1	45.8	51.9	60.8	71.3	85.7
HI8	41	71.8	78.3	71.1	70.6	61.8	47.8	45.6	31.0	22.2	38.8	47.0
HI9	28	52.5	61.7	59.9	66.4	64.1	64.5	48.8	49.5	52.4	59.6	71.1
Mean HI	37.7	40.9	45.5	45.2	45.7	46.2	45.4	39.0	40.5	44.5	55.3	63.1
SD HI	16.8	15.3	16.8	15.3	18.6	17.4	18.6	15.3	17.1	17.4	16.5	18.7
Mean NH	21.6	20.5	22.8	16.2	14.0	14.2	5.6	−0.3	6.3	7.3	15.7	24.4
SD NH	1.6	4.2	4.2	3.2	6.6	7.0	6.6	5.2	4.4	5.8	7.7	8.7

adaptive track remained constant until the listener obtained two consecutive correct responses.[4] Each trial block consisted of 12 reversals in the adaptive track and started with an initial step size of 4 dB that was reduced to 2 dB after the third reversal. Threshold for each block was determined by averaging the levels of the final eight reversals.

The masked threshold for each probability condition was determined by the mean masked threshold for five consecutive blocks of trials. Initial starting levels for each condition were 10 dB above the masked threshold estimated from at least one block of practice trials. Subsequent starting levels were set to 10 dB above the running average of all previous blocks. Each experimental session lasted about 45 min and consisted of one block of trials for each probability condition, the order of which was randomly permutated for each session for each listener. To minimize practice-effects, a block of trials for each probability condition was run as practice and was repeated until the listener could consistently perform above chance. Furthermore, the signal was presented in quiet before the start of every block to remind the listener of the signal characteristics.

Before masked thresholds were estimated, quiet thresholds for one block of trials were estimated for the signal and for each masker tone using the same response-feedback procedure described above. After completion of the IM experiment, quiet threshold estimates for two more trial blocks were obtained. Quiet threshold for each tone was based on the mean across the three trial blocks. Table I shows the quiet thresholds for each of the individual HI listeners along with their age. The mean quiet thresholds and standard deviations for the HI and NH listeners are also provided in Table I. The SLs of the signal and masker components for every trial were computed by subtracting the quiet threshold from the presentation level, which depended on p [cf. Eq. (5)].

4. Analyses

Experimental conditions in which the probability of occurrence for each masker component was less than $p=1.0$ were considered to be the IM conditions because the spectral composition of the masker was uncertain from trial to trial. The amount of IM in these conditions was obtained by subtracting the estimated proportion of masking attributable to energetic masking [cf. Eq. (1)]. One way to estimate the amount of energetic masking in this experiment is to measure signal threshold for each possible masker in isolation without uncertainty. The number of possible maskers makes this approach impractical. We have chosen instead to take the amount of masking for the $p=1.0$ condition as our estimate of energetic masking since the masker composition for this condition was constant from trial to trial. Because the tails of the auditory filters are thought to be relatively flat for frequencies remote from the signal (Patterson and Green, 1978; Patterson *et al.*, 1982) and because the average total power of the masker complex was constant for all the probability conditions, the average amount of masker energy falling in the auditory filter centered on the signal frequency (the amount of energetic masking) is also assumed to be roughly constant for all the probability conditions.

Predictions for IM were based on the analytical expression in Eq. (4) except that the standard deviation term was empirically derived from the actual trial-by-trial standard deviation of the decision variable values [cf. Eq. (2)] for trials in which the signal was absent.[5] Similarly, d' was empirically derived from the difference in a listener's Z-scores for the hit and false-alarm rates across all conditions and blocks.

It is assumed that a listener adopted the same weighting strategy throughout the experiment and that weights were fixed across the probability conditions (Lutfi *et al.*, 2003a). The weights in Eq. (4) were estimated by a logistic regression of a listener's trial-by-trial responses against the SLs of the signal and masker components taken across all conditions except $p=1.0$ (cf. Richards and Zhu, 1994; Lutfi, 1995).[6] Variation in masker SL arises because masker SL was adjusted for each condition to maintain a constant average total power [cf. Eq. (5)]. The variation in signal SLs arises because signal level was varied in the adaptive threshold track. When components were ''off'' for a given trial, SL was set to 0.

As has been done in past studies, the weights are normalized so that their magnitudes sum to unity (cf. Berg,

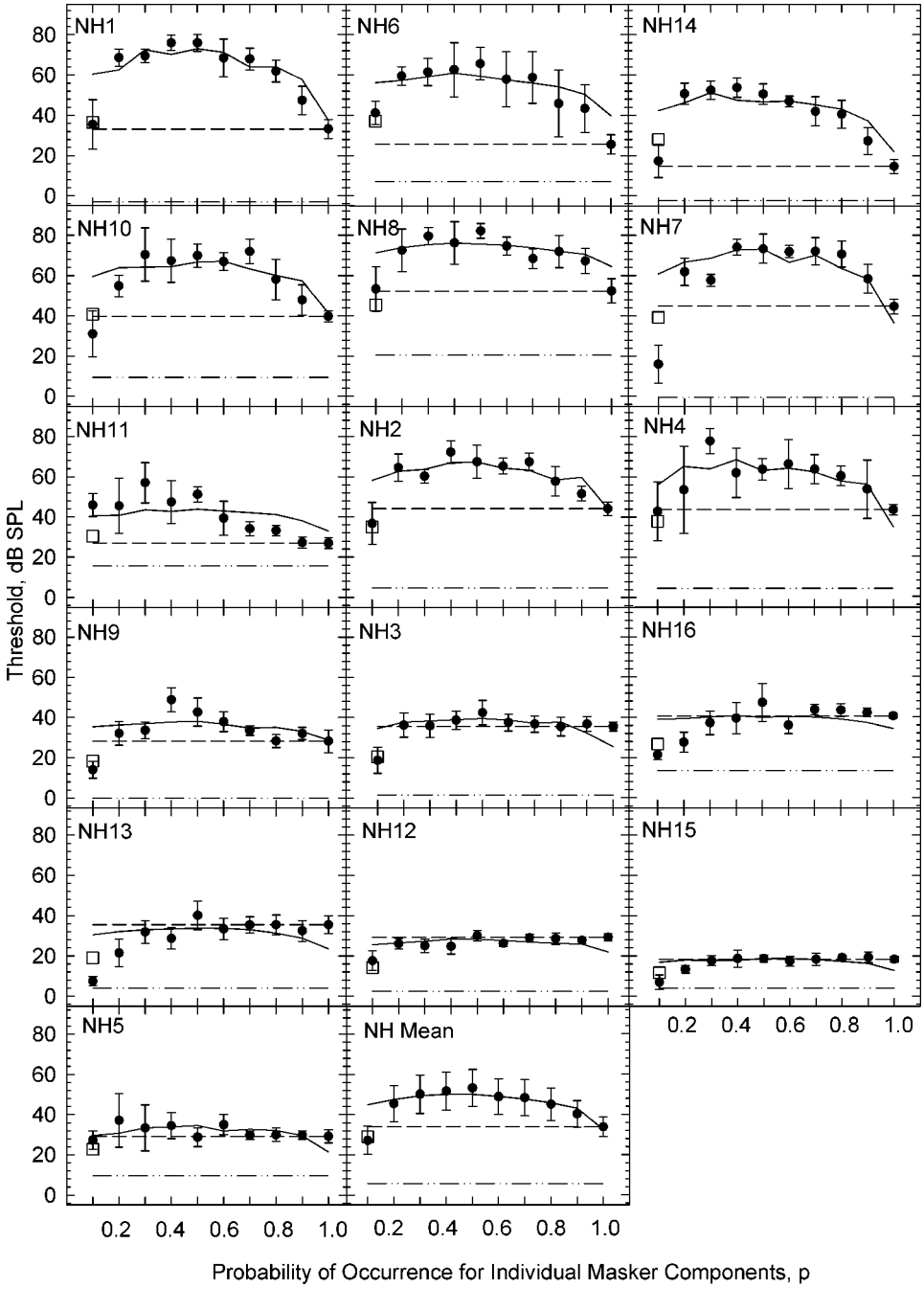


FIG. 1. Masked thresholds for the NH listeners are shown as a function of the probability of occurrence of the individual masker components. Circles represent the mean masked threshold for five trial blocks and the error bars represent the 95% confidence intervals. Listeners are arranged in descending order from left-to-right, top-to-bottom according to the amount of IM at $p=0.5$. The last panel provides the overall group mean and 95% confidence intervals. The amount of masking is represented by the difference between the masked threshold and the quiet threshold for the signal (the dash-dot line). The dashed line represents masked threshold corresponding to energetic masking. The difference between the circles and the energetic masking line is the estimated amount of IM. The solid line is the CoRE model prediction and the open square is the adjusted CoRE model prediction for $p=0.1$ (see footnote 9).

1990; Lutfi, 1992; Doherty and Lutfi, 1996, 1999; Willihnganz *et al.*, 1997; Stellmack *et al.*, 1997). Model predictions are the same for different normalization procedures if the raw regression coefficients for the signal and masker components are all multiplied by the same constant (e.g., normalizing with respect to the signal weight). We chose this particular normalization procedure so that the values can be interpreted as the *proportional* weight that the listeners devote to each component. Weighting efficiency for each listener was determined by computing the root-mean-square (rms) of the difference between the obtained normalized weights and the normalized weights of an ideal analytic listener in which the signal weight is 1 and the masker-component weights are 0 (cf. Dai and Berg, 1992; Willihnganz *et al.*, 1997; Stellmack *et al.*, 1997).[7] We define weighting efficiency as 1−rms so that efficiency ranges from 0 (listener always attends to the wrong auditory filter) to 1 (listener attends only to the signal).

B. Results

1. Information masking

a. Masked thresholds as a function of p. Figures 1 and 2 show the mean masked thresholds (dB SPL) for five blocks of trials as a function of the masker-component probability of occurrence for the NH and HI listeners, respectively. The listeners in each figure are arranged from left-to-right, top-to-bottom in descending order of the amount of IM observed at $p=0.5$, where masker uncertainty is predicted to be greatest. The estimate of energetic masking for all conditions is represented by the amount of masking at $p=1.0$, that is, the masked threshold at $p=1.0$ (the dashed line) minus the quiet

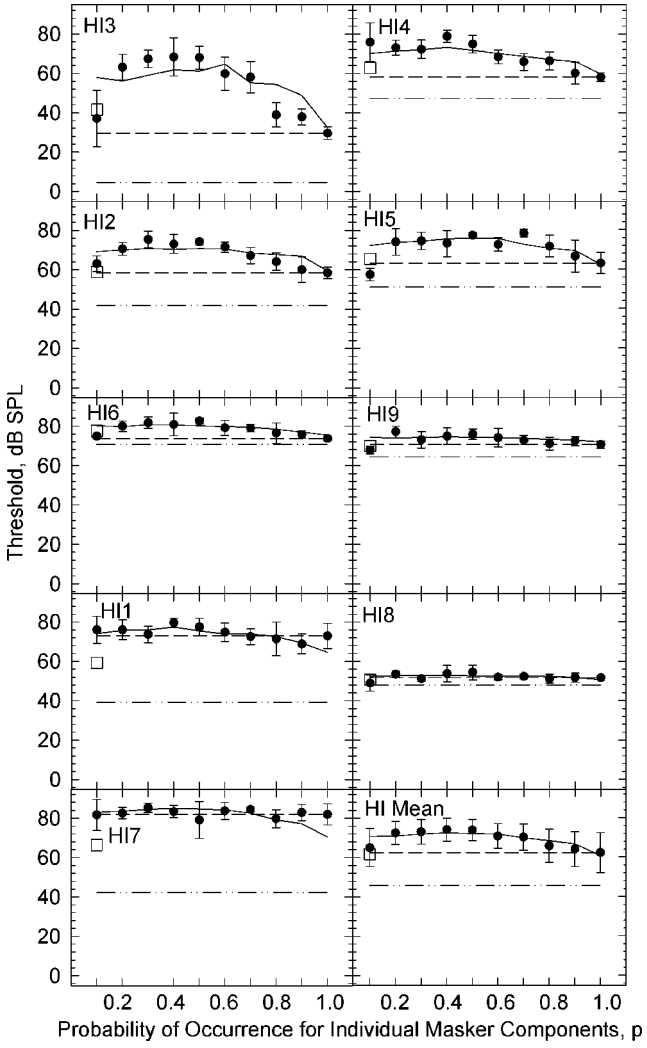


FIG. 2. Masked thresholds for the HI listeners plotted in the same manner as Fig. 1.

threshold for the signal (the dash-dot line). IM for $p<1.0$ is estimated by the difference between the masked thresholds (circles with 95% confidence intervals) and the dashed line for $p=1.0$. The last panel in each figure shows the overall group means with 95% confidence intervals.

For NH listeners masked thresholds for the $p=1.0$ condition ranged from 14.5 to 52.3 dB SPL. Such variability is quite uncharacteristic of what is commonly taken to be energetic masking (cf. Patterson *et al.*, 1982). A possible explanation for the large individual differences is that they reflect a perceptual grouping effect for some listeners in which the pure-tone signal is judged to be absent because it is perceived as belonging to the multitone masker. Such effects have been described in similar conditions by Kidd *et al.* (1994, 2001, 2002). If the masked thresholds for $p=1.0$ do not reflect the exclusive contribution of energetic masking, they still appear to represent a lower limit on masked thresholds for the other masking conditions. Note that the masking functions for $p<1.0$ often collapse on the dashed line representing the masked threshold for $p=1.0$ and, with the exception of the $p=0.1$ condition, individual masked thresholds rarely fall significantly below this line. For this reason we will continue to express the amount of IM relative to the $p=1.0$ condition, recognizing that this may be an underestimate of the amount of IM relative to the amount energetic masking. Some further comments, however, are in order for $p=0.1$ condition. With a few exceptions (NH6, NH11, and HI4), little or no IM was obtained in this condition. This condition is somewhat unusual in that zero or one masker components were present on about 75% of the trials. Seven of the NH and seven of the HI listeners had masked thresholds for $p=0.1$ that were not statistically different from the masked threshold at $p=1.0$ (the minimal uncertainty condition), possibly indicating that masker uncertainty was virtually nonexistent for this condition. (For an alternative explanation, see footnote 9.)

Compared to the NH listeners, the HI listeners in Fig. 2 appear to have lower amounts of IM. At $p=0.5$, NH listeners on average have about 8 dB more IM (M=19.3 dB, SD =14.8) than HI listeners (M=11.5 dB, SD=12.0), however, this difference is not statistically significant, $t(23)=1.34$, $p=0.19$. Notice that the masking function for HI3 is similar to the masking functions of the NH listeners who show moderate to large amounts of IM. This is also the only HI listener whose hearing loss was confined to regions above 2000 Hz, with normal hearing for the 2000-Hz signal and all the masker components below it. When the listeners with normal hearing for the signal frequency, including HI3, are compared to those with a hearing loss for the signal frequency, then the means are 20.4 dB (SD=15.0) and 8.2 dB (SD =7.0), respectively. Listeners who have a hearing loss for the signal frequency have significantly less IM (12.2 dB) than those who have normal hearing for the signal frequency, $t(23)=2.18$, $p=0.04$.

The amount of IM is quite variable among listeners in both groups, ranging from -0.5 to 42.8 dB at $p=0.5$ for the NH group and from -2.9 to 38.3 for the HI group.[8] This range of individual differences is comparable to that of other IM studies (Neff and Green, 1987; Neff and Dethlefs, 1995; Oh and Lutfi, 1998). Although individual differences in IM are more appropriately described by a continuous function (Lutfi *et al.*, 2003b), it is often convenient to describe listeners who show little or no IM as ''analytic'' or ''low-threshold'' listeners and those who are susceptible to IM as ''holistic'' or ''high-threshold'' listeners (Neff *et al.*, 1993). Listeners who show either high or low amounts of IM were distinguished by using an arbitrary criterion of 10 dB of IM at $p=0.5$. The mean masked thresholds for the high- and low-threshold listeners are shown in Fig. 3 for both the NH and HI listeners.

For the NH listeners ($n=10$) and HI listeners ($n=4$) who show high amounts of IM, panels (a) and (b) of Fig. 3, the masked thresholds depend on p. For these listeners, there is a broad maximum of IM between $p=0.3$ and 0.5, the points where variance in the masker ensemble (predicted masker uncertainty) are greatest. In addition, IM is greater for lower probability conditions than for higher probability conditions even though the standard deviation of the binomial distribution on which the predictions are based is symmetric. One likely reason for this is that the level per masker component (SL) increased as the probability of occurrence decreased in order to keep the expected total power

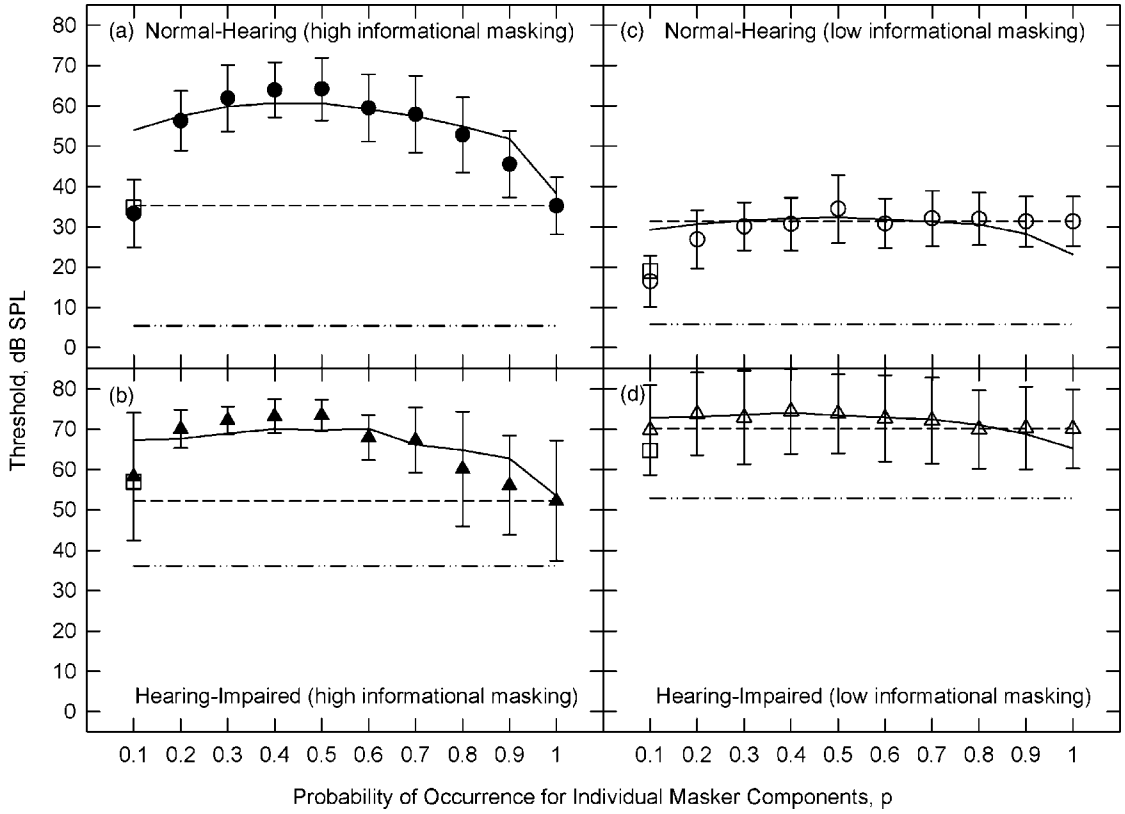


FIG. 3. Mean masked thresholds for the NH (circles) and HI listeners (triangles) plotted separately for listeners who had high (filled symbols) and low (open symbols) amounts of IM. Thresholds are plotted in a similar manner as Figs. 1 and 2 with error bars representing the 95% confidence intervals of the mean masked thresholds.

of the masker ensemble constant [cf. Eq. (5)]. An additional reason might be that as the probability of masker occurrence increased, the outputs of auditory filters became less independent of one another. This would reduce the unshared masker variance at the output of the auditory filters, and so would reduce the amount of IM as p increased.

Six of the 16 NH listeners (NH3, NH16, NH13, NH12, NH15, and NH5), about $\frac{1}{3}$, show little or no IM. This proportion is comparable to other studies (Neff and Green, 1987; Neff and Dethlefs, 1995; Oh and Lutfi, 1998). Five of the nine HI listeners (HI6, HI9, HI1, HI8, and HI7) also show little or no IM. For these listeners the masked thresholds in panels (c) and (d) of Fig. 3 are essentially equal to the masked threshold at $p=1.0$. This result is consistent with the assumption that the masked threshold for $p=1.0$ represents at least a lower bound on masked thresholds in these conditions, if not the exclusive effect of energetic masking. The two groups of NH listeners in panels (a) and (c) of Fig. 3 have similar mean thresholds for the signal in quiet and for $p=1.0$. However, the HI listeners who show little or no IM in panel (d) generally have higher thresholds for the signal in quiet and for $p=1.0$ than their counterparts in panel (b). Even after partialing out the effects of the signal quiet threshold, the amount of IM for the HI listeners is significantly correlated with the masked threshold for $p=1.0$, $r(6)=-0.71$, $p=0.049$. There is no such relationship for the NH listeners, $r(13)=0.14$, $p=0.61$. This pattern of results suggests that different factors might contribute to reduced amounts of IM in NH and HI listeners.

As will be shown in the next section, NH listeners with low amounts of IM have weighting functions that more closely approximate those of an ideal listener, whereas the same is not necessarily true for the HI listeners with low amounts of IM. One explanation is that ''low-threshold'' HI listeners, excluding HI8, could have used overall intensity level or loudness for the decision variable. Note, in particular, that masked thresholds for all probability conditions including $p=1.0$ are greater than the average total power of the masker ensemble. The signal levels corresponding to these thresholds are great enough to shift the overall intensity level of the signal+masker trials several jnd's above the overall intensity level for a majority of the masker-only trials, thus making it a reliable cue for signal detection. However, as noted earlier, where the signal levels adapted to lower thresholds as with the ''high-threshold'' HI listeners, loudness cannot reliably distinguish signal+masker trials from masker-only trials because the number of masker components, hence the overall intensity level, varied randomly from trial to trial. In addition, if overall loudness on each trial was the decision variable instead of the weighted sum of masker and signal SLs, then it might be expected that the variance of the decision variable, hence IM, would be greater than predicted because of the rapid change in loudness associated with loudness recruitment.

b. Practice effects. An effect of practice would be indicated by a decrease in IM across the five trial blocks. A least-squares linear regression analysis revealed that the amount of IM on average decreased by less than 1 dB for the

NH listeners ($b=-0.97$, SE=0.42) and by even less for HI listeners ($b=-0.26$, SE=0.36). Given the large variation in the amount of IM between listeners and the size of the observed effects, this rate of improvement does not alter the interpretation of the data.

c. Predicted amounts of informational masking. Model predictions are shown by the continuous lines in Figs. 1–3. The predicted amount of IM as function of p was first generated using the trial-by-trial standard deviation of the decision variable [Eq. (4)]. Then, the predicted IM function was fit to the data by selecting the intercept that yielded the least-squares error between $p=0.2$ and 0.9. The point at $p=0.1$ was not included in this prediction since, as earlier noted, zero or one masker components were present on about 75% of the trials for this condition.[9] The intercept of the fitted masking function can be taken as the model prediction for energetic masking and should approximately equal the estimated threshold at $p=1.0$. As indicated by the predictions for the individual listeners in Figs. 1 and 2, the model does an adequate job of predicting masked thresholds. The mean difference between the masking at $p=1.0$ and the predicted masking (model intercept) is 1.2 dB for the NH listeners (see last panel in Fig. 1) and 1.4 dB for the HI listeners (see last panel in Fig. 2).

2. Decision weights

a. The independence of weights and SLs. Doherty and Lutfi (1996) found that HI listeners tend to place relatively more weight on stimulus frequencies where hearing loss is greatest (where SLs are generally lowest). To examine whether there was a relationship between the relative weight and the relative SL of the masker components, masker weights and SLs were normalized for each listener so that the sum of their magnitudes equaled unity. No relationship was found between the two variables for the NH listeners, $R^2=0.006$, $F(1,158)=1.0$, $p>0.05$ and for the HI listeners, $R^2=0.01$, $F(1,62)=6.44$ $p=0.014$. These results indicate that SLs and decision weights can be treated as independent factors for both groups. The likely reason that these results fail to replicate the findings of Doherty and Lutfi (1996) is that, unlike the present study, their study involved intensity discrimination in which all the components were potential signals. A fairer comparison is to the companion study of Doherty and Lutfi (1999) in which the signal was a single tone. The results of that study are consistent with those reported here.

b. Weighting functions. Figures 4 and 5 show the individual weighting functions for the NH and HI listeners arranged in the same order as Figs. 1 and 2, respectively (from most to least amount of IM at $p=0.5$). Mean weighting functions for the two groups of listeners are also shown in the last panel of each figure. As the amount of IM decreases for the NH listeners, more weight is placed on the 2000-Hz signal and less weight on the masker frequencies. As noted earlier, listeners who do well on IM experiments have often been termed ''analytic listeners,'' meaning that the SL of the signal is a very strong predictor of a listener's ''yes/no'' responses. It is clear from the weighting functions in Fig. 4 that this term aptly applies to the six NH listeners in Fig. 1 who have little or no IM (NH3, NH16, NH13, NH12, NH15, and NH5). The weighting functions of these listeners are closer to the ideal weighting function with a mean weighting efficiency of 0.88 (SD=0.03) compared to the remainder of the NH listeners who have a mean weighting efficiency of 0.80 (SD=0.03), $t(14)=4.48$, $p\leqslant 0.001$.

The weighting functions for the HI listeners in Fig. 5 should be interpreted with caution because some masker components were presented at levels below quiet threshold (SL$\leqslant$0 dB) and were assigned a weight of zero. These cases are indicated by the open circles in Fig. 5. The scatterplot in Fig. 6 shows the amount of IM at $p=0.5$ as a function of the weighting efficiency for each listener. As can be seen by open circles in the scatterplot, there is no relationship between weighting efficiency and the amount of IM for the HI listeners, $R^2=0.05$, $F(1,7)<1.0$, $p>0.05$. It follows that the HI listeners with little or no IM in Fig. 2 (HI6, HI9, HI1, HI8, and HI7) have a mean weighting efficiency of 0.81 (SD=0.08) that is not significantly higher than the mean weighting efficiency of 0.77 (SD=0.02) from the rest of the HI listeners, $t(7)=1.10$, $p>0.05$.

The scatterplot in Fig. 6 for the NH listeners (filled circles) shows that there is a negative linear relationship between listeners' weighting efficiencies and the amount of IM, $R^2=0.65$, $F(1,14)=25.5$, $p\leqslant 0.001$. This relationship suggests that factors associated with analytic listening or attention are largely responsible for differences in the amount of IM for NH listeners. This conclusion is also consistent with the findings of Lutfi *et al.* (2003b) who preformed a principal components analysis on IM functions from 38 NH children and 46 NH adults and found that one factor was able to account for about 83% of the variance in masked thresholds within and across the age groups. This factor was highly correlated with the estimated number of monitored auditory filters, just as lower weighting efficiencies in this study correspond to the degree to which a listener attends to or monitors the output of nonsignal auditory filters instead of the auditory filter centered on the signal frequency.

C. Experiment I discussion

Results for the equal-SPL maskers in experiment I are summarized in Table II. The listeners are listed by group in descending order according to the amount of IM at $p=0.5$ (fourth column), the same ordering used for Figs. 1 and 2. The table provides the quiet threshold for the signal (dB SPL), the mean SL of the ten masker components that were presented in experiment I when $p=0.5$ along with their range, and the listener's weighting efficiency. It is apparent from Table II that the mean SL of the masker components was the primary difference between the NH listeners (M=48.9 dB SL, SD=3.0) and the HI listeners (M=16.0 dB SL, SD=9.4), $t(23)=13.1$, $p<0.001$. In contrast, there was a fair amount of overlap in the weighting efficiencies between the NH listeners (M=0.83, SD=0.05) and the HI listeners (M=0.80, SD=0.06), $t(23)=1.63$, $p=0.12$. This indicates that the difference in IM between NH and HI listeners is not likely explained by differences in weighting strategies.

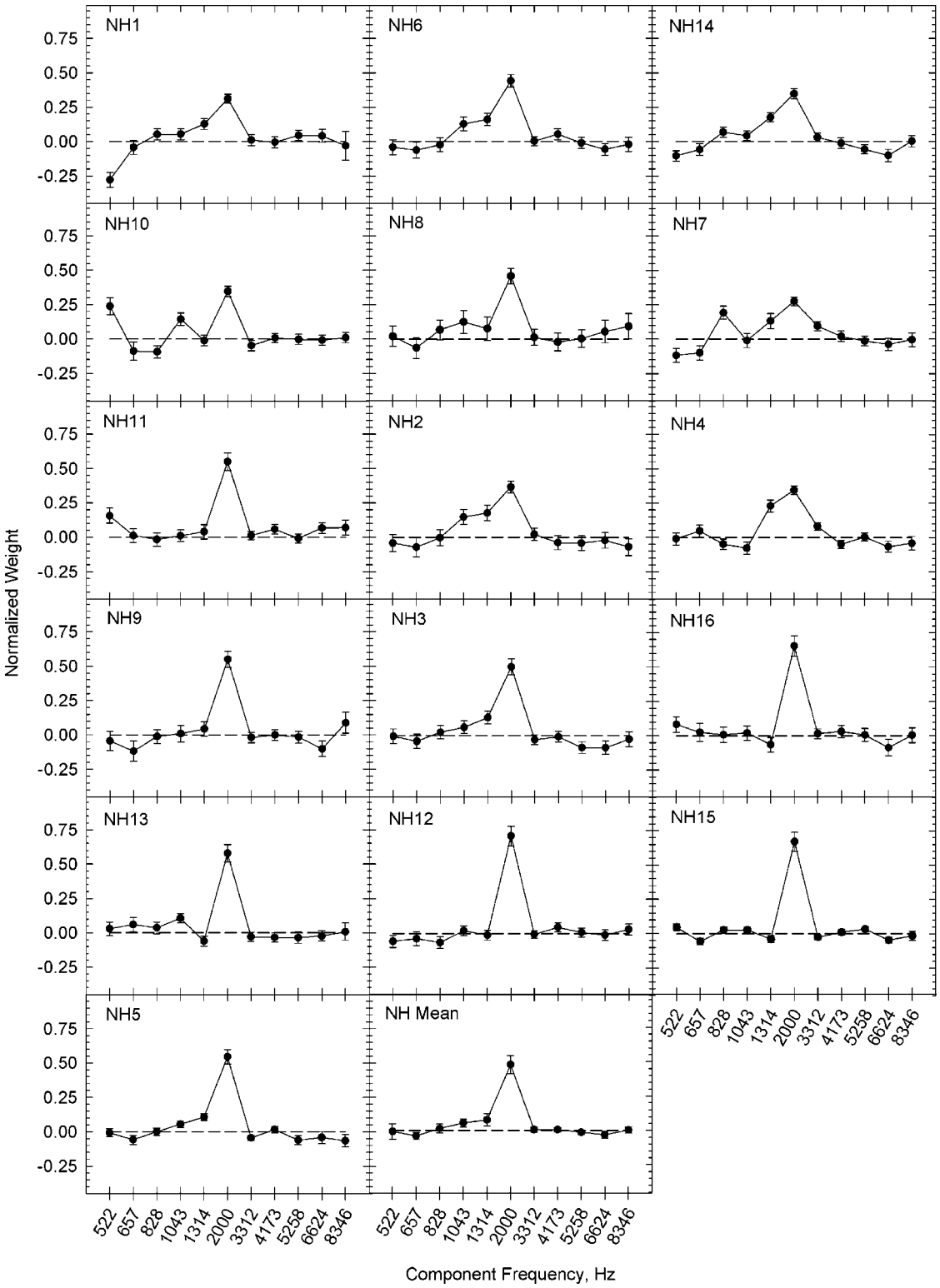


FIG. 4. Decision weights (filled circles) as a function of the 2000-Hz signal and of the masker components for the NH listeners plotted in the same order as Fig. 1, with error bars representing the 95% confidence intervals.

According to Eq. (4), the reduced masker-component SLs associated with hearing loss effectively reduce the trial-by-trial variance of the masker at the output of the auditory filters. That is, even though masker components were either on or off, the difference in the output level in the auditory filters between the on and off states was smaller for the HI listeners than for NH listeners. IM is the result of an interaction and requires that large weights be placed on masker components with relatively high SLs. Thus, two listeners can have similar weighting functions, but if the SLs of the masker components are greater for one listener than for another, then there can be a large difference in the amount of IM between the two. For example, the weighting functions for listeners NH14 and HI5 in Figs. 4 and 5 are somewhat

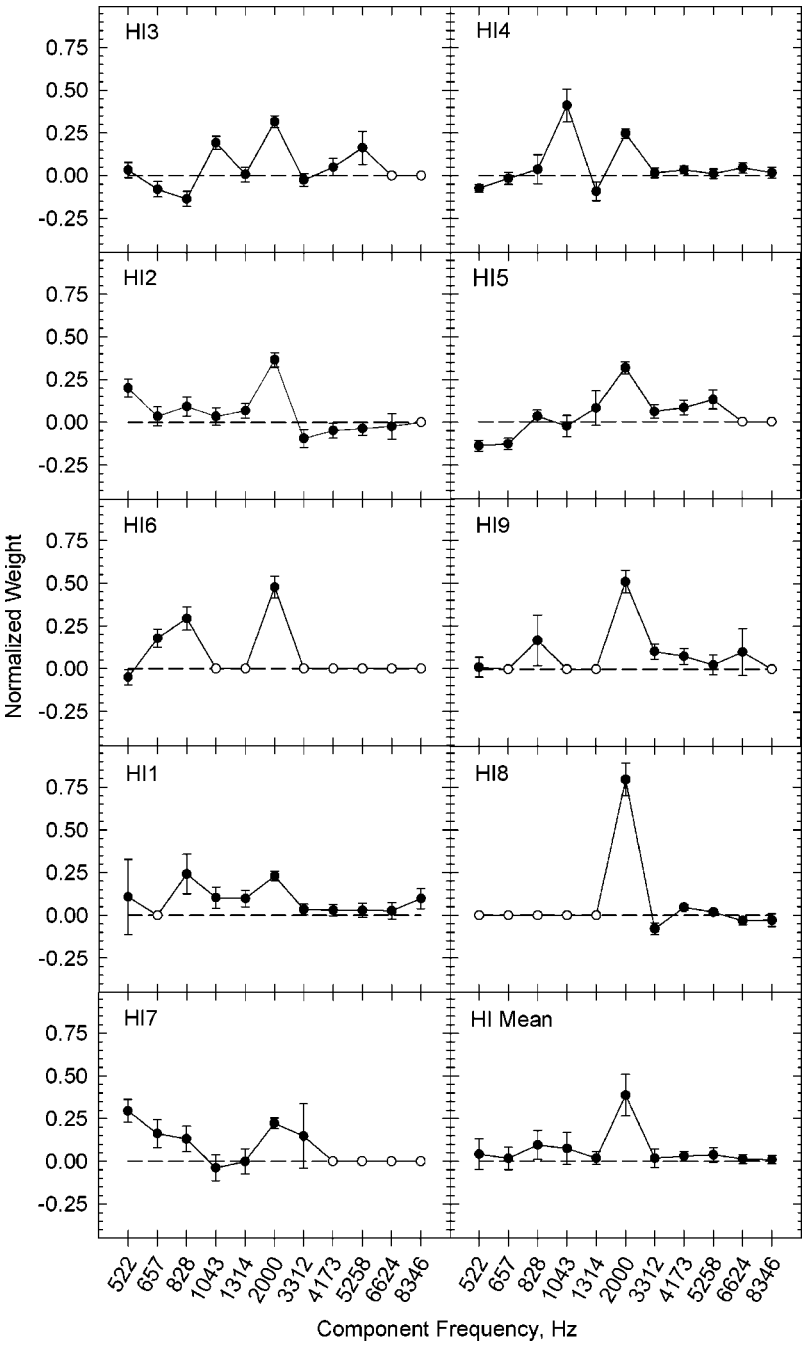


FIG. 5. Decision weights (filled circles) as a function of the 2000-Hz signal and of the masker components for the HI listeners plotted in the same order as Fig. 2, with error bars representing the 95% confidence intervals. Open circles indicate weights that were set to 0 because the masker components were presented at levels lower than their thresholds in quiet.

similar, but due to the differences in the mean SL of the masker components, 49.1 vs. 15.6 dB SL, NH14 has much more IM at $p=0.5$, 36.0 dB, than HI5 who has only 14.1 dB. Figure 6 shows that when two listeners, one from each group, have equally poor weighting efficiencies (about 0.80 or less) that the NH listener will usually have a greater amount of IM than the HI listener. The one exception is HI3 (the lone open circle in the upper-left hand corner of Fig. 6) who has as much IM as the NH listeners. This is likely because there were a sufficient number of masker components with high SLs for this listener who had normal hearing for the 2000-Hz signal and below.

Unlike for the NH listeners, weighting efficiency and IM are not strongly correlated among the HI listeners. One reason for this might be that there is more variability in the mean masker-component SLs for the HI listeners (SD=9.4 dB) than there is for the NH listeners (SD=3.0 dB). Because IM depends on both SLs and weights, when the SLs of the masker components are relatively homogenous among listeners as they are for the NH listeners, individual differences in the amount of IM will be more directly influenced by weighting efficiency. However, when the SLs of the masker components are relatively heterogeneous, as they are for the HI listeners, the amount of IM will be less directly influenced by weighting efficiency.

The results of this study suggest that when masker SL differs substantially between or within groups of listeners that it will be the dominant factor determining IM. One way to confirm this hypothesis would be to show substantial IM in HI listeners when masker SLs are the same as for the NH listeners. This is not practical in the present study since masked threshold for $p=1.0$ in many HI listeners was close

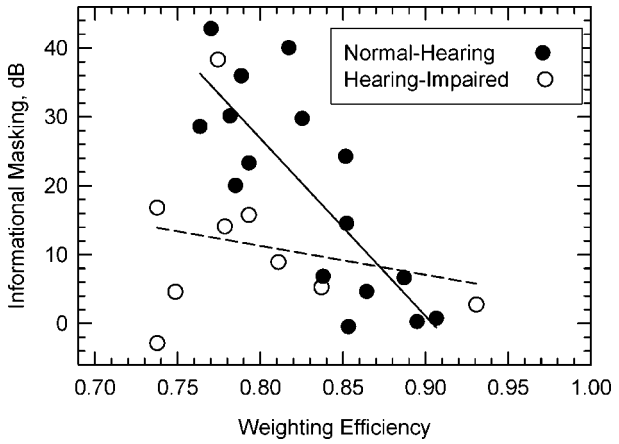


FIG. 6. IM at $p=0.5$ as a function of the weighting efficiency for the NH listeners (filled circles) and the HI listeners (open circles). The solid and dashed lines represent the least-squares linear regression fit for the NH and HI listeners, respectively.

to the maximum allowable signal level. Instead, two types of control experiments were conducted in which masker SLs were equated for NH and HI listeners.

III. EXPERIMENT II: SIMULATION OF REDUCED SL'S IN NORMAL-HEARING LISTENERS

A. Method

Four NH listeners who showed large amounts of IM in experiment I (NH1, NH4, NH7, and NH14) repeated the experiment with the masker levels adjusted to roughly equal the average masker SLs of the HI listeners. Masker levels were determined by subtracting the average hearing loss of the HI listeners from the otherwise equal-SPL masker levels.[10] The average total power of the maskers across all conditions was reduced by about 20 dB to a little more than 50 dB SPL. In order from lowest to highest frequency, the mean masker-component SLs for the four listeners were 26.2, 22.5, 20.6, 16.4, 17.8, 25.8, 21.4, 22.9, 7.0, and 3.7 dB. The mean masker SL per component was reduced from 47.7 to 18.4 dB, which was close to the mean for HI listeners, 16.0 dB.

B. Results

Weighting efficiency for NH1, NH4, NH7, and NH14 was 0.76, 0.81, 0.79, and 0.85, respectively and IM was 16.3, 13.3, 11.7, and 12.0 dB, respectively. Mean weighting efficiency was not much different in this experiment, 0.80, compared to experiment I, 0.78, $t(3)=1.73$, $p>0.05$. Figure 7 shows that when the SL of the masker components are adjusted to simulate the average hearing loss of the HI listeners in experiment I that IM as a function of p is reduced. At $p=0.5$ the mean amount of IM decreased from 31.9 dB in experiment I to 13.3 dB in experiment II, $t(3)=4.20$, $p=0.025$. The amount of IM for the NH listeners in this experiment was similar to the 11.5 dB of IM for the HI listeners in experiment I and indirectly supports the conclusion that reduced masker-component SLs were mostly responsible for the lower amounts of IM in the HI listeners.

IV. EXPERIMENT III: EQUAL-SL MASKER COMPONENTS

A. Method

This experiment addressed how equating the SLs of the masker components affects the difference in the amount of IM and weighting efficiency between 12 NH and 6 HI listeners from experiment I (the listeners are those in Table II with entries under the heading ''Equal-SL''). The signal and maskers were the same as in experiments I and II. Only the minimum ($p=1.0$) and maximum ($p=0.5$) uncertainty conditions were tested. For the $p=0.5$ condition, each masker component was played at 10 dB SL based on the estimates of quiet threshold obtained earlier. As before, the level per masker component in the $p=1.0$ condition was 3 dB less so that the average total power of the masker ensemble would be the same for both conditions. A constant 10 dB SL was used because higher masker levels would have prevented some HI listeners from detecting the signal above chance even at the most intense levels tested (about 90 dB SPL). For HI6 and HI7 the SL of the 8346-Hz masker component was set to 0 dB SL because setting it at 10 dB SL would have exceeded 90 dB SPL.

B. Results and discussion

For the equal-SL maskers in this experiment, Table II provides the masked thresholds for $p=0.5$ and $p=1.0$, the amount of IM (i.e., the difference between the two masked thresholds), and the weighting efficiency. IM for the NH listeners was significantly less in this experiment (M=5.0 dB, SD=3.7) compared to experiment I (M=18.8, SD=16.2), $t(11)=26.2$, $p<0.001$. This decrease was likely due to the difference in the masker SLs rather than weighting efficiency because weighting efficiency was significantly less in this experiment (M=0.78, SD=0.05) compared to experiment I (M=0.83, SD=0.05), $t(11)=3.16$, $p\leqslant 0.01$, which would otherwise be associated with greater IM. As with experiment I, weighting efficiency and IM in NH listeners were significantly correlated, $R^2=0.35$, $F(1,10)=5.38$, $p=0.04$. The relationship is smaller in this experiment compared to experiment I where masker-component SLs varied within and between listeners probably because there is less variance in the amount of IM.

IM for the HI listeners in this experiment (M=9.4 dB, SD=7.8) was not significantly different from experiment I (M=7.5 dB, SD=7.3) where the mean masker SL was 11.4 dB, $t(5)=1.10$, $p>0.05$. The results of this experiment also indicate that when masker-component SLs were equal for NH and HI listeners that IM and weighting efficiency were not significantly different between the two groups, $t(16)=1.68$, $p=0.11$, and $t(16)=1.82$, $p=0.09$, respectively. As with the NH listeners, weighting efficiency for the HI listeners was significantly less in this experiment, 0.74 (SD=0.03), compared to experiment I, 0.81 (SD=0.07), $t(11)=3.54$, $p=0.02$. This seems to indicate that equating SL across masker components makes masker components more difficult to ignore for both groups of listeners.

In experiment I the lack of a significant relationship between weighting efficiency and IM was attributed to unequal

TABLE II. Summary data for experiment I (equal-SPL maskers) and experiment III (equal-SL maskers) for normal-hearing (NH) and hearing-impaired listeners (HI) is shown. Listeners are listed in descending order according to the amount of IM at $p=0.5$ for the equal-SPL maskers (fourth column), the same ordering used for Figs. 1 and 2. Shown for each listener are the quiet threshold for the signal (dB SPL), the mean and range of the SLs for the ten masker components that were presented when $p=0.5$ for the equal-SPL maskers, and the weighting efficiency (see text). For the equal-SL maskers, the masked threshold when $p=0.5$ and when $p=1.0$ are given along with the amount of IM (i.e., the difference between the two) and the weighting efficiency.

Masker type		Equal-SPL			Equal-SL			
Listener	Signal quiet threshold (dB SPL)	Mean masker SL at $p=0.5$ (range)	IM at $p=0.5$	Weight efficiency	Masked threshold at $p=0.5$	Masked threshold at $p=1.0$	IM	Weight efficiency
NH1	−3.1	50.3(21.3–59.6)	42.8	0.77	17.52	12.33	5.2	0.72
NH6	7.1	52.0(40.6–69.3)	40.1	0.82	15.18	8.02	7.2	0.76
NH14	−2.5	49.1(39.7–58.2)	36.0	0.79	4.54	−2.25	6.8	0.80
NH10	9.4	50.7(31.8–63.1)	30.2	0.78				
NH8	20.3	44.2(35.3–56.3)	29.8	0.83	30.41	21.91	8.5	0.76
NH7	−0.6	49.9(41.1–65.6)	28.6	0.76	16.95	13.75	3.2	0.75
NH11	15.7	47.0(35.4–65.3)	24.3	0.85				
NH2	4.7	50.5(40.6–61.2)	23.3	0.79				
NH4	3.9	49.5(36.6–69.6)	20.1	0.79	16.12	8.91	7.2	0.72
NH9	−0.2	43.6(29.3–61.1)	14.6	0.85	13.59	1.96	11.6	0.76
NH3	1.5	50.6(42.7–64.6)	6.9	0.84	8.49	6.87	1.6	0.81
NH16	13.3	45.3(35.1–65.4)	6.6	0.89	22.32	16.87	5.5	0.77
NH13	4.0	49.1(30.5–63.7)	4.7	0.86				
NH12	2.3	55.1(42.1–72.7)	0.8	0.91	18.42	14.45	4.0	0.75
NH15	4.0	49.1(32.0–61.2)	0.3	0.90	4.58	5.98	−1.4	0.82
NH5	9.7	47.0(29.2–68.3)	−0.5	0.85	9.30	9.11	0.2	0.92
HI3	4.7	34.2(0.0–53.4)	38.3	0.77				
HI4	47.2	20.0(7.2–31.1)	16.8	0.74	68.68	48.18	20.5	0.73
HI2	41.7	23.0(2.0–31.1)	15.8	0.79				
HI5	50.8	15.7(0.0–28.9)	14.1	0.78	68.97	58.71	10.3	0.73
HI6	70.8	3.5(0.0–13.1)	8.9	0.81	89.88	82.19	7.7	0.73
HI9	64.5	5.7(0.0–14.2)	5.2	0.84	82.49	68.69	13.8	0.74
HI1	39.3	18.4(2.1–33.5)	4.6	0.75				
HI8	47.8	13.2(0.0–40.8)	2.7	0.93	56.99	49.67	7.3	0.78
HI7	42.1	10.3(0.0–21.2)	−2.9	0.74	88.58	91.58	−3.0	0.70

masker-component SLs within and across listeners. However, even though masker-component SLs were equated in this experiment, weighting efficiency alone did not reliably predict IM in HI listeners, $R^2=0.09$, $F(1,4)<1.0$. One possible reason for this weakened relationship is that the auditory-filter bandwidths for some HI listeners might be broad enough to increase the correlation between the auditory filter outputs, which would effectively reduce the masker variance. As we have noted, some correlation is likely, particularly at high values of p. Another reason, however, is that there is less variance in the amount of IM for HI listeners.

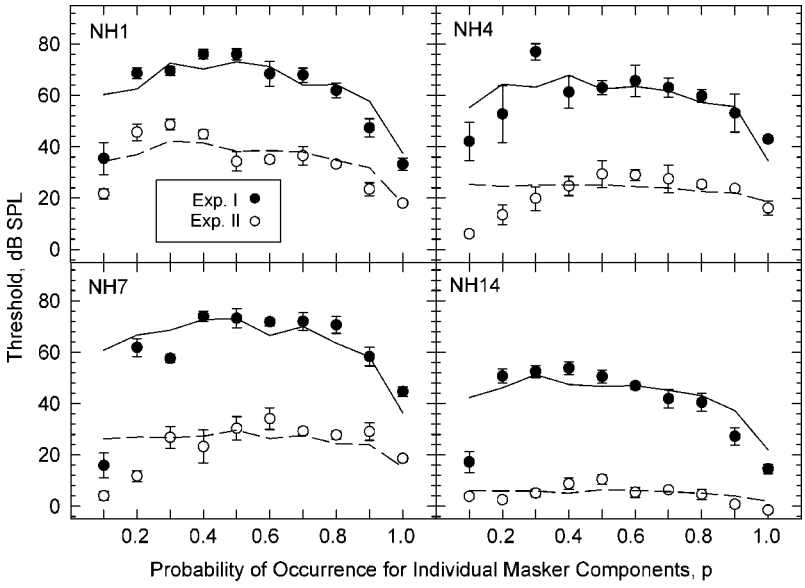


FIG. 7. Open circles indicate masked thresholds with 95% confidence intervals for the four NH listeners in experiment II where the level per masker component was adjusted to simulate the reduced SLs associated with hearing loss. For comparison, the experiment I results for these listeners are replotted from Fig. 1 (filled circles).

V. SUMMARY AND CONCLUSIONS

Interpreting the results of this study within the framework of a specific decision model [Eq. (4)] allows us to evaluate how IM is affected by a hearing loss in terms of lower masker SLs and in terms of perceptual differences in how the stimulus components are weighted. According to the model, differences in the amount of IM between and within NH and HI listeners is dependent on how decision weights are distributed across masker components with different SLs, with the greatest amounts of IM occurring when large weights are combined with high SLs. For the experiments in this study, it seems that the SLs and weights of the masker components are independent factors because a notable relationship between them is lacking for both groups.

The results of the experiments in this study are in general agreement with the findings of Micheyl *et al.* (2000). As in that study, differences in the amount of IM between our NH and HI listeners seem to be strongly affected by differences in the SLs of the masker components. When the signal was in a region of hearing loss, the amount of IM for the equal-SPL maskers never reached a level as high as that obtained by a majority of NH listeners. Because the dB SPL per masker component was equal for the NH and HI listeners, masker SLs were much lower for the HI listeners than for the NH listeners. The observed results are thus interpreted in terms of the resulting reduction in the variance of the assumed decision variable. The significant decrease in IM for the NH listeners when the masker SLs were reduced in experiments II and III confirm that the differences in masker SLs between the NH and HI listeners in experiment I was likely a major contributing factor to the differences in the amounts of IM. Further confirmation is provided by the results of experiment III where no significant differences in IM were found when masker-component SLs were equated within and across the two groups of listeners.

Although HI listeners generally had lower weighting efficiencies than the NH listeners in experiments I and III, our results do not lend strong support to the speculation by Kidd *et al.* (2001) that HI listeners use a less analytic listening (weighting) strategy than NH listeners. However, unlike for NH listeners, the relationship between weighting efficiency and IM was not significant for HI listeners even in experiment III where the effect of masker-component SLs was virtually eliminated. We suggest that in addition to reduced masker SLs, broadened auditory filters accompanying sensorineural hearing loss might also reduce IM by increasing the covariance between auditory filter outputs, thereby decreasing the masker ensemble variance.

ACKNOWLEDGMENTS

This research was supported by a grant to the second author from the NIDCD (R01 DC1262-10). We would like to thank Dr. Christophe Micheyl, Dr. Gerald Kidd, and an anonymous reviewer for their very useful comments on an earlier version of this manuscript.

[1]While the HI listeners were generally older than the NH listeners, observed differences between the two groups are unlikely related to the age difference because four of the nine HI listeners were of the same approximate age (<30 years) as the NH listeners.

[2]The Sennheiser earphone (test earphone) was calibrated with a TDH-50 earphone (standard earphone) using continuous tones and a binaural loudness balancing procedure (ANSI, 1996). Calibration at the octave frequencies between 250 and 8000 Hz was carried out by two normal-hearing experimenters. The output levels of tones played through the standard earphone were set to 60 dB SPL, then the test earphone was balanced using the method of adjustment and a programmable attenuator (TDT, PA4). Median RMS voltages across several independent adjustments of the output levels from the test earphone for each ear of the experimenters were taken as the matching values. The RMS voltage for the 1000-Hz tone was used as the reference for the Sennheiser earphone.

[3]When rounded to the nearest one-hundredth, actual masker probabilities averaged across all listeners were equal to the nominal probabilities except at $p=0.5$ where the actual probability was 0.51. The standard deviations of the actual probabilities across all listeners was 0.003 for $p=0.3$, 0.5, and 0.6 and was 0.002 for the other conditions.

[4]This occurred on 0.39% of the trials used in the computation of masked threshold for the NH listeners and on 1.82% of these trials for the HI listeners, excluding HI7 (34.2%) who had difficulty even at the reduced masker levels.

[5]The analytic expression in Eq. (4) yields almost identical predictions as the empirical estimate.

[6]Data were fit to a generalized linear model and a binomial distribution using the ''glmfit'' function in Matlab® to obtain the weights. Note that Eq. (4) assumes that the weights are normalized with respect to the signal weight, w_s, which is 1 in this instance. When the weights are instead normalized to sum to 1 as they are for this report, the right-hand side of Eq. (4) must be divided by w_s.

[7]Note that this method of quantifying efficiency or η is different from the method outlined by Berg (1990) and Tanner and Birdsall (1958) in which the obtained d' (squared) is divided by the d' (squared) from an ideal observer:

$$\eta_{obt}=\eta_{wgt}\times\eta_{noise}\Leftrightarrow\frac{(d'_{obt})^2}{(d'_{ideal})^2}=\frac{(d'_{wgt})^2}{(d'_{ideal})^2}\times\frac{(d'_{obt})^2}{(d'_{wgt})^2},$$

where η_{obt} represents the degree to which a listener's obtained sensitivity matches that of an ideal observer, where η_{wgt} represents the degree to which the sensitivity from a hypothetical listener who utilizes the measured weights matches the sensitivity of an ideal observer, and where η_{noise} represents the degree to which a listener's obtained sensitivity matches that of a hypothetical listener who utilizes the measured weights. This computation is not meaningful in the present study because an ideal observer is equivalent to a filter matched to the signal frequency and would therefore commit no errors; that is, the denominator d' ideal is infinite.

[8]''Negative'' amounts of IM are possible because IM is operationally defined as the difference between two estimates of masked threshold, one for $p<1.0$ and another for $p=1.0$. Because of variability or errors in the estimates, predictions for negative IM are inevitable. These values were left uncorrected (i.e., were not set equal to zero) so as not to artificially deflate the variability in the estimates, especially for the purposes of statistical analysis.

[9]Since the total amount of masking is predicted to be the dB sum of energetic and informational masking, the failure of the model to predict the masked threshold at $p=0.1$ must be attributed to one or both of these factors. If the predicted amounts of IM were incorrect, this would imply that the listeners used a different, more analytic, weighting strategy for this condition. However, it is also reasonable to consider that the prediction for energetic masking must be inaccurate to some extent because for some listeners in Figs. 1 and 2 the estimated thresholds for $p=0.1$ are lower than the $p=1.0$ estimate of energetic masking. One parsimonious solution for this discrepancy is to assume that energetic masking at $p=0.1$ is some constant proportion, k, of the CoRE model prediction for energetic masking at $p=1.0$ (in dB). As shown by the open squares plotted at $p=0.1$ in Figs. 1–3, arbitrarily choosing $k=0.4$ yields accurate predictions for most listeners. This adjustment dramatically improves the model predictions by decreasing the mean discrepancy between the estimated and the predicted thresholds at $p=0.1$ to only -1.9 dB for the NH listeners and to 3.4 dB for the HI listeners.

[10]The amount of hearing loss was determined by subtracting the mean quiet thresholds of the HI listeners from laboratory norms for the frequencies used in the experiments. Norms were obtained from the 10 NH listeners

who had hearing levels no greater than 10 dB HL for all the octave frequencies during a routine audiologic evaluation. So that the computed average hearing loss would more closely reflect the actual masker-component SLs, quiet thresholds that were greater than the presentation level for a masker component were substituted with the level that corresponded to 0 dB SL (i.e., the presentation level). Note that the effect of restricting the amounts of hearing loss that entered into the total was to make the average hearing loss appear more similar across frequency, especially in the high-frequency regions.

Allen, P., and Wightman, F. L. (**1995**). ''Effects of signal and masker uncertainty on children's detection,'' J. Speech Hear. Res. **38**, 503–511.

ANSI (**1996**). ANSI S3.6-1996, ''American National Standards specification for audiometers'' (American National Standards Institute, New York).

Berg, B. G. (**1989**). ''Analysis of weights in multiple observation tasks,'' J. Acoust. Soc. Am. **86**, 1743–1746.

Berg, B. G. (**1990**). ''Observer efficiency and weights in a multiple observation task,'' J. Acoust. Soc. Am. **88**, 149–158.

Dai, H. and Berg, B. G. (**1992**). ''Spectral and temporal weights in spectral-shape discrimination,'' J. Acoust. Soc. Am. **92**, 1346–1355.

Doherty, K. A., and Lutfi, R. A. (**1996**). ''Spectral weights for overall level discrimination in listeners with sensorineural hearing loss,'' J. Acoust. Soc. Am. **99**, 1053–1057.

Doherty, K. A., and Lutfi, R. A. (**1999**). ''Level discrimination of single tones in a multitone complex by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **105**, 1831–1840.

Kidd, Jr., G., Mason, C. R., and Arbogast, T. L. (**2002**). ''Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **111**, 1367–1376.

Kidd, Jr., G., Arbogast, T. L., Mason, C. R., and Walsh, M. (**2001**). ''Informational masking in listeners with sensorineural hearing loss,'' J. Assoc. Res. Otolaryngol. **3**, 107–119.

Kidd, Jr., G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). ''Reducing informational masking by sound segregation,'' J. Acoust. Soc. Am. **95**, 3475–3480.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Lutfi, R. A. (**1990**). ''How much masking is informational masking?'' J. Acoust. Soc. Am. **88**, 2607–2610.

Lutfi, R. A. (**1992**). ''Informational processing of complex sound. III. Interference,'' J. Acoust. Soc. Am. **91**, 3391–3401.

Lutfi, R. A. (**1993**). ''A model of auditory pattern analysis based on component-relative-entropy,'' J. Acoust. Soc. Am. **94**, 748–758.

Lutfi, R. A. (**1995**). ''Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks,'' J. Acoust. Soc. Am. **97**, 1333–1334.

Lutfi, R. A., Kistler, D. J., Callahan, M. R., and Wightman, F. L. (**2003a**). ''Psychometric functions for informational masking,'' J. Acoust. Soc. Am. **114**, 3273–3282.

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (**2003b**). ''One factor underlies individual differences in auditory informational masking within and across age groups,'' Percept. Psychophys. **65**, 396–406.

Micheyl, C., Arthaud, P., Reinhart, C., and Collet, L. (**2000**). ''Informational masking in normal-hearing and hearing-impaired listeners,'' Acta Oto-Laryngol. **120**, 242–246.

Neff, D. L., and Callaghan, B. P. (**1988**). ''Effective properties of multicomponent simultaneous maskers under conditions of uncertainty,'' J. Acoust. Soc. Am. **83**, 1833–1838.

Neff, D. L., and Dethlefs, T. M. (**1995**). ''Individual differences in simultaneous masking with random-frequency, multicomponent maskers,'' J. Acoust. Soc. Am. **98**, 125–134.

Neff, D. L., and Green, D. M. (**1987**). ''Masking produced by spectral uncertainty with multicomponent maskers,'' Percept. Psychophys. **41**, 409–415.

Neff, D. L., Dethlefs, T. M., and Jesteadt, W. (**1993**). ''Informational masking for multicomponent maskers with spectral gaps,'' J. Acoust. Soc. Am. **94**, 3112–3126.

Oh, E. L., and Lutfi, R. A. (**1998**). ''Nonmonotonicity of informational masking,'' J. Acoust. Soc. Am. **104**, 3489–3499.

Oh, E. L., Wightman, F. L., and Lutfi, R. A. (**2001**). ''Children's detection of pure-tone signals with random multitone maskers,'' J. Acoust. Soc. Am. **109**, 2888–2895.

Patterson, R. D., and Green, D. M. (**1978**). ''Auditory masking,'' in *Handbook of Perception, Volume IV*, edited by E. Carterette and M. Friedman (Academic, New York), pp. 337–361.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.

Richards, V. M., and Zhu, S. (**1994**). ''Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients,'' J. Acoust. Soc. Am. **95**, 423–434.

Richards, V. M., Tang, Z., and Kidd, Jr., G. (**2002**). ''Informational masking with small set sizes,'' J. Acoust. Soc. Am. **111**, 1359–1366.

Stellmack, M. A., Willihnganz, M. S., Wightman, F. L., and Lutfi, R. A. (**1997**). ''Spectral weights in level discrimination by preschool children: Analytic listening conditions,'' J. Acoust. Soc. Am. **101**, 2811–2821.

Tang, Z., and Richards, V. M. (**2002**). ''Internal noise invariance across two informational masking tasks,'' J. Acoust. Soc. Am. **111**, 2470.

Tanner, W. P., and Birdsall, T. G. (**1958**). ''Definitions of d' and η as psychological measures,'' J. Acoust. Soc. Am. **30**, 922–928.

Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). ''Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty,'' J. Acoust. Soc. Am. **60**, 1176–1186.

Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). ''Children's detection of pure-tone signals: Informational masking with contralateral maskers,'' J. Acoust. Soc. Am. **113**, 3297–3305.

Willihnganz, M. S., Stellmack, M. A., Lutfi, R. A., and Wightman, F. L. (**1997**). ''Spectral weights in level discrimination by preschool children: Synthetic listening conditions,'' J. Acoust. Soc. Am. **101**, 2803–2810.

Wright, B. A., and Saberi, K. (**1999**). ''Strategies used to detect auditory signals in small sets of random maskers,'' J. Acoust. Soc. Am. **105**, 1765–1775.

Masker phase effects in normal-hearing and hearing-impaired listeners: Evidence for peripheral compression at low signal frequencies

Andrew J. Oxenham[a)]
Research Laboratory of Electronics, and Harvard-MIT Division of Health Sciences and Technology, Speech and Hearing Bioscience and Technology Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Torsten Dau
Centre for Applied Hearing Research, Acoustic Technology, Ørsted•DTU, Technical University of Denmark, DK-2800 Lyngby, Denmark

(Received 19 December 2003; revised 29 June 2004; accepted 9 July 2004)

The presence of cochlear-based compression at low frequencies was investigated by measuring phase effects in harmonic maskers. In normal-hearing listeners, the amount of masking produced depends strongly on the phase relationships between the individual masker components. This effect is thought to be determined primarily by properties of the cochlea, including the phase dispersion and compressive input–output function of the basilar membrane. Thresholds for signals of 250 and 1000 Hz were measured in harmonic maskers with fundamental frequencies of 12.5 and 100 Hz as a function of the masker phase curvature. Results from 12 listeners with sensorineural hearing loss showed reduced masker phase effects, when compared with data from normal-hearing listeners, at both 250- and 1000-Hz signal frequencies. The effects of hearing impairment on phase-related masking differences were not well simulated in normal-hearing listeners by an additive white noise, suggesting that the effects of hearing impairment are not simply due to reduced sensation level. Maximum differences in masked threshold were correlated with auditory filter bandwidths at the respective frequencies, suggesting that both measures are affected by a common underlying mechanism, presumably related to cochlear outer hair cell function. The results also suggest that normal peripheral compression remains strong even at 250 Hz. © *2004 Acoustical Society of America*. [DOI: 10.1121/1.1786852]

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Sr [GK] Pages: 2248–2257

I. INTRODUCTION

Changing the phase relationships of components within a harmonic tone complex masker can have dramatic effects on the ability to detect a signal embedded in the masker. In general, waveforms that produce highly modulated patterns of temporal activity in the auditory periphery lead to lower thresholds than waveforms that produce a relatively unmodulated temporal envelope. Differences in masked threshold between such temporally ''peaked'' or ''flat'' maskers can in some cases exceed 25 dB. These masking effects can be explained in two (not mutually exclusive) ways. First, listeners may be able to make use of the momentary high signal-to-masker ratios in low-level epochs of the modulated masker envelopes to improve signal detection by ''listening in the valleys'' (Buus, 1985; Kohlrausch and Sander, 1995). Second, fast-acting compression of the waveform in the auditory periphery reduces the effective rms level of a modulated masker, relative to that of an unmodulated masker (Carlyon and Datta, 1997a; Oxenham and Dau, 2001a). The second explanation, based on peripheral compression, is consistent with the finding that hearing-impaired listeners generally show little or no effect of masker phase on signal threshold (Summers and Leek, 1998; Summers, 2000). A common cause of hearing impairment is a loss of outer hair cell function, which in turn leads to a reduction or loss of the basilar membrane's compressive input-output function (e.g., Ruggero and Rich, 1991; Oxenham and Plack, 1997). Thus, masker phase effects may provide an indirect measure of basilar-membrane compression.

One apparent anomaly in the compression theory of masker phase effects is that large effects continue to be found for normal-hearing listeners down to signal frequencies as low as 125 Hz (Oxenham and Dau, 2001b). In contrast, earlier physiological and psychophysical studies had concluded that cochlear nonlinearity is reduced or even absent at low characteristic frequencies (CFs) (Cooper and Yates, 1994; Hicks and Bacon, 1999; Plack and Oxenham, 2000). At least three interpretations of this apparent discrepancy are possible.

The first interpretation relates to the masker fundamental frequencies ($F0$s) used at low signal frequencies. At a signal frequency of 125 Hz, Oxenham and Dau (2001b) used a masker $F0$ of 12.5 Hz, which corresponds to a period of 80 ms. As shown in an earlier study (Oxenham and Dau, 2001a), if the period of the masker is large compared with the integration period used by the auditory system to detect the signal, compression ceases to be necessary to explain masker phase effects. According to this interpretation, there-

a)Electronic mail: oxenham@mit.edu

fore, the continued masker phase effects at very low signal frequencies may be due simply to the very long masker periods, enabling listening in the valleys (explanation 1, above), rather than to peripheral compression.

The second interpretation relates to whether peripheral compression is indeed reduced at low CFs. The question has not yet been settled conclusively via direct measurements of basilar-membrane motion because of the technical difficulties associated with accessing the apex of the cochlea without causing structural damage (Rhode and Cooper, 1996). As pointed out in an earlier study (Plack and Oxenham, 2000), most psychophysical (Oxenham and Plack, 1997; Nelson *et al.*, 2001) and indirect physiological (Cooper and Yates, 1994) estimates of basilar-membrane compression rely on the assumption that a given place along the basilar membrane responds linearly to tones well below its CF. While this assumption is supported by basilar-membrane measurements in the base of the cochlea, its extrapolation to apical mechanics may not be justified. In fact, recent psychophysical studies using estimates of compression that do not make the assumption of off-frequency linearity have concluded that compression remains strong down to at least 250 Hz (Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003; Plack and O'Hanlon, 2003a). Thus, according to this interpretation, the large masker phase effects at low signal frequencies do in fact reflect strong peripheral compression at low CFs, and earlier studies suggesting reduced compression at apex of the cochlea are in error.

The third interpretation is that the masker phase differences are due to compression, but that the compression is located more centrally than the basilar membrane, at least for low signal frequencies. Along these lines, Zeng and Shannon (1994), using loudness functions derived from cochlear-implant patients, proposed that a lack of peripheral compression at low CFs may be compensated for by stronger central compression. The psychophysical studies inferring strong compression at low frequencies in normal-hearing listeners (Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003; Plack and O'Hanlon, 2003) cannot rule out the possibility that the compression actually occurs at a later stage in the auditory pathway.

This study attempts to distinguish between these interpretations by comparing the results from normal-hearing listeners with those from listeners with sensorineural hearing impairment. The rationale is as follows: if peripheral compression is present at low frequencies, then listeners with cochlear hearing loss at low frequencies are likely to exhibit reduced peripheral compression, presumably due to outer hair cell loss or dysfunction (Oxenham and Plack, 1997; Moore *et al.*, 1999). Thus, if peripheral compression can account for masker phase differences (interpretation 2), then the reduction in masker phase effects found for hearing-impaired listeners at high signal frequencies (Summers and Leek, 1998; Summers, 2000) should also be found at low signal frequencies. Conversely, if the masker phase effects found in normal-hearing listeners at low frequencies are not due to peripheral compression, but to the long masker periods resulting from very low fundamental frequencies (interpretation 1), or to a more central compression (interpretation 3) then a loss of peripheral compression due to hearing loss should not affect results, and so the data from normal-hearing and hearing-impaired listeners should be similar.

The present study also tests the relationship between masker phase effects and auditory filter bandwidths, which have been found to correlate strongly with estimates of peripheral compression (Moore *et al.*, 1999). If peripheral compression underlies masker phase effects at all signal frequencies, then a strong correlation should exist between auditory filter bandwidth and the size of the masker phase effects at all frequencies. On the other hand, if peripheral compression underlies masker phase effects only at high frequencies, then the correlation should exist only at high signal frequencies. Finally, if peripheral compression is not responsible for masker phase effects, we might expect little or no correlation between the two measures.

Data were collected at signal frequencies of 250 and 1000 Hz from 12 hearing-impaired listeners with mild and moderate losses at the test frequencies. Three normal-hearing listeners served as controls. Thresholds were measured at both signal frequencies for harmonic complex maskers with an $F0$ of 12.5 Hz (to measure masker phase effects) and for notched-noise maskers (to measure auditory filter bandwidths). At the higher signal frequency (1000 Hz), thresholds were also measured with a higher masker $F0$ of 100 Hz to allow a more direct comparison with previous studies using hearing-impaired listeners (Summers and Leek, 1998).

II. EXPERIMENT 1: MASKER PHASE EFFECTS AT 250 AND 1000 Hz

A. Methods

1. Stimuli

Thresholds were measured for a sinusoidal signal in the presence of a simultaneous masker. Signal frequencies were 250 and 1000 Hz. The total signal duration was 260 ms, gated with 30-ms raised-cosine ramps and temporally centered in a 320-ms (total duration) masker, which was gated with 10-ms raised-cosine ramps. The signal was added to the masker with a starting phase that was selected randomly from trial to trial. The masker was a harmonic tone complex with components between $0.4f_s$ and $1.6f_s$, where f_s is the signal frequency. The phases of the components were selected according to a modification of Schroeder's (1970) equation, as used previously by Lentz and Leek (2001) and Oxenham and Dau (2001b)

$$\theta_n = C\pi n(n-1)/N, \quad -1 \leqslant C \leqslant 1. \tag{1}$$

A Schroeder positive (m_+) or Schroeder negative (m_-) complex is generated when $C=1$ or $C=-1$, respectively. Schroeder complexes are notable for their very flat temporal envelopes and can be characterized as either a rising (m_-) or falling (m_+) linear frequency glide, which repeats at a rate corresponding to the $F0$. When $C=0$, a sine-phase complex is generated. The phase curvature of the complex is

$$\frac{d^2\theta}{df^2} = C\frac{2\pi}{NF0^2}. \tag{2}$$

TABLE I. Ages and audiometric thresholds of the hearing-impaired listeners.

Subject	Age	Ear tested	Audiometric threshold in test ear (dB HL)					
			250	500	1000	2000	4000	8000
HI1	83	R	45	55	70	60	60	80
HI2	37	L	50	45	45	45	40	45
HI3	80	L	55	60	60	45	45	30
HI4	34	R	30	30	35	45	50	55
HI5	36	R	50	50	55	55	55	65
HI6	60	R	30	40	45	40	45	60
HI7	47	L	35	40	50	60	50	75
HI8	51	L	50	50	45	55	45	35
HI9	33	R	40	50	45	60	85	>110
HI10	36	R	55	55	60	60	65	70
HI11	53	L	50	45	50	50	55	65
HI12	50	R	55	65	55	70	60	55

By varying the C value from -1 to 1, a range of masker phase curvatures, or frequency sweep rates, can be generated. This leads to systematic changes in signal threshold, which are thought to reflect the interactions between the masker phase curvature and that of the basilar membrane, which in turn produce varying degrees of envelope modulation in the auditory response to the masker (Oxenham and Dau, 2001b). The overall level of the masker was set to 93 dB SPL. This relatively high level was selected in an attempt to ensure that all masked thresholds would lie well above the absolute thresholds of the hearing-impaired listeners.

The stimuli were generated digitally at a sampling rate of 48 kHz and converted to analog signals by a 24-bit DAC including reconstruction filtering (SEKD ADSP 2496). The stimuli were presented to the listener via Sennheiser HD 580 headphones. Listeners were seated in a double-walled sound-attenuating booth. Sounds were presented monaurally. For the hearing-impaired listeners the ear with the lower audiometric thresholds at the two test frequencies was chosen for testing.

2. Procedure

An adaptive three-interval three-alternative forced-choice procedure was used in conjunction with a 2-down, 1-up tracking rule to estimate the 70.7%-correct point on the psychometric function (Levitt, 1971). Each interval in a trial was separated by an interstimulus interval (ISI) of 500 ms. The intervals were marked on a computer monitor and feedback was provided after each trial. Listeners responded via the computer keyboard or mouse. The initial step size was 5 dB, which was reduced to 2 dB after the first four reversals. Threshold was defined as the mean of the remaining six reversals. Reported thresholds are the mean of four runs. Absolute thresholds for the two signal frequencies were also collected for each subject, using the same adaptive procedure as for the masked thresholds, with the exception that each threshold reported is the mean of three, rather than four, runs. The conditions were presented in random order, with all conditions being presented before embarking on the next repetition. The order of presentation of the conditions was selected randomly for each listener and each repetition. Measurements were made in 2-h sessions, including many short breaks. No more than one session per listener was completed in any one day.

3. Listeners

Twelve hearing-impaired and three normal-hearing listeners took part in the study. The second author served as one of the normal-hearing listeners (NH3). All the hearing-impaired listeners were paid for their participation, while the other two normal-hearing listeners were students who participated on a voluntary basis. The hearing-impaired listeners all had bilateral hearing loss. All losses were diagnosed as being of cochlear origin, as indicated by lack of an air–bone gap and evidence of loudness recruitment. Audiometric thresholds for the hearing-impaired listeners, along with their ages, are shown in Table I. The normal-hearing listeners were aged between 25 and 36 years and had audiometric thresholds of no more than 10 dB HL at the test frequencies. All the listeners (both hearing impaired and normal hearing) had some previous experience in psychoacoustic tests, including masking experiments. They were given practice until their performance appeared to be stable. This generally took about 2 h.

B. Results

The results are shown in Fig. 1 for the individual subjects. Data from the three normal-hearing listeners are shown in the bottom row; the remaining data are from the 12 hearing-impaired listeners. Circles represent thresholds with a 250-Hz signal; squares represent thresholds with a 1000-Hz signal. Filled and open symbols denote thresholds using a 12.5-Hz and 100-Hz masker $F0$, respectively. Absolute thresholds are also shown for the hearing-impaired listeners and are shaded gray. Absolute thresholds for the normal-hearing listeners were all below 20 dB SPL (mean thresholds: 12.1 dB SPL at 250 Hz and 6.4 dB SPL at 1000 Hz). With the exception of HI12 at 1000 Hz, masked thresholds were always at least 6 dB (and in most cases more than 10 dB) above absolute threshold.

In the normal-hearing listeners, the pattern of results is similar to that found in previous studies (e.g., Oxenham and Dau, 2001b): thresholds vary considerably with masker

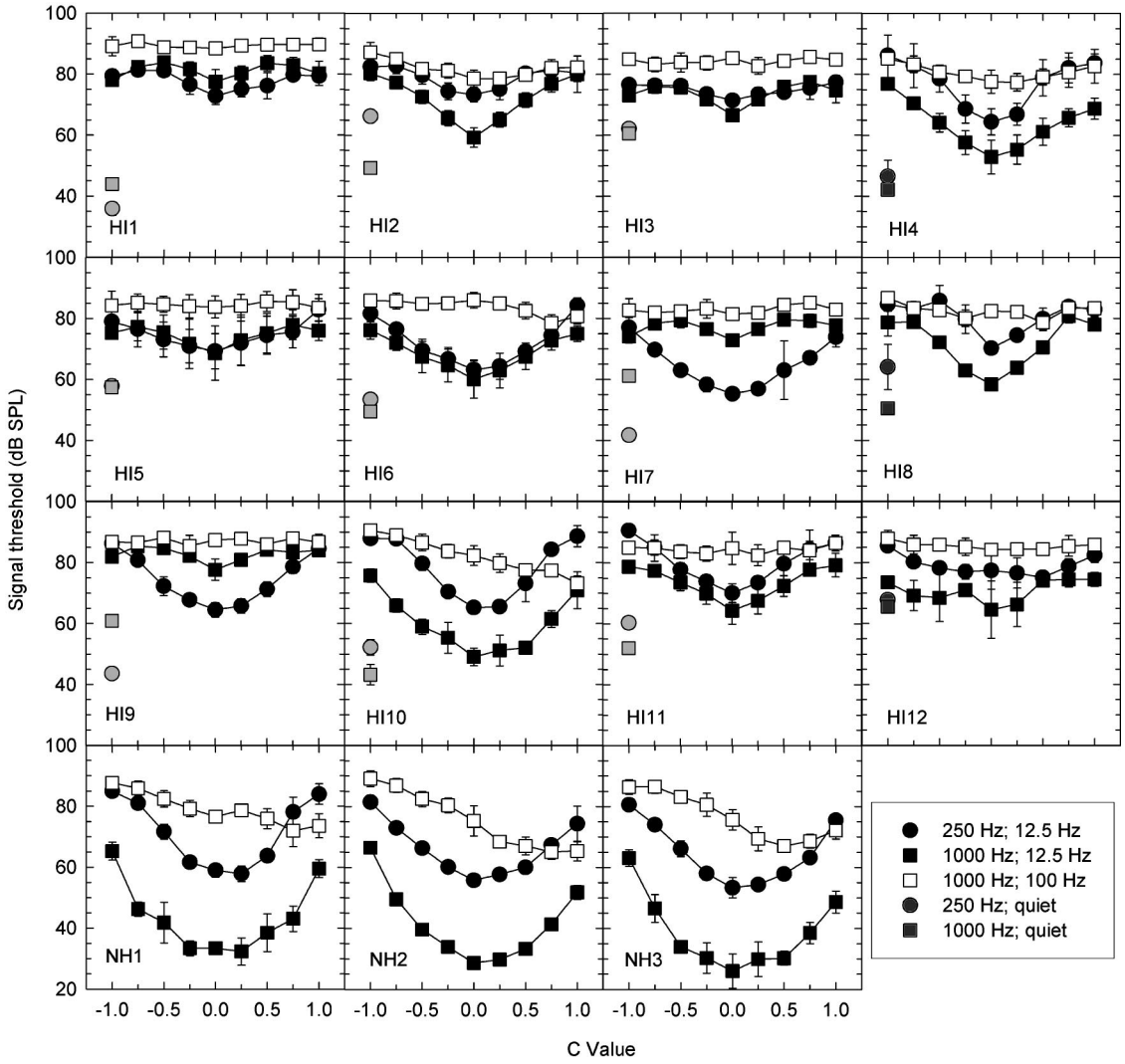


FIG. 1. Individual masked thresholds in the harmonic tone complex maskers, as a function of masker phase curvature, denoted by C. Different symbols represent different conditions, as shown in the legend. Error bars represent ± 1 s.d. of the mean.

phase curvature and reach a minimum at a C value that depends on the interaction between the masker phase curvature and the phase curvature of the auditory periphery. In good agreement with the current data, phase curvature estimates from our previous study (Oxenham and Dau, 2001b) predict that the C value at the minimum for the 1000-Hz signal should be around 1 for the 100-Hz $F0$ and between 0 and 0.25 for the 12.5-Hz $F0$. The minimum around 0–0.25 found with a signal frequency of 250 Hz is also in good agreement with the results of our earlier study. The larger difference between maximum and minimum threshold found with decreasing $F0$ at a given signal frequency has been reported in earlier studies (Kohlrausch and Sander, 1995; Oxenham and Dau, 2001b).

Results from the hearing-impaired listeners show a somewhat different pattern of results. Where minima in the masking functions are visible, they tend to occur at the same C values as for normal-hearing listeners. In general, however, the effect of masker phase curvature on thresholds is much reduced in the hearing-impaired listeners at both signal frequencies. The effects of masker phase relationships were quantified for each listener by subtracting the minimum from the maximum masked threshold in a given condition. These differences are shown in Table II. It can be seen that most hearing-impaired listeners exhibit reduced masker phase effects in all three conditions. These data and their implications for the possible underlying mechanisms are considered further below (Sec. IV).

C. Effects of background noise in normal hearing

In general, hearing loss influences masker phase effects at both 250 and 1000 Hz. Following from the reasoning laid out in the Introduction, this suggests that cochlear compression may remain relatively strong in normal hearing, even at frequencies as low as 250 Hz, and that this compression may be reduced in listeners with hearing impairment. However, before proceeding under this assumption, it is important to test for the possibility that the effects can be accounted for simply by the elevated absolute thresholds in the hearing-impaired listeners. Studies attempting to simulate hearing impairment by using a background noise to elevate thresholds have come to various conclusions. Although most researchers agree that cochlear hearing impairment cannot be

TABLE II. Difference in dB between the maximum and minimum thresholds in each condition of experiment 1. This value provides an indication of the overall masker phase effect in individual listeners.

	Condition		
Listener	250 Hz; 12.5 Hz	1000 Hz; 12.5 Hz	1000 Hz; 100 Hz
HI1	8.5	6.3	2.4
HI2	9.4	21.0	8.7
HI3	5.9	10.9	2.9
HI4	21.8	24.0	7.8
HI5	13.6	7.3	2.1
HI6	21.1	16.1	7.5
HI7	21.8	6.8	3.8
HI8	15.6	22.4	8.1
HI9	21.8	7.8	2.8
HI10	23.4	26.7	17.4
HI11	20.6	14.9	4.0
HI12	10.3	9.9	3.7
NH1	27.0	32.9	15.7
NH2	25.7	37.8	24.1
NH3	26.2	37.1	19.4

recreated in normal hearing simply by adding masking noise (Fabry and Van Tasell, 1986; Leek *et al.*, 1987; Dubno and Schaefer, 1992; Leek and Summers, 1993), there have been some studies showing certain aspects of the impairment can be successfully simulated in that way (e.g., Humes *et al.*, 1987; Zurek and Delhorne, 1987; Dubno and Schaefer, 1992). Also, masker phase effects in normal-hearing listeners are reduced at low overall levels (Carlyon and Datta, 1997b; Oxenham and Dau, 2001b). This has been assumed to be due to more linear cochlear processing at low levels, but might conceivably be due to some other cause associated with low sensation levels.

1. Methods

To test the effect of sensation level on masker phase effects, the three normal-hearing listeners from experiment 1 were tested again in an additional background of white Gaussian noise, designed to raise the threshold of the signal to levels comparable to the absolute thresholds of some of our hearing-impaired listeners. Signal thresholds were measured for all the harmonic tone complex conditions tested in the main part of experiment 1. The stimuli and procedure were identical to those used there, with the exception that a Gaussian white noise, bandpass filtered between 100 and 6000 Hz with a spectrum level of 27 dB SPL in the passband, was gated on and off synchronously with the complex-tone masker. This raised detection thresholds of the signal in the noise alone to around 45 dB SPL at both signal frequencies for the three normal-hearing listeners.

2. Results

Individual results from the three normal-hearing listeners are shown in the three rows of Fig. 2. Columns 1–3 represent the 250-Hz, 1000-Hz (12.5-Hz $F0$), and 1000-Hz (100-Hz $F0$), respectively. Corresponding data from experiment 1, without the background noise, are replotted in each panel as lines without symbols. Gray symbols show the signal thresholds in the noise alone for the 250-Hz (column 1) and 1000-Hz (columns 2 and 3) signals. In the 250-Hz signal condition (left column) and the 1000-Hz signal condition with the 100-Hz masker $F0$ (right column), the additional noise masker has essentially no effect on masked thresholds. It is only in the 1000-Hz signal condition with the 12.5-Hz masker, where thresholds in the complex-tone masker drop below the signal threshold in noise alone, that the effect of

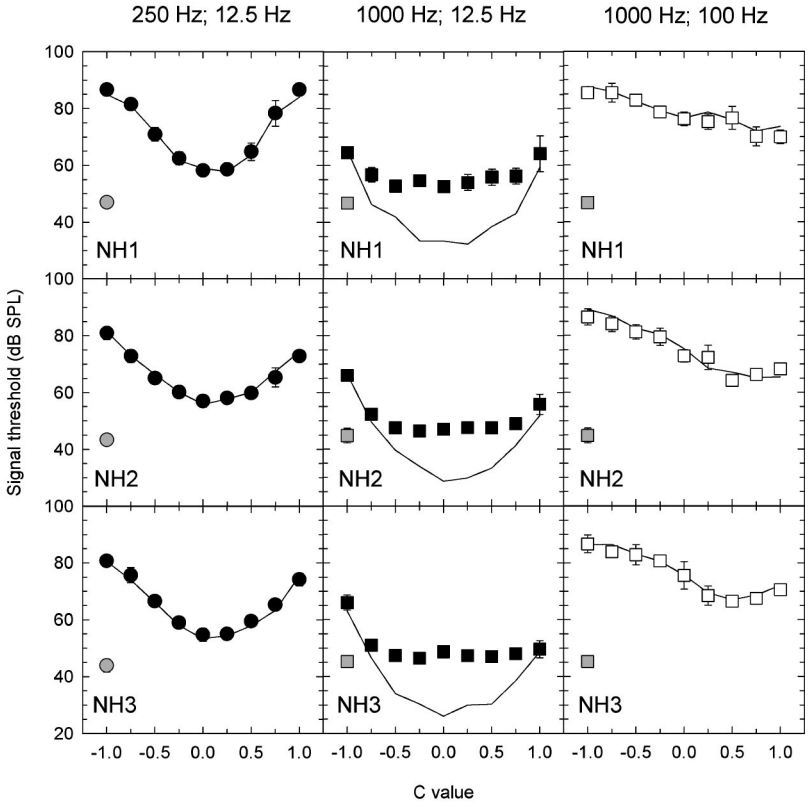


FIG. 2. Individual masked thresholds from three normal-hearing listeners in the harmonic tone complex maskers, combined with a Gaussian white noise. Each column presents data from a different condition, as shown in the column headers. Thresholds in the noise alone are shown as gray symbols; thresholds in the harmonic tone complex alone are replotted from Fig. 1 as lines.

the noise masker becomes apparent. Even here, it is only when the signal threshold would otherwise be within about 5 dB of its masked threshold in the noise alone that thresholds are affected.

The pattern of results found in the normal-hearing listeners with an additional background noise does not resemble the general pattern of results of the hearing-impaired listeners (see Fig. 1). With the hearing-impaired listeners, masker phase effects were generally less than those found for normal-hearing listeners, even though masked thresholds for the hearing-impaired listeners were in most cases more than 10 dB above absolute threshold. Thus, it appears that the signal's reduced sensation level cannot account for the differences found between normal-hearing and hearing-impaired listeners. This is interesting as it provides an example of an auditory task in which additive noise does not provide an adequate simulation of moderate hearing loss in normal-hearing listeners. Other examples include temporal integration, or the decrease in signal threshold with increasing signal duration (e.g., Oxenham *et al.*, 1997), and the ability to recognize speech in amplitude-modulated backgrounds (Eisenberg *et al.*, 1995).

TABLE III. Individual equivalent rectangular bandwidths (ERBs) of auditory filters measured at 250 and 1000 Hz, using a fixed spectrum-level noise of 60 and 50 dB SPL, respectively. The values are normalized to the signal frequency (i.e., ERB/f_s). The values in brackets are the rms errors (in dB) of the signal threshold predictions using the fitted filter function, which was a symmetric roex(p) filter. Missing values (···) indicate conditions in which more than one data point was within 5 dB of absolute threshold.

	Signal frequency	
Listener	250 Hz	1000 Hz
HI1	0.43 (0.47)	1.23 (1.20)
HI2	0.92 (0.55)	0.38 (1.47)
HI3	0.52 (1.27)	···
HI4	0.37 (0.77)	0.27 (2.09)
HI5	0.41 (0.89)	0.36 (1.18)
HI6	0.38 (1.10)	0.32 (2.13)
HI7	0.35 (0.55)	2.00 (0.68)
HI8	0.44 (0.71)	0.35 (1.32)
HI9	0.43 (0.49)	2.00 (0.45)
HI10	0.29 (0.60)	0.16 (1.39)
HI11	0.79 (1.16)	0.57 (1.10)
HI12	···	···
NH3	0.39 (0.84)	0.24 (1.71)

III. EXPERIMENT 2: AUDITORY FILTER SHAPES AT 250 AND 1000 Hz

This experiment measured auditory filter bandwidths at 250 and 1000 Hz in the same 12 hearing-impaired listeners as used in experiment 1. Only one normal-hearing listener (NH3) was tested. As mentioned in the Introduction, if a loss of peripheral compression in the hearing-impaired listeners can account for the reduced masker phase effects, then we should find a correlation between the difference in maximum and minimum thresholds found in experiment 1 and the auditory filter bandwidths, which are also thought to be highly correlated with peripheral compression (Moore *et al.*, 1999).

A. Method

The signals and the temporal parameters of both the masker and signal were identical to those used in experiment 1. However, instead of the masker being a harmonic tone complex, it consisted of two spectral bands of Gaussian noise, each with a bandwidth of $0.4f_s$, centered below and above the signal frequency. The two noise bands were either contiguous (0 notch condition) or were spaced apart in frequency, thereby introducing a spectral notch between them. The notches were defined in terms of the frequency separation (Δf) between the signal frequency and the nearer edge of one of the noise bands, normalized by the signal frequency (i.e., $\Delta f/f_s$). The noise bands were placed either symmetrically or asymmetrically around the signal frequency. In the symmetric conditions, the notch values were 0, 0.2, and 0.4. In the asymmetric conditions, the notch on one side of the signal was 0.2, while on the other side it was 0.4. This led to the same five conditions (three symmetric and two asymmetric) that were deemed by Stone *et al.* (1992) sufficient to produce reasonably accurate estimates of auditory filter shape. The spectrum level of the masker was 60 dB SPL at 250 Hz and 50 dB SPL at 1000 Hz, producing an overall level of around 83 dB SPL and 80 dB SPL at 250 and 1000 Hz, respectively.

The method of testing was also identical to that used in experiment 1, and again, four repetitions of each condition were averaged to compute the individual threshold estimates. Two exceptions occurred (both with subject HI10), where one run resulted in a threshold that exceeded the mean of the other three runs by more than 4 standard deviations. These two outliers were discarded and the mean of the remaining three runs in each of the two conditions was recorded.

B. Results

As expected from the extensive literature on the topic (e.g., Patterson *et al.*, 1982; Tyler *et al.*, 1984; Moore and Glasberg, 1986; Dubno and Schaefer, 1995; Moore *et al.*, 1999; Baker and Rosen, 2002), changes in the masked signal thresholds of the hearing-impaired listeners as a function of masker notch width were less than those found for normal-hearing listeners. Masked thresholds were on the whole well above the absolute thresholds of the individual listeners. On average, the lowest thresholds in the notched-noise conditions exceeded absolute thresholds by more than 10 dB. However, in 4 of the 12 hearing-impaired listeners, at least one masked threshold fell within 5 dB of absolute threshold. For the purposes of analysis, these thresholds were excluded from the fitting procedure. If more than one of the five thresholds within a condition was excluded, a fit was not attempted. This procedure resulted in the exclusion of the 1000-Hz data of listener HI3 and of both the 250-Hz and 1000-Hz data of listener HI12.

Frequency selectivity was quantified in the individual listeners by deriving auditory filter shapes from the data, using the rounded-exponential, roex(p), model (e.g., Patterson *et al.*, 1982). Initially, we allowed the parameter p, which defines the slope of the filter function, to vary inde-

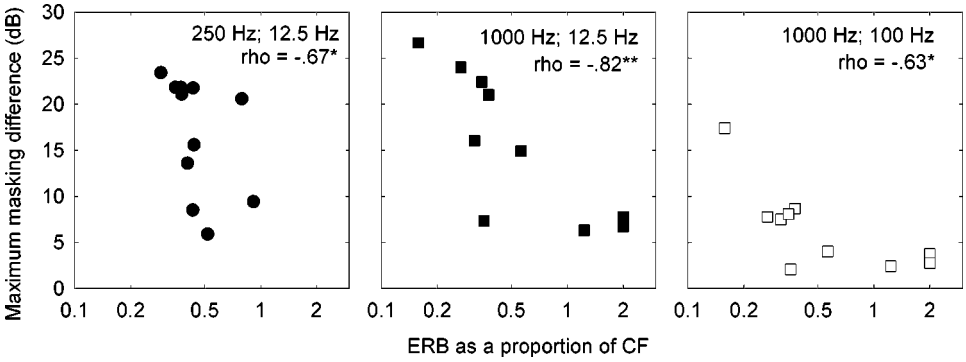


FIG. 3. Maximum masking difference in harmonic tone complex maskers, plotted as a function of ERB. Each data point represents one hearing-impaired listener. The frequencies shown are the signal frequency and masker $F0$, respectively. The Spearman rank correlation coefficient is also shown in each panel. One and two asterisks denote statistical significance at the $p=0.05$ and $p=0.01$ levels, respectively (one-tailed tests).

pendently on the lower and upper side of the filter. However, for some of the hearing-impaired listeners, where the slope of the masking function was very shallow, this led to very poorly defined filter estimates, probably because of the small number of data points (5) used to fit the model. We then restricted the model to a single free parameter, p, which defined the slope on both sides of the filter. By using only one parameter, we assume the filter is symmetric. Note, however, that this does not imply that the model cannot predict threshold differences in the asymmetric notch conditions. This is because the model ''detects'' the signal in the filter with the lowest signal-to-noise ratio, which does not necessarily have a CF corresponding to the signal frequency (''off-frequency listening''). Furthermore, this model simplification led to only a small increase in prediction errors (mean increase in rms error <0.2 dB at both 250 and 1000 Hz).

The results from the fitting procedure were used to derive equivalent rectangular bandwidths (ERBs) for each listener at each of the two signal frequencies. These values are given in Table III, normalized by the signal frequency. The ERBs were calculated by integrating the area under the derived filter from 0 to 2 f_s Hz. Thus, the maximum possible normalized ERB value was 2, which essentially implies no measurable frequency selectivity. This was the case for 2 of the 12 subjects (HI7 and HI9) at a signal frequency of 1000 Hz. For these two subjects, thresholds for a total notch width of 800 Hz were within 1 or 2 dB of those found with no notch. The ERBs from the normal-hearing listener are larger than the normative values provided by Glasberg and Moore (1990) of 0.21 and 0.13 at 250 and 1000 Hz, respectively. This is at least in part due to the high overall level used in the current experiments, which is known to increase estimated filter bandwidths.

IV. ANALYSIS

Experiment 1 found that masker phase effects were greatly reduced in hearing-impaired listeners at both 250 and 1000 Hz. A previous study showed a similar lack of phase effects in hearing-impaired listeners (Summers and Leek, 1998). However, in that study, only two masker phase curvatures were tested ($C=1$ or -1), leaving open the possibility that the smaller differences may have reflected only a shift along the C axis in the maximum and minimum of the masking functions, rather than a reduction in the size of the effect itself. Such a shift might result from a change in the phase response of the impaired cochlea. For instance, for the 1000-Hz signal in the 100-Hz $F0$ masker, the minimum of the masking function might shift to $C=0$ if the phase curvature of the impaired cochlea tended toward zero. Although there might be a hint of such a shift in 2 of the 12 listeners (HI2 and HI4), two other listeners (HI6 and HI10) show more normal minima, while the other 8 listeners show no reliable minima. Thus, the reduced phase effects in the hearing-impaired listeners do not seem to be due to a simple change in the phase curvature of the impaired cochlea. It is possible that a more complex change in the phase response, resulting in markedly nonuniform phase curvature in the impaired cochlea, could lead to none of the masker waveforms (which all have constant phase curvature) achieving maximal ''peakiness.'' In that case, at least part of the reduction in phase effects could be due to complex changes in the phase response, rather than to changes in compression. However, current data suggest that the phase curvature of the normal cochlea is rather invariant over a large range of levels (Oxenham and Dau, 2001b; Shera, 2001), suggesting that perhaps it may remain relatively invariant after cochlear damage.

If the size of the masker phase effect instead provides an indirect estimate of cochlear nonlinearity or compression, then one might expect a strong correlation between it and auditory filter bandwidth estimates in the same subjects. This is because both measures are thought to reflect the functioning of the cochlea's outer hair cells. To test this, the relation between the masker phase effect, quantified using the maximum–minimum difference in threshold shown in Table II, and the ERBs given in Table III, was analyzed. Both measures were also compared with individual absolute thresholds. These relationships for the hearing-impaired listeners are shown in Figs. 3, 4, and 5. The numbers in the panels show the stimulus parameters and the Spearman rank correlation coefficients.

In Fig. 3, the maximum–minimum masking difference is plotted as a function of the ERB values, on a log scale. Correlations in all three conditions are significant, although the relationships are not very strong, with all three conditions exhibiting considerable scatter. Similarly weak but significant correlations were found between absolute threshold and ERB (Fig. 4), as has been found previously (e.g., Moore *et al.*, 1999). The relationship between absolute threshold and the maximum–minimum masking difference (Fig. 5) appeared to be weaker, with only two of the three conditions exhibiting a statistically significant relationship.

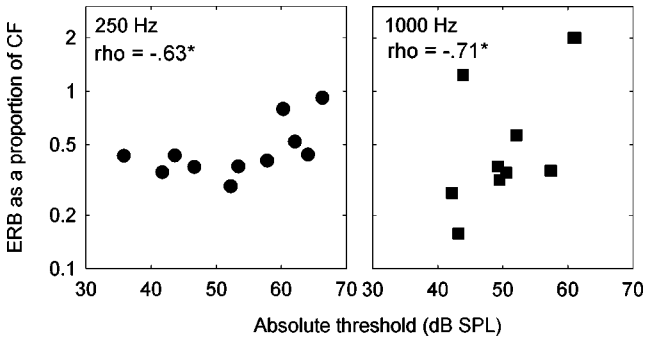


FIG. 4. Individual ERBs, plotted as a function of absolute threshold. Stimulus parameters and correlation coefficients are shown, as in Fig. 3.

Other potential correlations were also examined as controls. No measure at 250 Hz (absolute thresholds, maximum–minimum threshold difference, or ERB) was significantly correlated with any measure at 1 kHz. There were also no strong relationships between listener age and psychoacoustic measure. The correlation coefficient between listener age and maximum–minimum masking difference at 250 Hz did reach statistical significance (rho$=0.62$; $p=0.015$, one-tailed test; uncorrected for multiple comparisons), but the six other potential correlations with age (ERB at 250 and 1000 Hz, maximum masking difference at 1000 Hz for 12.5 and 100 Hz $F0$, absolute threshold at 250 and 1000 Hz) were all below rho$=0.3$ ($p>0.2$).

V. DISCUSSION

The results of experiment 1 suggest that the large masker phase effects observed at very low signal frequencies in our earlier study (Oxenham and Dau, 2001b) were not simply due to the low $F0$ used. Instead, peripheral compression is likely to play a role. This conclusion is based on the fact that hearing-impaired listeners exhibited similarly reduced masker phase effects with an $F0$ of 12.5 Hz at both signal frequencies of 250 and 1000 Hz.

Our conclusion is in line with other recent studies, presenting evidence for strong compression in normal hearing at very low frequencies (Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003; Plack and O'Hanlon, 2003). The results of the present study extend these findings by showing that compression at low frequencies seems to be affected by cochlear hearing loss in the same way as compression at high frequencies. This makes retrocochlear compression (e.g., Zeng and Shannon, 1994) a less likely explanation for the data from normal-hearing listeners in this and other studies of low-frequency compression.

According to our hypothesis, frequency selectivity and masker phase sensitivity are both largely determined by the functioning of the cochlea's outer hair cells. In contrast, absolute threshold may reflect a number of different factors, including outer hair cell and inner hair cell loss or dysfunction, as well as possible retrocochlear components that are not easily distinguished from cochlear hearing loss. Thus, we might have expected ERB values and the maximum masking difference to be more strongly correlated with each other than with absolute threshold (see also Moore *et al.*, 1999). This was not the case: essentially all three measures were weakly, but mostly significantly, correlated with each other. One explanation is that the measure of maximum masking difference does not reflect solely the functional status of the cochlear outer hair cells, as is likely to be the case for the ERB values. For instance, detection of a tone in a fluctuating periodic masker may involve some degree of temporal resolution. Although temporal resolution is generally not thought to be strongly affected by cochlear hearing loss, there exist certain patients, currently diagnosed with ''auditory neuropathy'' (Sininger and Starr, 2001), who exhibit extremely poor temporal resolution, as measured using amplitude modulation detection and gap detection (Zeng *et al.*, 1999, 2001). It is at least conceivable that some more subtle temporal deficits, which are not apparent in standard tests for cochlear hearing loss, may affect thresholds in our masker phase thresholds, but not ERBs.

Another possible explanation for the relatively weak correlations is that the relationship between masker phase effects and the degree of outer hair cell loss is complicated by the contribution of filter bandwidth. Increasing the effective filter bandwidth leads to more masker components interacting and hence to greater maximum peakiness of the temporal envelope, which in turn leads to larger potential masking differences. Thus, outer hair cell damage or dysfunction results in a loss of peripheral compression, which may *decrease* masker phase effects, but also results in a widening of filter bandwidths, which may *increase* masker phase effects. It is conceivable that these two competing effects may interact differently in different listeners, resulting in weaker-than-expected correlations.

In summary, the maximum masking difference does not provide an easily interpretable quantitative estimate of co-

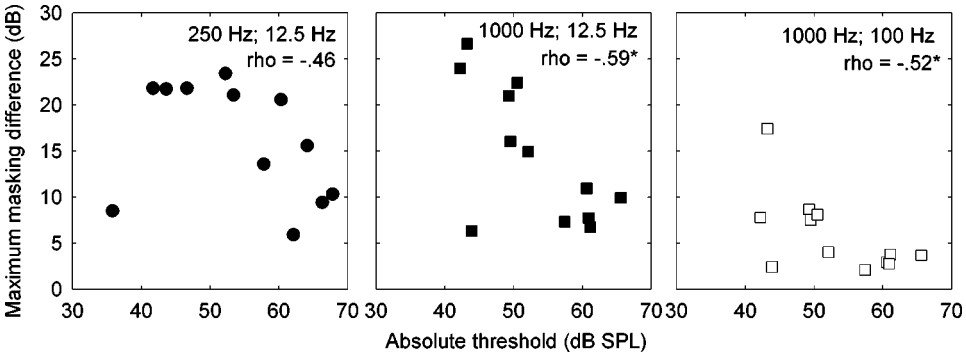


FIG. 5. Individual maximum masking differences, plotted as a function of absolute threshold. Stimulus parameters and correlation coefficients are shown, as in Fig. 3.

chlear compression; measures such as masking additivity (Plack and O'Hanlon, 2003), and temporal masking curves (Lopez-Poveda *et al.*, 2003; Plack and Drga, 2003) are probably more suitable. Nevertheless, while masker phase masking differences are not easily quantified in terms of amount of cochlear compression, the results from our hearing-impaired and normal-hearing listeners do provide reasonably strong qualitative support for continued peripheral compression at low CFs in normal hearing.

VI. SUMMARY AND CONCLUSIONS

The results and analysis from experiments 1 and 2 can be summarized as follows:

(1) At both signal frequencies tested (250 and 1000 Hz), effects of masker phase on signal thresholds were substantially less in hearing-impaired listeners than in normal-hearing listeners. This suggests that the masker phase differences found in normal hearing at low signal frequencies may be due to continued strong compression in apical regions of the normal cochlea, which is reduced or no longer present in the impaired cochlea.
(2) Additive white noise did not materially affect masker phase differences in normal-hearing listeners, unless the signal threshold was within 5 dB of its threshold in the noise alone. This suggests that the reduction in masker phase effects with hearing impairment is not due simply to the lower sensation level at which the stimuli were presented to the hearing-impaired listeners. The results provide a further example of a situation where moderate hearing loss is not adequately simulated by presenting masking noise to normal-hearing listeners.
(3) Correlations between the auditory filter bandwidth (ERB) and the maximum masking differences in the hearing-impaired listeners were statistically significant, but not overwhelming, suggesting that thresholds in the two tasks are mediated by overlapping, but not necessarily identical, mechanisms.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIDCD Grant R01 DC 03909 to A.J.O.) and the Deutsche Forschungsgemeinschaft (T.D.). The experiments were carried out at the laboratory of the work group Medizinische Physik at Oldenburg University while the first author was a fellow at the Hanse Institute for Advanced Study, Delmenhorst, Germany. We thank Müge Kaya from the Hörzentrum Oldenburg for assistance in recruiting and collecting data from the hearing-impaired listeners. Comments by Gerald Kidd and the reviewers helped improve the paper.

Baker, R. J., and Rosen, S. (**2002**). ''Auditory filter nonlinearity in mild/moderate hearing impairment,'' J. Acoust. Soc. Am. **111**, 1330–1339.
Buus, S. (**1985**). ''Release from masking caused by envelope fluctuations,'' J. Acoust. Soc. Am. **78**, 1958–1965.
Carlyon, R. P., and Datta, A. J. (**1997a**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.
Carlyon, R. P., and Datta, A. J. (**1997b**). ''Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components,'' J. Acoust. Soc. Am. **101**, 3648–3657.
Cooper, N. P., and Yates, G. K. (**1994**). ''Nonlinear input–output functions derived from the responses of guinea-pig cochlear nerve fibers: Variations with characteristic frequency,'' Hear. Res. **78**, 221–234.
Dubno, J. R., and Schaefer, A. B. (**1992**). ''Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners,'' J. Acoust. Soc. Am. **91**, 2110–2121.
Dubno, J. R., and Schaefer, A. B. (**1995**). ''Frequency selectivity and consonant recognition for hearing-impaired and normal-hearing listeners with equivalent masked thresholds,'' J. Acoust. Soc. Am. **97**, 1165–1174.
Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). ''Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing,'' J. Speech Hear. Res. **38**, 222–233.
Fabry, D. A., and Van Tasell, D. J. (**1986**). ''Masked and filtered simulation of hearing loss: Effects on consonant recognition,'' J. Speech Hear. Res. **29**, 170–178.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Hicks, M. L., and Bacon, S. P. (**1999**). ''Psychophysical measures of auditory nonlinearities as a function of frequency in individuals with normal hearing,'' J. Acoust. Soc. Am. **105**, 326–338.
Humes, L. E., Dirks, D. D., and Kincaid, G. E. (**1987**). ''Recognition of nonsense syllables by hearing-impaired listeners and by noise masked normal listeners,'' J. Acoust. Soc. Am. **81**, 765–773.
Kohlrausch, A., and Sander, A. (**1995**). ''Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets,'' J. Acoust. Soc. Am. **97**, 1817–1829.
Leek, M. R., and Summers, V. (**1993**). ''Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise,'' J. Acoust. Soc. Am. **94**, 3127–3137.
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.
Lentz, J. J., and Leek, M. R. (**2001**). ''Psychophysical estimates of cochlear phase response: Masking by harmonic complexes,'' J. Assoc. Res. Otolaryngol. **2**, 408–422.
Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). ''Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing,'' J. Acoust. Soc. Am. **113**, 951–960.
Moore, B. C. J., and Glasberg, B. R. (**1986**). ''Comparisons of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments,'' J. Acoust. Soc. Am. **80**, 93–107.
Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). ''Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism,'' J. Acoust. Soc. Am. **106**, 2761–2778.
Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). ''A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **110**, 2045–2064.
Oxenham, A. J., and Dau, T. (**2001a**). ''Reconciling frequency selectivity and phase effects in masking,'' J. Acoust. Soc. Am. **110**, 1525–1538.
Oxenham, A. J., and Dau, T. (**2001b**). ''Towards a measure of auditory-filter phase response,'' J. Acoust. Soc. Am. **110**, 3169–3178.
Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Oxenham, A. J., Moore, B. C. J., and Vickers, D. A. (**1997**). ''Short-term temporal integration: Evidence for the influence of peripheral compression,'' J. Acoust. Soc. Am. **101**, 3676–3687.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). ''The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold,'' J. Acoust. Soc. Am. **72**, 1788–1803.
Plack, C. J., and Drga, V. (**2003**). ''Psychophysical evidence for auditory compression at low characteristic frequencies,'' J. Acoust. Soc. Am. **113**, 1574–1586.
Plack, C. J., and O'Hanlon, C. G. (**2003**). ''Forward masking additivity and auditory compression at low and high frequencies,'' J. Assoc. Res. Otolaryngol. **4**, 405–415.
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., and Cooper, N. P. (**1996**). ''Nonlinear mechanics in the apical turn of the chinchilla cochlea in vivo,'' Aud. Neurosci. **3**, 101–121.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Schroeder, M. R. (**1970**). ''Synthesis of low peak-factor signals and binary sequences with low autocorrelation,'' IEEE Trans. Inf. Theory **16**, 85–89.
Shera, C. A. (**2001**). ''Intensity invariance of fine time structure in basilar-membrane click response: Implications for cochlear mechanics,'' J. Acoust. Soc. Am. **110**, 332–348.
Sininger, Y., and Starr, A. (**2001**). *Auditory Neuropathy*. (Singular, San Diego).
Stone, M. A., Glasberg, B. R., and Moore, B. C. J. (**1992**). ''Simplified measurement of impaired auditory filter shapes using the notched-noise method,'' Br. J. Audiol. **26**, 329–334.
Summers, V. (**2000**). ''Effects of hearing impairment and presentation level on masking period patterns for Schroeder-phase harmonic complexes,'' J. Acoust. Soc. Am. **108**, 2307–2317.
Summers, V., and Leek, M. R. (**1998**). ''Masking of tones and speech by Schroeder-phase harmonic complexes in normally hearing and hearing-impaired listeners,'' Hear. Res. **118**, 139–150.
Tyler, R. S., Hall, J. W., Glasberg, B. R., Moore, B. C. J., and Patterson, R. D. (**1984**). ''Auditory filter asymmetry in the hearing impaired,'' J. Acoust. Soc. Am. **76**, 1363–1368.
Zeng, F.-G., and Shannon, R. V. (**1994**). ''Loudness-coding mechanisms inferred from electric stimulation of the human auditory system,'' Science **264**, 564–566.
Zeng, F. G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (**1999**). ''Temporal and speech processing deficits in auditory neuropathy,'' NeuroReport **10**, 3429–3435.
Zeng, F. G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (**2001**). ''Psychoacoustics and speech perception in auditory neuropathy,'' in *Auditory Neuropathy*, edited by Y. Sininger and A. Starr (Singular, San Diego).
Zurek, P. M., and Delhorne, L. A., (**1987**). ''Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment,'' J. Acoust. Soc. Am. **82**, 1548–1559.

Effects of pulse rate and electrode array design on intensity discrimination in cochlear implant users[a)]

Heather A. Kreft,[b)] Gail S. Donaldson, and David A. Nelson
Clinical Psychoacoustics Laboratory, University of Minnesota, Department of Otolaryngology, MMC 396, Rm 8-323 PWB, 420 Delaware St. SE, Minneapolis, Minnesota 55455

(Received 29 September 2003; revised 8 July 2004; accepted 9 July 2004)

The effects of pulse rate on intensity discrimination were evaluated in 14 subjects with Clarion C-I cochlear implants. Subjects had a standard [Clarion spiral electrode array (SPRL group)] or perimodiolar electrode array [Clarion HiFocus electrode array with electrode positioning system (HF+EPS group)]. Weber fractions for intensity discrimination [$Wf_{\mathrm{dB}}=10\log\Delta I/I$] were evaluated at five levels over dynamic range at each of three pulse rates (200, 1625 and 6500 pps) using monopolar stimulation. Weber fractions were smaller for 200 pps stimuli than for 1625 or 6500 pps stimuli in both groups. Weber fractions were significantly smaller for SPRL subjects (mean $Wf\mathrm{dB}=-9.1$ dB) than for HF+EPS subjects (mean $Wf\mathrm{dB}=-6.7$ dB). Intensity difference limens (DLs) expressed as a percentage of dynamic range (DR) ($\Delta I_{\%\mathrm{DR}}=\Delta I/\mathrm{DR}_{\mathrm{dB}}*100$) did not vary systematically with pulse rate in either group. Larger intensity DLs combined with smaller dynamic ranges led to fewer intensity steps over the dynamic range for HF+EPS subjects (average 9 steps) compared to SPRL subjects (average 23 steps). The observed effects of pulse rate and electrode array design may stem primarily from an inverse relationship between absolute current amplitude and the size of intensity DLs. The combination of smaller dynamic ranges and larger Weber fractions in HF+EPS subjects could be the result of increased variability of neural outputs in these subjects. [DOI: 10.1121/1.1786871]

PACS numbers: 43.66.Fe, 43.66.Ts [NFV] Pages: 2258–2268

I. INTRODUCTION

Cochlear implants have undergone several changes over the past decade that contribute to higher levels of benefit for their recipients. One such change is the use of faster rates of pulsatile stimulation. Stimulation rates have increased from approximately 250 pulses per second (pps) per electrode in first-generation cochlear implants to 2000 to 5000 pps per electrode in contemporary devices. Another important change is the use of electrode arrays that are designed to position electrode contacts close to the cochlear modiolus. Perimodiolar electrode arrays have several advantages over traditional electrode arrays, including lower current requirements, greater insertion depth, and the potential for less overlap of current fields generated by adjacent electrodes.

Previous studies have shown that pulse rate and electrode array design can influence threshold (THS), maximum acceptable loudness level (MAL), and dynamic range (DR) for electrical stimulation in cochlear implant users. Kreft *et al.* (2004) found that increasing pulse rate from 200 to 6500 pps in a group of Clarion cochlear implant listeners reduced THSs by an average of 11 dB and reduced MALs by an average of 3 dB, thereby increasing DR by an average of 8 dB. Similar findings have been reported by other investigators for narrower ranges of pulse rates (Simmons, 1966; Shannon, 1985; Shannon, 1989; Skinner *et al.*, 2000; Vandali *et al.*, 2000). The Kreft *et al.* (2004) study also compared THSs and MALs for subjects with a standard electrode array (Clarion Spiral electrode array [SPRL]) to those for subjects with a perimodiolar electrode array (Clarion HiFocus electrode array with electrode positioning system [HF+EPS]). Average THSs were similar for subjects with the two types of electrode arrays; however, average MALs were 6 dB lower for subjects with the HF+EPS array than for those with the SPRL array. As a result, average DRs were considerably smaller for subjects with the HF+EPS electrode array (9.2 dB) than for subjects with the SPRL electrode array (15.2 dB). This finding is consistent with previous studies indicating substantially lower MALs (or comfortable loudness levels) in subjects with perimodiolar electrode arrays as compared to those with traditional electrode arrays (Donaldson *et al.*, 2001; Osberger *et al.*, 1999; Lesinski-Shiedat *et al.*, 2000; Lenarz *et al.*, 2001; Young and Grohne, 2001; Xu and Pfingst, 2002; Parkinson *et al.*, 2002; Saunders *et al.*, 2002).

The present study was designed to extend the findings of Kreft *et al.* (2004) to measures of intensity discrimination. Previous studies have shown that cochlear implant listeners have substantially fewer discriminable intensity steps across the DR of hearing than normal-hearing acoustic listeners (Nelson *et al.*, 1996), at least for low pulse rates and traditional electrode arrays. Although reduced intensity resolution does not appear to have a substantial effect on speech recognition in quiet (Loizou *et al.*, 2000b), it is likely to have a greater impact on speech recognition in noise, music appreciation, and binaural tasks such as sound source localization.

It is not known how pulse rate or electrode array design influences intensity discrimination for cochlear implant lis-

[a)]Portions of these data were presented at the 25th Midwinter Research Meeting, Association for Research in Otolaryngology. St. Petersburg Beach, FL (2002).

[b)]Electronic mail: plumx002@umn.edu

TABLE I. Subject information. Gender, age when tested for the present study, etiology of deafness, duration of bilateral severe-to-profound hearing loss prior to implantation, duration of implant use prior to the study, percent-correct score on NU-6 monosyllabic word test in quiet, and transmitted information (TI) for consonants and vowels in quiet. Speech scores were obtained at a presentation level of 60 dB SPL with subjects using their own speech processor and clinical speech processing strategy (CIS, PPS, or SAS).

SPRL Group								
Subject	**M/F**	**Age**	**Etiology of deafness**	**Dur. (years)**	**CI Use (yrs)**	**NU-6 (% C)**	**Consonants (TI)**	**Vowels (TI)**
C02	M	37	Unknown	19	4.6	88	0.801	0.933
C03	F	49	Progressive, familial	27	3.8	68	0.789	0.889
C05	M	42	Unknown, sudden	1	3.3	46	0.806	0.894
C06	M	62	Unknown	12	3.3	24	0.555	0.512
C07	F	58	Progressive, familial	35	2.7	34	0.722	0.823
C12	F	47	Otosclerosis	13	2.0	N/A	0.712	0.863
C13	M	78	Noise	6	1.7	50	0.548	0.420
MEAN		**53.3**		**16.1**	**3.1**	**51.7**	**0.705**	**0.762**
S.E.		**5.3**		**4.5**	**0.4**	**9.5**	**0.042**	**0.078**

HF+EPS Group								
Subject	**M/F**	**Age**	**Etiology of deafness**	**Dur. (years)**	**CI Use (yrs)**	**NU-6 (% C)**	**Consonants (TI)**	**Vowels (TI)**
C14	M	63	Unknown	47	1.6	76	0.730	0.901
C15	F	40	Unknown	7	1.5	68	0.755	0.817
C18	M	66	Otosclerosis	33	1.3	42	0.631	0.697
C19	M	62	Progressive, familial	32	0.6	36	0.489	0.584
C20	M	59	Progressive, familial	28	1.6	48	0.645	0.600
C21	F	56	Measles	3	1.8	64	0.654	0.833
C25	F	80	Otosclerosis	21	1.3	20	N/A	0.562
MEAN		**60.9**		**24.4**	**1.4**	**50.6**	**0.651**	**0.713**
S.E.		**4.5**		**5.8**	**0.1**	**7.5**	**0.038**	**0.052**

teners. Accordingly, the goal of this study was to determine whether these two factors have a systematic effect on Weber fractions for intensity discrimination ($Wf_{\mathrm{dB}} = 10 \log \Delta I/I$) or on the number of discriminable intensity steps across the electrical dynamic range.

II. METHODS

A. Subjects and implants

Subjects were 14 postlingually deafened adults with a Clarion C-I cochlear implant, divided into two groups according to the type of implanted electrode array. Seven subjects had a Clarion Spiral electrode array with no electrode positioner (SPRL group), and the other seven subjects had a Clarion HiFocus electrode array with an independent or attached electrode positioning system (HF+EPS group). The SPRL array is a traditional array that is designed to reside midway between the lateral and medial walls of scala tympani. It consists of eight pairs of radially offset ball electrodes with approximately 2 mm spacing between consecutive lateral or medial electrodes. In contrast, the HiFocus array with electrode positioning system (EPS) is designed to hug the modiolar wall (due to its slightly curved design and use of the EPS). It consists of eight pairs of longitudinally aligned plate electrodes with center-to-center distances of approximately 1 mm. Full insertion of the electrode array (25 mm) was achieved in all subjects in both subject groups with the exception of subject C07, whose insertion depth was estimated at the time of surgery to be 22 mm. Table I provides additional subject information. Note that average duration of implant use was significantly longer for subjects in the SPRL group than for subjects in the HF+EPS group (t-test, $p < 0.01$), but that average duration of deafness was similar for the two groups (t-test, $p = 0.28$). Average word and phoneme recognition in quiet was also similar for subjects in each group.

For the present study, stimulation was monopolar, with the active intracochlear electrode referenced to a ground electrode on the case of the internal receiver-stimulator. One electrode near the middle of the array (typically rEL7 or rEL8)[1] was tested in each subject. Subjects with the SPRL array were always tested on a medially positioned electrode.

B. Stimuli and procedures

Experiments were controlled by a personal computer (PC) running the Clarion SCLIN '98 FOR WINDOWS clinical programming software (Advanced Bionics Corporation, 1996–1997), and a second PC running custom software that controlled a research interface provided by Advanced Bionics Corporation for the Clarion C-I intracochlear stimulator. Stimuli were 200-ms pulse trains comprised of 77-μs/ph, cathodic-first, biphasic pulses with no interphase gap, presented in monopolar mode. Pulse rates were 200, 1625, and 6500 pps. Nominal current amplitudes specified in the clinical and research stimulation software were translated to calibrated amplitudes using a set of tables developed in our laboratory. These tables compensated for nonlinearities in the current source that were functions of electrode impedance and pulse rate.[2] Electrical impedances for all test electrodes

were measured at the beginning and end of each data collection session using the SCLIN '98 FOR WINDOWS software. Impedance measures typically varied about ±10% across sessions.[3] The average value of impedance measures obtained at all test sessions was used to compute the calibrated amplitudes for a given electrode.

Each pulse rate was tested in a single, two-hour data collection session or in two consecutive sessions, with the order of pulse rates randomized across subjects. In each test session, THS and MAL were first measured using the SCLIN '98 FOR WINDOWS software with an ascending method of adjustment procedure. For these measurements, the 200-ms pulse train stimuli were presented continuously, separated by 500-ms silent intervals. Starting at a level below THS, the tester slowly increased current amplitude until the subject indicated that a sound was barely audible (THS) and then continued to increase current amplitude until the subject indicated that the sound had reached a level that was just below an uncomfortable level of loudness (MAL). This procedure was repeated 2–3 times, and average values were computed for THS and MAL. Initial estimates of THS and final values of MAL were based on these values.

Additional measures of THS were obtained using a three-interval forced-choice (3IFC) adaptive procedure. The initial THS estimate obtained with the ascending method of adjustment procedure was used to set the starting current level of the adaptive track, which used a three-down, one-up final stepping rule to estimate the current level corresponding to 79.4% correct detection (Levitt, 1971). Stimulus intervals were cued visually on a video monitor and subjects used a computer mouse to select the interval containing a sound. Correct-answer feedback was given immediately after each response. Current amplitudes corresponding to the final eight reversals of the adaptive track were averaged to obtain a single THS estimate. THS estimates from 3–5 tracks were averaged to obtain a final value of THS.

Intensity difference limens (DLs) were obtained using a 3IFC adaptive procedure at five current levels spanning the electrical dynamic range (approximately 17, 35, 50, 65, and 83% of the dynamic range in dB) for each pulse rate. The procedure was similar to that used for THS measurements, described above. Stimuli were 200-ms pulse trains with no rise/fall envelopes. Two intervals contained fixed-amplitude stimuli and the third interval, selected at random on each trial, contained a higher-amplitude comparison stimulus. The subject was instructed to select the interval with the loudest sound. A one-up, one-down initial stepping rule was used for the first four reversals in order to quickly move the track to the region of threshold. A three-down, one-up final stepping rule was used for the subsequent eight reversals. Probe levels associated with these eight reversals were averaged to obtain the intensity discrimination threshold corresponding to 79.4% correct detection (Levitt, 1971). Step size was equal to two stimulus units (SUs)[4] for the initial four reversals and a single SU for the final eight reversals. Data were obtained in sets, where each set was comprised of one adaptive track at each stimulus intensity in increasing order. Four to six threshold estimates were averaged to obtain a single intensity discrimination threshold at each level.

Intensity discrimination thresholds were expressed as Weber fractions in dB [$Wf_{\mathrm{dB}}=10\log(\Delta I/I)$]. Since stimulus level was defined in units of current amplitude, the Weber fraction ($\Delta I/I$) was specified as $Wf=(\Delta A/A)^2+2\Delta A/A$, where A is the current amplitude of the reference stimulus and $A+\Delta A$ is the amplitude of the incremented stimulus. Current amplitude was measured in microamperes (μA). Wfs are commonly used to describe intensity discrimination in acoustic hearing and have been shown to be a valid metric of intensity discrimination in electric hearing (Nelson *et al.*, 1996). Mean Wf_{dB}s for each pulse rate were plotted as functions of stimulus level in dB and percent dynamic range ($\%\mathrm{DR}_{\mathrm{dB}}$).

Intensity discrimination thresholds were also expressed as a percentage of DR by normalizing intensity DLs relative to DR $\{\Delta I_{\%\mathrm{DR}}=\Delta I_{\mathrm{dB}}/\mathrm{DR}_{\mathrm{dB}}{}^{*}100$, where $\Delta I_{\mathrm{dB}}=10\log[2A\Delta A+(\Delta A)^2]$ and $\mathrm{DR}_{\mathrm{dB}}=\mathrm{MAL}_{\mathrm{dB}}-\mathrm{THS}_{\mathrm{dB}}\}$. Normalized DLs were plotted as a function of stimulus level in %DR. In addition, normalized DLs at each of five probe levels over DR were averaged to produce a single measure of intensity discrimination for each subject and pulse rate, expressed as a percentage of DR ($\Delta I_{\%\mathrm{DR\ avg}}$). The number of discriminable intensity steps over DR was quantified by dividing the DR by this value ($\mathrm{STEPS}=100\%/\Delta I_{\%\mathrm{DR\ avg}}$). This method for estimating the number of discriminable steps was chosen over a cumulative DL procedure because step-counts based on cumulative DLs may be strongly weighted by excessively large DLs near threshold or excessively small DLs near MAL (for example, see Nelson *et al.*, 1996).

III. RESULTS

A. THS, MAL, and DR

THS, MAL, and DR measures for each subject and pulse rate are summarized in Table II.[5] Mean values for each group are also shown. Absolute THSs were similar for subjects with the SPRL and HF+EPS electrode arrays, but MALs and DRs were significantly larger for subjects with the SPRL array than for subjects with the HF+EPS array. DR was an increasing function of pulse rate for both subject groups. As pulse rate increased from 200 to 6500 pps, mean DR increased from 10.7 to 19.4 dB for subjects with the SPRL array and from 7.5 to 12.5 dB for subjects with the HF+EPS array.

B. Weber fractions

Figure 1 shows Weber fractions as a function of stimulus level for the three pulse rates tested (200, 1625, and 6500 pps). Each panel shows the data for one subject. Data points at the resolution limits of the implant system are indicated by an **x**. Actual Weber fractions at these stimulus levels may be smaller than the values indicated, but could not be measured due to limitations in the minimum current step that could be delivered by the device. In general, the Weber fractions for a given individual and pulse rate decrease with increasing current amplitude, as reported in earlier studies. However, exceptions are common at 200 pps, where several subjects (e.g., C03, C20) exhibit Weber fractions that decrease with

TABLE II. THS, MAL, and DR by group (SPRL vs HF+EPS electrode array) and pulse rate. All current levels are expressed in dB re: 1 μA.

SPRL Group									
	THS			MAL			DR		
Subject	**200**	**1625**	**6500**	**200**	**1625**	**6500**	**200**	**1625**	**6500**
C02	38.6	29.6	26.0	50.4	49.6	48.5	11.7	20.0	22.5
C03	40.4	32.1	28.1	51.0	50.5	48.6	10.7	18.4	20.5
C05	37.8	29.6	27.0	49.3	49.1	47.6	11.5	19.5	20.6
C06	34.6	27.0	23.2	45.5	44.7	41.8	11.0	17.7	18.6
C07	35.7	27.3	24.6	44.0	42.1	39.2	8.3	14.8	14.6
C12	44.0	35.5	31.0	54.2	53.6	49.2	10.2	18.1	18.1
C13	39.5	32.3	28.2	51.1	51.0	49.3	11.6	18.7	21.1
MEAN	**38.7**	**30.5**	**26.9**	**49.4**	**48.7**	**46.3**	**10.7**	**18.2**	**19.4**
S.E.	**1.18**	**1.14**	**0.97**	**1.33**	**1.49**	**1.54**	**0.45**	**0.64**	**0.98**

HF+EPS Group									
	THS			MAL			DR		
Subject	**200**	**1625**	**6500**	**200**	**1625**	**6500**	**200**	**1625**	**6500**
C14	37.8	29.7	27.0	45.2	42.2	40.3	7.4	12.5	13.3
C15	33.7	26.1	22.5	38.4	30.8	28.8	4.7	4.7	6.3
C18	44.1	35.0	31.6	51.8	47.0	42.6	7.6	11.9	11.0
C19	41.9	34.8	31.1	46.2	43.3	41.1	4.2	8.5	10.1
C20	34.7	27.9	25.6	44.2	41.6	39.5	9.5	13.7	13.9
C21	39.4	31.2	28.1	50.1	49.1	49.0	10.7	17.9	20.9
C25	35.2	28.3	26.2	43.4	40.3	38.3	8.2	12.1	12.1
MEAN	**38.1**	**30.4**	**27.4**	**45.6**	**42.0**	**40.0**	**7.5**	**11.6**	**12.5**
S.E.	**1.48**	**1.30**	**1.20**	**1.68**	**2.21**	**2.28**	**0.89**	**1.56**	**1.69**

increasing stimulus level at low-to-moderate levels, but then increase with level at the highest stimulus intensities. These upturns at high levels may reflect saturation of the outputs of auditory nerve fibers or saturation in recruitment of new fibers with increasing current amplitude. It is unlikely that they reflect voltage compliance limitations of the device, since stimulus levels fell within the linear, nonsaturating portion of calibration input-output functions. We have observed similar upturns in a few subjects with the Nucleus-22 implant (Nelson *et al.*, 1996; Donaldson and Nelson, 1997).

Weber fractions for subjects with the SPRL electrode array range from approximately -2 dB at low levels to -15 dB at high levels. These ranges are fairly consistent across subjects. Weber fractions for subjects with the HF+EPS electrode array are less consistent in their ranges: Three subjects with smaller DRs and lower THSs and MALs (C14, C15, and C20) show a restricted range of relatively large Weber fractions (-1 to -9 dB), whereas four subjects with larger DRs and higher THSs and MALs (C18, C19, C21, and C25) show a wider range of Weber fractions (-2 to -14 dB) similar to that demonstrated by subjects with the SPRL array. This trend suggests that Weber fractions are at least partly determined by overall current amplitude, becoming smaller at high amplitudes.

For most subjects in both groups, Weber fractions for 1625 and 6500 pps stimuli are similar to one another, but Weber fractions for the 200 pps stimulus are slightly or considerably smaller. A few subjects (C07, C12 C18, C19, and C25), show a very sharp decrease in Weber fractions for the 200 pps stimulus at high levels. This reflects the trend for Weber fractions to become smaller at high current amplitudes, noted above, since THSs and MALs occur at substantially higher current amplitudes for the 200 pps stimulus than for 1625 and 6500 pps stimuli in these subjects.

In order to compare the Weber fractions for different pulse rates at equivalent levels within the DR, Fig. 2(a) replots the data from Fig. 1 with the stimulus level expressed as a percentage of DR (%DR_{dB}) for each pulse rate. For most subjects, Weber fractions are similar for the 1625 and 6500 pps stimuli at equivalent levels within the DR, but are smaller for the 200 pps stimulus at one or more levels. In three subjects (C06, C07, and C15) there is no clear difference between the size of Weber fractions at 200 pps and at the higher pulse rates. The average data shown in Fig. 2(a) (bottom right panel for each group) and Fig. 2(b) (which directly compares the mean data for each group at each pulse rate) confirm the trend for Weber fractions to be more sensitive for 200 pps stimuli than for stimuli with higher pulse rates. The group mean data also indicate that Weber fractions at moderate and high levels within the DR are more sensitive for subjects with the SPRL array than for subjects with the HF+EPS array.

A two-way (group$\times$pulse rate, repeated measures within subjects) analysis of variance (ANOVA) confirmed a significant main effect of pulse rate on the size of the average Weber fraction (collapsed across level) [$F(2,24)=30.68$, $p<0.001$]. As expected, post-hoc testing indicated that Weber fractions for the 200 pps stimulus were significantly smaller than Weber fractions for the 1625 and 6500 pps stimuli, but that Weber fractions for 1625 and 6500 pps stimuli were not

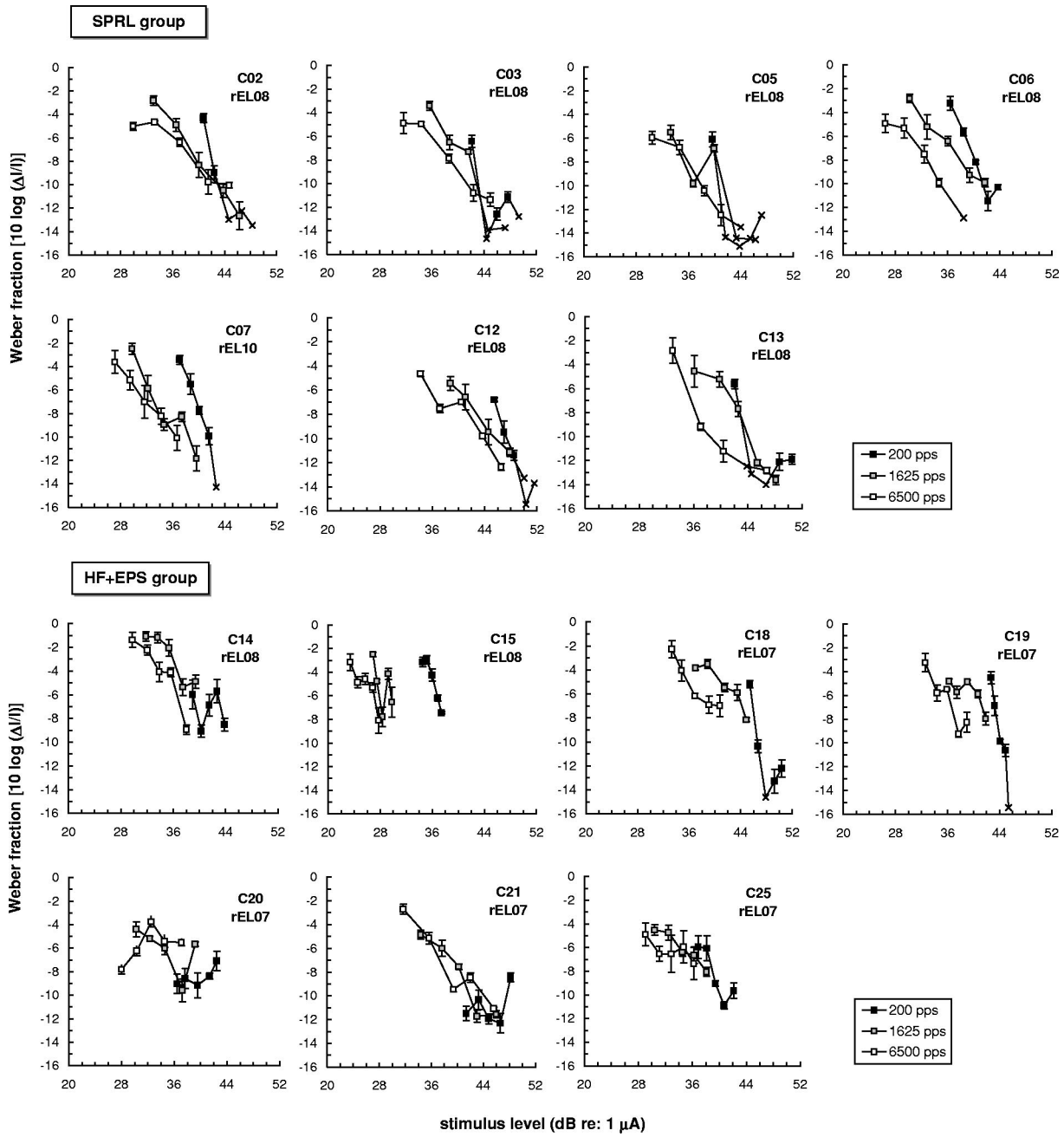


FIG. 1. Weber fractions [10 log ($\Delta I/I$)] as a function of stimulus level (dB re: 1 μA) for each subject at each of three pulse rates. Data points limited by device resolution are indicated by an **x**. Error bars represent 1 s.e. of the mean.

significantly different from one another (Tukey test, $p < 0.001$). The two-way (group$\times$pulse rate, repeated measure within subjects) ANOVA also confirmed a significant main effect of group [$F(1,24) = 10.97, p < 0.01$]. Weber fractions (collapsed across pulse rate and level) had a mean value of -9.1 dB [standard error(s.e.) $= 0.4$] for subjects with the SPRL electrode array and -6.7 dB (s.e. $= 0.5$) for subjects with the HF+EPS array. Smaller Weber fractions for subjects in the SPRL group may be attributable to the fact that their THSs and MALs extend to higher current amplitudes than those for HF+EPS subjects (see Table II). There was no interaction between group and pulse rate.

To better evaluate the relationship between current amplitude and the size of Weber fractions, all of the individual data points from Fig. 1 are replotted in Fig. 3 as a function of current amplitude. Points at the resolution limits of the device are shown as filled symbols. The linear regression line indicates that there is a strong inverse relationship between absolute current amplitude and the size of Weber fractions and that current amplitude accounts for more than half of the variability in Weber fraction magnitudes (slope $= -0.43$; $R^2 = 0.56$; $p < 0.001$). This regression line was based on all points shown in the figure, including those at the resolution limits of the device. However, the linear regression remains significant (slope $= -0.35$; $R^2 = 0.484$; $p < 0.001$) even when such data points are excluded.

Linear regression lines were also fit to the data in Fig. 3 separately by group collapsed across pulse rate [SPRL

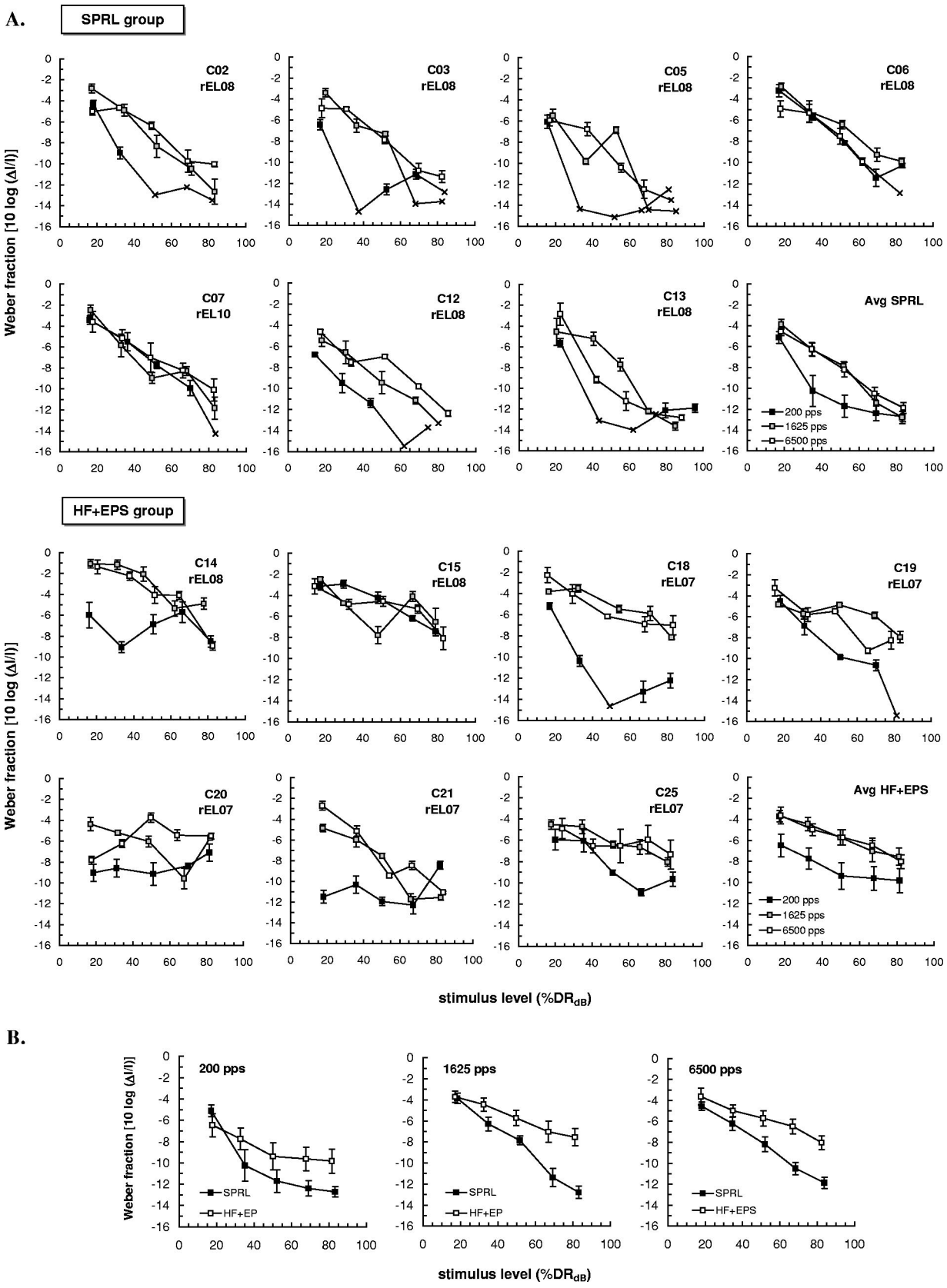


FIG. 2. (a) Weber fractions [10 log ($\Delta I/I$)] as a function of stimulus level in percent dynamic range [$\%DR_{dB}$] for each subject at each of three pulse rates. Data points limited by device resolution are indicated by an **x**. Error bars represent 1 s.e. of the mean. The average data are shown in the bottom right panel for each group. (b) Weber fractions [10 log ($\Delta I/I$)] as a function of stimulus level in percent dynamic range ($\%DR_{dB}$) for average group data compared as a function of pulse rate. Error bars represent 1 s.e. of the mean.

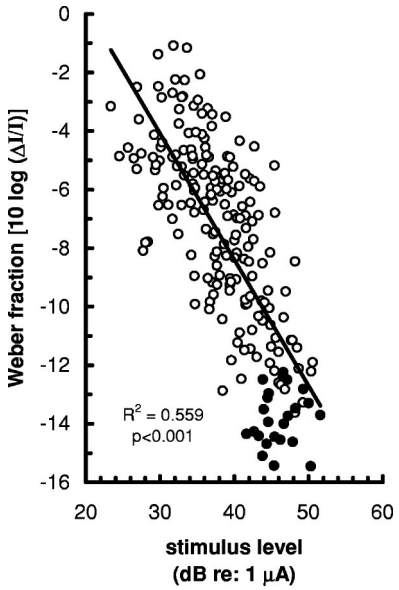


FIG. 3. Scatterplot of Weber fractions [$10 \log (\Delta I/I)$] versus current amplitude (dB re: 1 μA) for all subjects and pulse rates. Filled symbols represent points limited by device resolution. The regression coefficient was computed using all data points.

(slope $=-0.48$, $R^2=0.61$); HF+EPS (slope $=-0.34$, $R^2=0.45$)] and by pulse rate collapsed across group [200 pps (slope $=-0.59$, $R^2=0.50$); 1625 pps (slope $=-0.46$, $R^2=0.56$); 6500 pps (slope $=-0.40$, $R^2=0.56$)]. The slopes of the regression lines obtained did not differ significantly either by group ($F=0.85$, $p<0.001$) or by pulse rate ($F=0.97$, $p<0.001$).

C. Normalized intensity DLs

Normalized intensity DLs that express intensity discrimination thresholds as a percentage of DR ($\Delta I_{\%\mathrm{DR}} = \Delta I_{\mathrm{dB}}/\mathrm{DR}_{\mathrm{dB}}*100$) are shown in Fig. 4 for each subject and pulse rate. Note that graph ordinates have a wider range for HF+EPS subjects than for SPRL subjects. To facilitate visual comparison of the individual data across groups, a dashed line is plotted in each graph at a level of 5%DR. The range of normalized DLs is considerably smaller for SPRL subjects (1–20%DR) than for HF+EPS subjects (2–42%DR). Five out of seven subjects in the SPRL group have normalized intensity DLs that fall primarily below the dashed line at 5%, whereas six of seven subjects in the HF+EPS group have normalized DLs that fall primarily above this line.

For most subjects, normalized DLs are fairly similar in magnitude for each of the three pulse rates. However, four subjects show smaller intensity DLs at some pulse rates than others at some stimulus levels. Three of these subjects (C06, C07, and C19) show a clear advantage of 1625 and 6500 pps stimuli over the 200 pps stimulus at low levels within the DR. This is the result of DRs being smaller for the 200 pps stimulus. One additional subject (C18) shows a clear advantage of 200 pps stimuli over the two higher rate stimuli at moderate and higher stimulus levels.

D. Discriminable intensity steps

Figure 5 plots the average number of discriminable intensity steps over DR estimated for SPRL and HF+EPS subjects at each pulse rate. These values were obtained by computing the ratio of DR to mean normalized intensity DL for each subject and condition, and then computing average values for each group. The number of discriminable intensity steps was similar across pulse rate within each group. For the SPRL group, step counts ranged from 20 to 25 steps, averaging 23 steps. For the HF+EPS group, step counts were much smaller, ranging from 8.5 to 9.5 steps and averaging 9 steps. A two-way (group$\times$pulse rate, repeated measures within subjects) ANOVA confirmed that differences in step counts between SPRL and HF+EPS subjects were statistically significant at all pulse rates [$F(1,24)=11.65$, $p<0.01$]. The observed difference in average step counts between subjects with the SPRL electrode array and subjects with the HF+EPS array is substantial and may have implications for device benefit (see below).

IV. DISCUSSION

The primary goal of this study was to evaluate the effects of pulse rate and electrode array design on intensity discrimination in Clarion cochlear implant listeners. Intensity discrimination was measured for stimuli with slow (200 pps), intermediate (1625 pps), and fast (6500 pps) pulse rates in subjects with each of two types of electrode arrays: a standard array (SPRL) that resides in the middle to midlateral region of scala tympani, and a perimodiolar array (HF+EPS) that resides close to the modiolar wall. Intensity discrimination thresholds were considered both in terms of Weber fractions and the number of discriminable steps over the dynamic range of hearing.

A. Effects of pulse rate

Pulse rate was found to have a significant effect on intensity discrimination, with the 200 pps stimulus producing smaller Weber fractions than the 1625 and 6500 pps stimuli. This effect appears to stem from an inverse relationship between the magnitude of Weber fractions and absolute current amplitude: Figure 3 showed that approximately 56% of variability in Weber fractions can be accounted for by current level. Because THSs and MALs for the 200 pps stimuli occurred at higher current amplitudes than those for the 1625 and 6500 pps stimuli, average Weber fractions across the dynamic range of hearing were smaller at 200 pps than at the higher pulse rates. The inverse relationship between Weber fractions and current level may reflect the same physiologic mechanism that underlies increases in the slopes of derived loudness growth functions at high stimulus levels (McKay *et al.*, 2003). A likely candidate for this mechanism is the rapid recruitment of auditory nerve fibers in the cochlear modiolus at high current amplitudes (Nelson *et al.*, 1996; Frijns *et al.*, 2001; McKay *et al.*, 2003). Reduced variance of auditory nerve fiber responses for 200 pps stimuli may also contribute to smaller Weber fractions at high current amplitudes (see below; Javel and Viemeister, 2000).

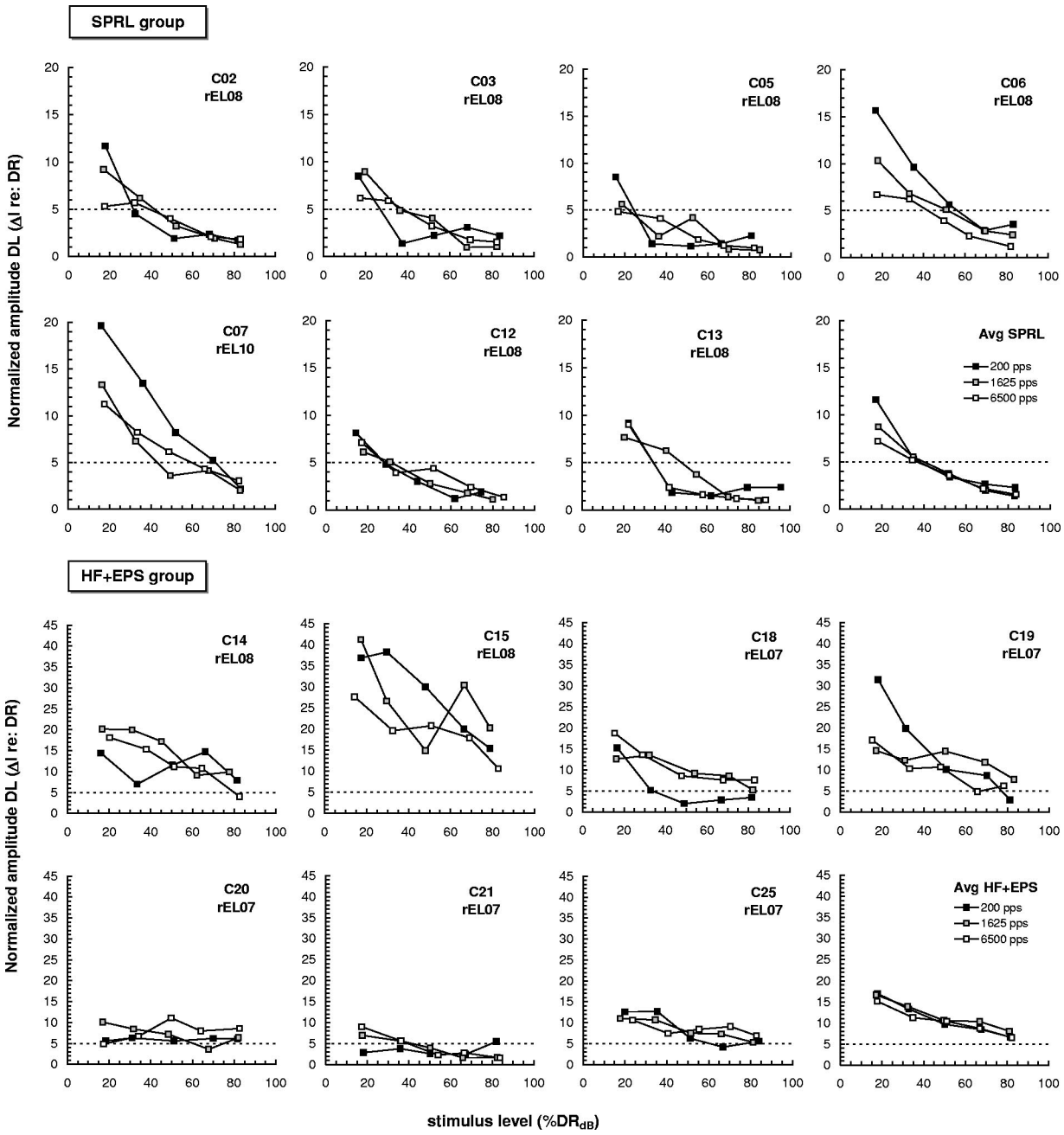


FIG. 4. Normalized amplitude DLs (ΔI re: DR) for each subject and pulse rate. Note that an expanded ordinate is used for HF+EPS subjects. Dashed lines at 5% DR are shown for reference.

Pulse rate had no systematic effect on the number of discriminable intensity steps over the dynamic range of hearing. The apparent advantage of 200 pps stimuli in producing smaller Weber fractions was offset by the smaller electrical dynamic range at this pulse rate. Thus, changes in stimulus pulse rate over the range evaluated here (200 to 6500 pps) do not have a systematic effect on the number of intensity steps that can be coded within the dynamic range. A few subjects showed advantages of a particular pulse rate (Fig. 3); however, most individual subjects and the mean data showed no benefit of one pulse rate over another (Figs. 3 and 5).

B. Effects of electrode array design

Electrode array design was also found to have a significant effect on the size of Weber fractions, with average Weber fractions over the dynamic range being larger for subjects with the HF+EPS electrode array than for subjects with the SPRL array. This finding is consistent with the inverse relationship observed between current amplitude and Weber fractions, since MALs are considerably lower for subjects with the HF+EPS array than for subjects with the SPRL array at all pulse rates.

Perhaps the most striking finding of this study was that the combined effects of smaller dynamic range and larger Weber fractions resulted in HF+EPS subjects having a very limited number of discriminable intensity steps over the dynamic range of hearing. Subjects with this electrode array had an average of only 9 discriminable intensity steps as compared to 23 discriminable steps in subjects with the SPRL electrode array.

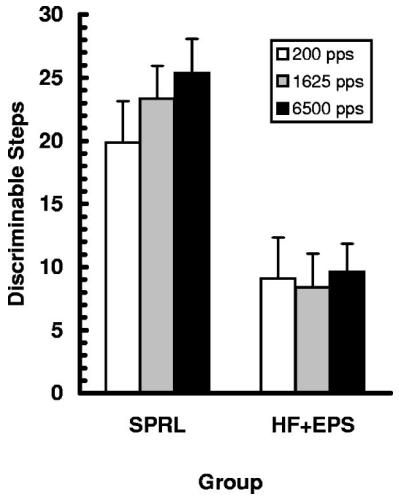


FIG. 5. Mean number of discriminable steps by group and pulse rate. Error bars represent 1 s.e. of the mean.

An interesting aspect of this result is the inverse relationship between Weber fractions and dynamic range. From an intuitive perspective, it might be expected that subjects having smaller dynamic ranges would exhibit proportionally smaller Weber fractions than those having larger dynamic ranges. This stems from the notion that loudness grows more rapidly with level in the case of a smaller dynamic range and the assumption that Weber fractions are inversely proportional to the rates of loudness growth. In the present study, however, subjects with smaller dynamic ranges had *larger* Weber fractions. One possibility suggested by this finding is that the variability of neural responses to identical pulse trains is greater for subjects with the HF+EPS electrode array. Response variability or ''internal noise'' is thought to limit discrimination abilities in all sensory systems and is incorporated in models of intensity discrimination (McGill and Goldberg, 1968; Penner, 1972; Teich and Lachs, 1979; Carlyon and Moore, 1984; Viemeister, 1988). Substantially increased response variability with the HF+EPS electrode array could account for large Weber fractions even in the presence of more rapid response growth. It is not clear what factors could underlie increased response variability in subjects with the HF+EPS array. Theoretically, variability will be higher if neural responses are elicited from smaller diameter neural targets (Verveen, 1962); for example, peripheral processes of auditory nerve fibers as compared to central processes or spiral ganglion cell bodies. However, there is little data to suggest a relationship between electrode position within scala tympani (standard versus perimodiolar electrode) and neural response site. Recent modeling data of Frijns *et al.* (2001) indicates that there is no clear difference in spike initiation site as a function of current level for different electrode positions in an idealized human cochlea.

The reduced number of intensity steps achieved by subjects with the HF+EPS electrode array does not appear to adversely affect speech recognition in quiet: Average word and phoneme recognition scores were similar for subjects in the HF+EPS and SPRL groups (Table I), even though subjects with the SPRL electrode array had many more intensity steps. This finding is consistent with previous studies indicating that a high degree of intensity resolution is not needed for speech recognition in quiet (Loizou *et al.*, 2000a). Poor intensity resolution among subjects with the HF+EPS electrode array may have a greater impact on the ability of these subjects to take advantage of fine level cues needed for complex listening tasks and sound source localization.

C. Monopolar versus bipolar stimulation

Weber fractions reported in the present study vary from approximately -5 to -15 dB across subjects and stimulus conditions, covering a similar range of values as Weber fractions described previously for bipolar stimulation (e.g., Nelson *et al.*, 1996; Cohen *et al.*, 2001). Previous studies of monopolar versus bipolar stimulation within the same subjects have had mixed results. Shannon (1983) found that intensity DLs were similar for monopolar and bipolar stimulation in one subject at pulse rates of 100 and 1000 pps. Preliminary data by Drennan *et al.* (2003) showed no consistent effect of electrode configuration (bipolar versus monopolar) on intensity DLs for 250 pps stimuli in ten Nucleus 24 Contour users, but found that some individual subjects demonstrated better intensity resolution with a particular configuration. In contrast, Morris and Pfingst (2000) found that monopolar stimulation produced better intensity discrimination than bipolar stimulation in three of four implanted monkeys. These existing data suggest that electrode configuration can have a significant effect on intensity resolution in some individual cochlear implant users, but that there is no systematic effect across subjects. As yet there are no monopolar-bipolar comparisons of intensity discrimination for stimulus rates above 1000 pps. Thus, future studies should evaluate the effects of electrode configuration on intensity discrimination for the higher-rate stimuli now in common clinical use.

V. SUMMARY

(1) Absolute intensity DLs (Weber fractions) were significantly smaller for 200 pps stimuli than for 1625 or 6500 pps stimuli.
(2) The number of discriminable intensity steps over the DR of electrical hearing was similar for 200, 1625, and 6500 pps stimuli. The advantage of 200 pps stimuli in producing smaller Weber fractions was offset by smaller DRs at this pulse rate.
(3) Average intensity DLs were smaller for subjects with the standard, SPRL electrode array ($Wf_{\mathrm{dB}} = -9.1$ dB) than for subjects with the perimodiolar HF+EPS electrode array ($Wf_{\mathrm{dB}} = -6.7$ dB).
(4) Subjects with the perimodiolar HF+EPS electrode array had significantly fewer discriminable intensity steps over DR than subjects with the standard SPRL electrode array. On average, subjects with the SPRL array could discriminate 23 intensity steps whereas subjects with the HF+EPS array could discriminate only 9 steps.
(5) There is a strong, inverse relationship between absolute current amplitude and the magnitude of Weber fractions. This relationship may largely explain the effects of pulse rate and electrode array design on intensity DLs in the present study. Increasing pulse rate lowers THS and

MAL, so that faster pulse rates are associated with larger Weber fractions than slower pulse rates. Similarly, perimodiolar electrode arrays are associated with lower MALs than standard electrode arrays, and are associated with larger intensity DLs.

(6) Smaller dynamic ranges, which imply steeper loudness growth functions, are not associated with proportionally smaller Weber fractions in subjects with the HF+EPS electrode array. This may be explained by larger response variance in these subjects than in subjects with the SPRL electrode array.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant No. DC00110 and the Lions 5M International Hearing Foundation. Preliminary findings were presented at the 25th Midwinter Research Meeting, Association for Research in Otolaryngology, St. Petersburg Beach, FL (2002). The authors thank Advanced Bionics Corporation for providing the C-I research interface, Eric Javel for software development, Suzanne Hansel and Shanna Allen for assistance in data collection, and Sharon Smith for clinical support. The authors wish to extend special thanks to the subjects who participated in this study. Neal Viemeister and two anonymous reviewers provided helpful comments on an earlier version of this manuscript.

[1]Our research numbering system orders electrodes sequentially from low-to-high along the apical-to-basal extent of the array. Research electrode numbers are identical to clinical electrode numbers for the HiFocus electrode. For the SPRL electrode, lateral clinical electrodes are converted to research electrodes by using the formula {(2*clinical electrode #) −1} and medial clinical electrodes are converted to research electrodes using the formula {2*clinical electrode #.} For example, clinical electrode 1L corresponds to research electrode 1 and clinical electrode 8M corresponds to research electrode 16.

[2]Voltage measurements were recorded from a digital oscilloscope for amplitude steps of approximately 50 CU (nominal μA) over the range of amplitudes evaluated in this study (50–1000 CU). Measurements were made for resistive loads of 1.1, 3.6, 7.5, 10, 12, 15, 18, 22, 27, 33, and 47 kohms and for pulse rates of 50, 100, 200, 325, 500, 1000, 1625, 3250 and 6500 pps (all possible combinations). Three different intracochlear stimulators were evaluated, and were found to produce almost identical outputs.

[3]Impedances were measured using a 1 kHz sinusoid at approximately 10 μA. Although impedance varies with current density, it is expected to be relatively constant for the range of current amplitudes evaluated in the present study (Geddes and Baker, 1989). Our calibration procedure does not account for possible changes in impedance with stimulus waveform. Because impedance for intracochlear electrodes is capacitive, it would be expected to decrease with pulse rate for the stimuli used here, due to the higher-frequency spectral content of higher-pulse rate stimuli. This could lead to an overestimation of absolute THS and MAL (actual THS and MAL would be higher than values reported) for high pulse-rate stimuli, however, this effect should be small (Advanced Bionics Corporation, personal communication). Variations in electrode impedance over time or as a function of pulse rate have essentially no effect on Weber fractions over the current amplitudes encompassed by the present data, which are well within the linear portion of input-output functions for the Clarion CI-I device.

[4]Stimulus units (SUs) are current amplitude units which vary in 0.3 dB re: 1 μA increments. Perfect performance for intensity discrimination with the Clarion C-I device represents a 0.5 SU change in current amplitude, corresponding to a Weber fraction of approximately −15 dB.

[5]The results summarized in Table II are not the same values obtained in Kreft *et al.* (2004). These values were obtained from additional measurements following that experiment (1 month to 1 year later). Because both studies used nearly the same subjects, the present THS, MAL, and DR results do not provide new data regarding the effects of pulse rate and electrode array design on these measures.

Advancd Bionics Corporation (**1996–1997**). *SCLIN '98 for Windows [cochlear implant programming software*], Sylmar, CA.

Carlyon, R. P., and Moore, B. C. J. (**1984**). "Intensity discrimination: A severe departure from Weber's law," J. Acoust. Soc. Am. **76**, 1369–1376.

Cohen, L. T., Saunders, E., and Clark, G. M. (**2001**). "Psychophysics of a prototype peri-modiolar cochlear implant electrode array," Hear. Res. **155**, 63–81.

Donaldson, G. S., Peters, M. D., Ellis, M. R., Friedman, B. J., Levine, S. C., and Rimell, F. L. (**2001**). "Effects of the Clarion Electrode Positioning System on auditory thresholds and comfortable loudness levels in pediatric patients with cochlear implants," Arch. Otolaryngol. Head Neck Surg. **127**, 956–960.

Donaldson, G. S., and Nelson, D. A. (**1997**). "Intensity resolution for pulse amplitude and pulse duration: Individual differences and stimulus-parameter effects," Abstracts of the 1997 Conference on Implantable Auditory Prostheses, Pacific Grove, CA.

Drennan, W. R., Pfingst, B. E., and Xu, L. (**2003**). "Intensity discrimination using bipolar and monopolar electrode configurations in Nucleus Contour cochlear implants," Abstracts of the 26th Midwinter Research Meeting, Association for Research in Otolaryngology. Daytona Beach, FL.

Frijns, J. H., Briaire, J. J., and Grote, J. J., (**2001**). "The importance of human cochlear anatomy for the results of modiolus-hugging multichannel cochlear implants," Otology and Neurotology **22**, 340–349.

Geddes, L. A., and Bake, L. E. (**1989**). *Principles of Applied Biomedical Instrumentation, 3rd ed.* (Wiley, New York), pp. 332–337.

Javel, E., and Viemeister, N. F. (**2000**). "Stochastic properties of cat auditory nerve responses to electric and acoustic stimuli and application of intensity discrimination," J. Acoust. Soc. Am. **107**, 908–921.

Kreft, H. A., Donaldson, G. S., and Nelson, D. A. (**2004**). "Effects of pulse rate on threshold and dynamic range in Clarion cochlear implant users," J. Acoust. Soc. Am. **115**, 1885–1888.

Lenarz, T., Tasche, C., Frohne-Buechner, C., Buechner, A., Illg, A., Strauss-Schier, A. and Battmer, R.-D. (**2001**). "Comparison of different Clarion and Nucleus electrode systems within matched-pairs groups," Abstracts of the 2001 Conference on Implantable Auditory Prostheses, Pacific Grove, CA, p. 116.

Lesinski-Schiedat, A., Goldring, J. E., Gupta, S., Battmer, R.-D., Frohme, C., Weber, B. P. and Lenarz, T. (**2000**). "Evaluation of Clarion Electrode Positioning System with the preformed and Hifocus electrode," Abstracts of the C12000: The 6th International Cochlear Implant Conference, Miami Beach, FL.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Loizou, P. C., Dorman, M., and Fitzke, J. (**2000a**). "The effects of reduced dynamic range on speech understanding: Implications for patients with cochlear implants," Ear Hear. **21**, 25–31.

Loizou, P. C., Dorman, M., Poroy, O., and Spahr, T. (**2000b**). "Speech recognition by normal-hearing and cochlear implant listeners as a function of intensity resolution," J. Acoust. Soc. Am. **108**, 2377–2387.

McGill, W. J., and Goldberg, J. P. (**1968**). "Pure-tone intensity discrimination and energy detection," J. Acoust. Soc. Am. **44**, 576–581.

McKay, C. M., Henshall, K. R., Farrell, R. J., and McDermott, H. J. (**2003**). "A practical method of predicting the loudness of complex electrical stimuli," J. Acoust. Soc. Am. **113**, 2054–2063.

Morris, D. J., and Pfingst, B. E. (**2000**). "Effects of electrode configuration and stimulus level on rate and level discrimination with cochlear implants," J. Assoc. Res. Otolaryngol. **1**, 211–223.

Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (**1996**). "Intensity discrimination as a function of stimulus level with electric stimulation," J. Acoust. Soc. Am. **100**, 2393–2414.

Osberger, M. J., Koch, D., Fisher, L., and Zimmerman-Phillips, S. (**1999**). "Clinical results in patients implanted with the Clarion electrode positioning system," Abstracts of the 1999 Conference on Implantable Auditory Prostheses, Pacific Grove, CA.

Parkinson, A. J., Arcaroli, J., Staller, S. J., Arndt, P. L., Cosgriff, A., and Ebinger, K. (**2002**). "The Nucleus 24 Contour cochlear implant system: adult clinical trial results," Ear Hear. **23**, 41S–48S.

Penner, M. J. (**1972**). "Neural or energy summation in a Poisson counting model," Percept. Psychophys. **15**, 566–570.

Saunders, E., Cohen, L., Aschendorff, A., Shapiro, W., Knight, M., Stecker, M., Richter, B., Waltzman, S., Tykocinski, M., Roland, T., Laszig, R., and

Cowan, R. (**2002**). ''Threshold, comfortable level and impedance changes as a function of electrode-modiolar distance,'' Ear Hear. **23**, 28S–40S.
Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.
Shannon, R. V. (**1985**). ''Threshold and loudness functions for pulsatile stimulation of cochlear implants,'' Hear. Res. **18**, 135–143.
Shannon, R. V. (**1989**). ''A model of threshold for pulsatile electrical stimulation of cochlear implants,'' Hear. Res. **40**, 197–204.
Simmons, F. B. (**1966**). ''Electrical stimulation of the auditory nerve in man,'' Arch. Otolaryngol. **84**, 24–76.
Skinner, M. W., Holden, L. K., Holden, T. A., and Demorest, M. E. (**2000**). ''Effect of stimulation rate on cochlear implant recipients' thresholds and maximum acceptable loudness levels,'' J. Am. Acad. Audiol **11**, 203–213.
Teich, M. C., and Lachs, G. (**1979**). ''A neural-counting model incorporating refractoriness and spread of excitation. I. Application to intensity discrimination,'' J. Acoust. Soc. Am. **66**, 1738–1749.
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.
Verveen, A. A. (**1962**). ''Axon diameter and fluctuation in excitability,'' Acta Morphol. Neerl Scand. **5**, 79–85.
Viemeister, N. F. (**1988**). ''Psychophysical aspects of auditory intensity coding,'' in *Auditory function: Neurobiological bases of hearing*, edited by G. M. Edelman, W. E. Gall and W. M. Cowan (Wiley, New York), pp. 213–241.
Xu, L., and Pfingst, B. E. (**2002**). ''Variation in thresholds and comfort levels across cochlear implant stimulation sites: effects of electrode configuration and stimulus level,'' Abstracts of the 25th Midwinter Research Meeting, Association for Research in Otolaryngology, St. Petersburg Beach, FL, p. 122.
Young, N. M., and Grohne, K. M. (**2001**). ''Comparison of pediatric Clarion recipients with and without the electrode positioner,'' Otol. Neurotol. **22**, 195–199.

Frequency modulation detection in cochlear implant subjects[a)]

Hongbin Chen[b)] and Fan-Gang Zeng[c)]

Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences and Otolaryngology—Head and Neck Surgery, University of California, Irvine

(Received 16 January 2004; revised 30 June 2004; accepted 1 July 2004)

Frequency modulation (FM) detection was investigated in acoustic and electric hearing to characterize cochlear-implant subjects' ability to detect dynamic frequency changes and to assess the relative contributions of temporal and spectral cues to frequency processing. Difference limens were measured for frequency upward sweeps, downward sweeps, and sinusoidal FM as a function of standard frequency and modulation rate. In electric hearing, factors including electrode position and stimulation level were also studied. Electric hearing data showed that the difference limen increased monotonically as a function of standard frequency regardless of the modulation type, the modulation rate, the electrode position, and the stimulation level. In contrast, acoustic hearing data showed that the difference limen was nearly a constant as a function of standard frequency. This difference was interpreted to mean that temporal cues are used only at low standard frequencies and at low modulation rates. At higher standard frequencies and modulation rates, the reliance on the place cue is increased, accounting for the better performance in acoustic hearing than for electric hearing with single-electrode stimulation. The present data suggest a speech processing strategy that encodes slow frequency changes using lower stimulation rates than those typically employed by contemporary cochlear-implant speech processors. [DOI: 10.1121/1.1785833]

PACS numbers: 43.66.Hg, 43.66.Fe, 43.66.Ts [NFV] Pages: 2269–2277

I. INTRODUCTION

Similar to the Fourier transform in which a signal can be decomposed into a series of sinusoids, a signal can also be decomposed into a combination of amplitude and frequency modulation components (Loughin and Tacer, 1996). Frequency modulation (FM) dynamically changes a signal's instantaneous frequency without necessarily affecting the instantaneous amplitude. The FM direction and rate, two important features that influence speech perception, are usually determined by two basic components: the carrier and the modulator. In speech and music sounds, frequency modulations in the form of formant transitions, fundamental frequency changes, and fine structure changes carry critical information for speech recognition, speaker identification, and music appreciation.

FM detection has been systematically studied in normal-hearing listeners for frequency sweeps (Sergeant and Harris, 1962; Pollack, 1968; Nabelek and Hirsh, 1969; Tsumura *et al.*, 1973; Arlinger *et al.*, 1977; Tyler *et al.*, 1983; Schouten, 1985; Dooley and Moore, 1988; Madden and Fire, 1996) and sinusoidal FM (Hartmann and Hnath, 1982; Demany and Semal, 1989; Moore and Glasberg, 1989; Edwards and Viemeister, 1994a,b; Sek and Moore, 1995; Moore and Sek, 1996). Generally, difference limens for frequency sweeps are positively proportional to the standard frequency and are relatively independent of the stimulation level. Dooley and Moore (1988) showed that the thresholds of upward and downward sweeps changed from approximately 4 Hz at standard frequency of 500 Hz to 7 Hz at a standard frequency of 1000 Hz. Previous studies were not consistent in reporting whether there is an asymmetry in detection of the downward and upward frequency glides. Schouten (1985) found that discrimination of upward sweeps required shorter durations and slower sweep rates than that of falling sweeps. Collins and Cullen (1978) reported an asymmetry in the detectability of upward and downward glides of short duration, with upward sweeps detected at lower signal intensities in the frequency ranges 200–700 Hz and 1200–1700 Hz. On the other hand, Tsumura *et al.* (1973) and Arlinger *et al.* (1977) found no significant difference between upward and downward sweeps.

Demany and Semal (1989) measured detection of sinusoidal frequency modulation as a function of carrier frequency from 250 to 4000 Hz and as a function of modulation frequency from 1 to 64 Hz. They found relatively independent effects of carrier frequency and modulation frequency on the detection threshold when the data were expressed in absolute Hz and plotted on a log scale; the difference limen increased monotonically with the carrier frequency, but remained relatively constant for different modulation frequencies. Other data suggested that the average difference limen was consistently larger when the amplitude of the frequency-modulated stimuli was either sinusoidally or randomly modulated (Grant, 1987).

Two theories have been put forward to explain the frequency coding mechanism. One theory relies on the ''place code'' in the cochlea (Zwicker, 1956; Henning, 1967) while

[a)]Portions of this work were presented at the 26th Midwinter Meeting of the Association for Research in Otolaryngology, Daytona Beach, FL, 2003.

[b)]Address for correspondence: 364 Med Surge II, University of California, Irvine, CA 92697-1275. Electronic mail: hchen@uci.edu

[c)]Electronic mail: fzeng@uci.edu

TABLE I. Subject information of three Nucleus-22 cochlear-implant users who participated in this study.

Subject	Gender	Age (years)	Cause of deafness	Duration of implant use (years)	Vowel recognition	Consonant recognition
S1	F	69	Cochlear otosclerosis	13	69%	69%
S2	M	59	Hereditary	6	59%	72%
S3	M	45	Trauma	10	71%	79%

the other uses the ''temporal code'' or ''phase locking code'' in the auditory nerve (Siebert, 1970; Goldstein and Srulovicz, 1977). In the place theory, the cochlea is modeled as a bank of filters and the cochlear filters' response is called an excitation pattern. When the stimulus frequency is changed, the filters' output is changed systematically dependent upon the filters' bandwidth and slope. Based on this theory, a common mechanism is used for processing both frequency discrimination and frequency modulation. However, Moore and Glasberg (1989) compared the difference limen between pulsed sinusoids and frequency modulations. The data showed that difference limens of pulsed sinusoids (DLFs) varied more with frequency than did the difference limens of frequency modulation (FMDLs). In addition, DLFs were less affected by the random variation in level and by the addition of bandpass noise than the FMDLs. These data were not consistent with the prediction of the excitation model, which posits that different mechanisms are involved in frequency discrimination and frequency modulation detection. Different from the place code, a temporal code extracts frequency changes from the temporal firing patterns of the nerve fibers that innervate the same cochlear place (Rose *et al.*, 1967). There is a body of evidence for a temporally based mechanism in FM detection, particularly at low-modulation rates (e.g., $<$20 Hz) and in hearing-impaired listeners (Edwards and Viemeister, 1994a; Moore and Sek, 1995, 1996). However, phase locking to sinusoids only occurs at frequencies lower than 4000–5000 Hz in the mammalian auditory nerve (Palmer and Russell, 1986), and it is unlikely that temporal cues can encode FM at high frequencies.

In acoustic hearing, one of the critical issues has been to reduce the interaction between ''temporal coding'' and ''place coding'' so that the two mechanisms can be studied separately. Electric stimulation of the auditory nerve provides a unique opportunity to address the relative contributions of the place and temporal mechanisms to frequency coding. For frequency discrimination tasks, the electric stimulation rate can be varied and delivered to the same electrode, presumably producing a change in the firing rate of the auditory nerve without any change in the excitation place. Under these conditions, electrical rate discrimination tasks have shown that some cochlear-implant users can detect pitch differences only up to 300–500 Hz (Bilger, 1977; Eddington *et al.*, 1978; Simmons *et al.*, 1981; Shannon, 1983; Fearn *et al.*, 1999; Zeng, 2002) while others show that pitch saturates at about 1000 Hz (Hochmair-Desoyer *et al.*, 1983; Townshend *et al.*, 1987). These data have been interpreted as an indication of the 500–1000-Hz upper limit for purely temporally based mechanism for pitch encoding. To our knowledge, there is only one study in the literature that has measured FM detection in cochlear-implant users (Tong *et al.*, 1982). The difference limen was about 10% of the standard frequency for a 100-Hz carrier with 10-Hz modulation frequency in two cochlear-implant users. In general, studies in electric hearing indicate that temporal cues might play a role in frequency discrimination and FM detection at low frequencies.

To further investigate the mechanism underlying dynamic frequency encoding, the present study systematically measured FM detection in cochlear-implant subjects. The goals of this study were twofold. One goal was to characterize cochlear-implant users' ability to detect frequency modulation. This goal was driven by a need in applications. Contemporary cochlear implants typically extract the temporal envelope but discard the temporal fine structure in the acoustic stimulus. The temporal envelope is typically used to amplitude modulate a fixed-rate carrier, which clearly does not reflect the natural way by which the acoustic information is transmitted to the brain via the auditory nerve. This study hopes to provide psychophysical evidence for using FM to convey additional speech and music information (Nie *et al.*, 2004b). The second goal, which has a more theoretical underpinning, was to use the electric hearing data to help delineate the mechanisms underlying FM detection in acoustic hearing.

II. METHODS

A. Subjects

Three post-lingually deafened adults with Nucleus-22 cochlear implants participated in this study. The subjects ranged in age from 45 to 69 years with a mean age of 58 years and were all native speakers of American English. All subjects had 6 or more years of experience with the device and consonant recognition scores of at least 69% correct. They also had extensive previous experience in psychophysical and speech tests. Detailed subject information is presented in Table I.

Three normal-hearing subjects also participated in this study. Their ages ranged from 20 to 25 years with a mean age of 23 years. None of the normal-hearing subjects had prior experience in psychophysical tests. They gave informed consent and were paid for their participation.

B. Stimuli

Three types of frequency modulation, upward sweep, downward sweep, and sinusoidal frequency modulation,

were used in the experiment. For the upward sweep and downward sweep, the electric stimulation parameters studied were electrode position (apical versus basal), stimulation level (most comfortable level versus soft level), and different standard frequencies. Loudness was measured using unmodulated stimuli at each standard frequency. The subjects first received subthreshold stimulation and then had to indicate when the soft and the most comfortable levels (MCL) were reached with gradually increasing current level. The MCL was 65%–70% of the maximum loudness level that subjects could tolerate, while the soft level was 25%–30% of the maximum loudness level. An additional parameter studied was the FM rate for the sinusoidal FM pattern. To avoid aliasing, the modulation rate was set at least 40% less than the standard frequency, being systematically varied from 5 to 320 Hz at the 1000-Hz standard frequency, from 5 to 160 Hz at 500 Hz, from 5 to 80 Hz at 250 Hz, from 5 to 40 Hz at 125 Hz, and from 5 to 10 Hz at 75 Hz. All stimuli were represented as 300-ms trains of biphasic pulses and were presented to a single, bipolar electrode pair in BP+1 mode (1.5-mm spacing between the active and reference electrodes). The phase duration and temporal separation between opposite phases of each pulse were 200 and 20 μs, respectively. The electric stimuli were delivered to the subject and controlled via a customized research interface[1] (Shannon *et al.*, 1990).

Figure 1 shows three types of FM stimuli in the frequency-time domain. The standard is represented by the solid line while the signal is represented by the dotted line. The parameter to be measured is the frequency difference between the signal and the standard at the end for the upward sweep, at the beginning for the downward sweep, and modulation depth for the sinusoidal FM stimuli.

To minimize the loudness cue due to the change in the stimulation rate, the instantaneous amplitude of each pulse in the electric stimuli was roved by a value that was uniformly distributed between −1 and 0 dB. Amplitude roving was applied to both standard and FM stimuli, so that levels of all three stimuli in each trial were randomized. Although 1 dB may be small, it corresponds to 10–20 dB roving in acoustic hearing because current cochlear-implant users typically have a narrow dynamic range of 10–20 dB (e.g., see Zeng and Galvin, 1999; Zeng *et al.*, 2002). Zeng and Shannon (1999) reported essentially no loudness change as a function of pulse rate from 100 to 3000 Hz in six cochlear-implant users. For example, the maximum difference in current levels producing equal loudness was only 0.40 dB between 1000 to 2000 Hz. With 30% or less frequency modulation and 1-dB roving in this study, loudness was unlikely to be a confounding cue.

In acoustic stimulation for the normal-hearing subjects, the carrier was a sinusoid generated using TDT System III equipment (RP2.1) and presented at 65 dB SPL monaurally via Sennheiser headphones (200A). Neither level roving nor background noise was used. Moore and Glasberg (1989) showed a nonsignificant effect of level randomization on FM detection in normal-hearing listeners with a relatively small range of level variations (3 dB). Emmerich *et al.* (1989) suggested that level randomization with a large range of variations could produce pitch shift with level, providing subjects with additional cues in frequency discrimination.

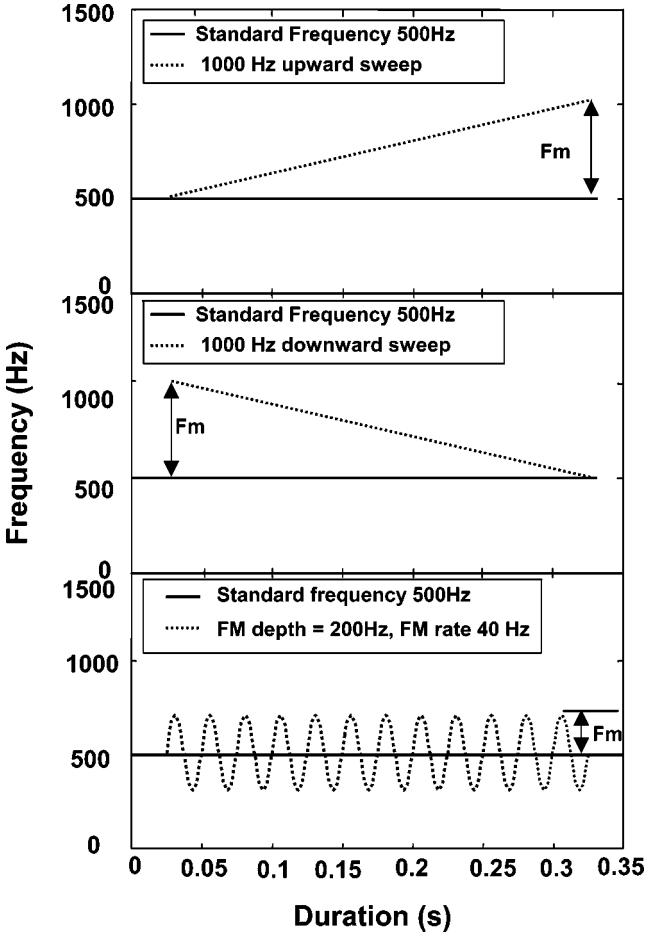


FIG. 1. Three patterns of frequency modulation examined in the present study: upward sweep (top panel), downward sweep (middle panel), and sinusoidal frequency modulation (bottom panel).

C. Procedures

FM difference limens were measured using a three-interval, forced-choice, adaptive procedure. A three-down, one-up decision rule was employed to track the 79.4% percent correct point on the psychometric function. In each trial, a subject heard three sounds including two steady-frequency standard signals and one frequency-modulated signal. The order of presentation was randomized. The subject was asked to identify the interval with the greatest change in pitch by pressing a button on a computer monitor. Visual feedback was given after each trial. The run terminated after 13 reversals or 60 trials with at least eight reversals. The step size was about 25% of the standard frequency for the first four reversals and reduced to 3%–5% thereafter. The difference limen was averaged over the last eight reversals. All subjects completed at least three successful runs with an average standard deviation of about 50% of the mean value.

III. RESULTS

A. Frequency sweep

Figure 2 shows frequency sweep difference limens in Hertz (Hz) as a function of standard frequency for the up-

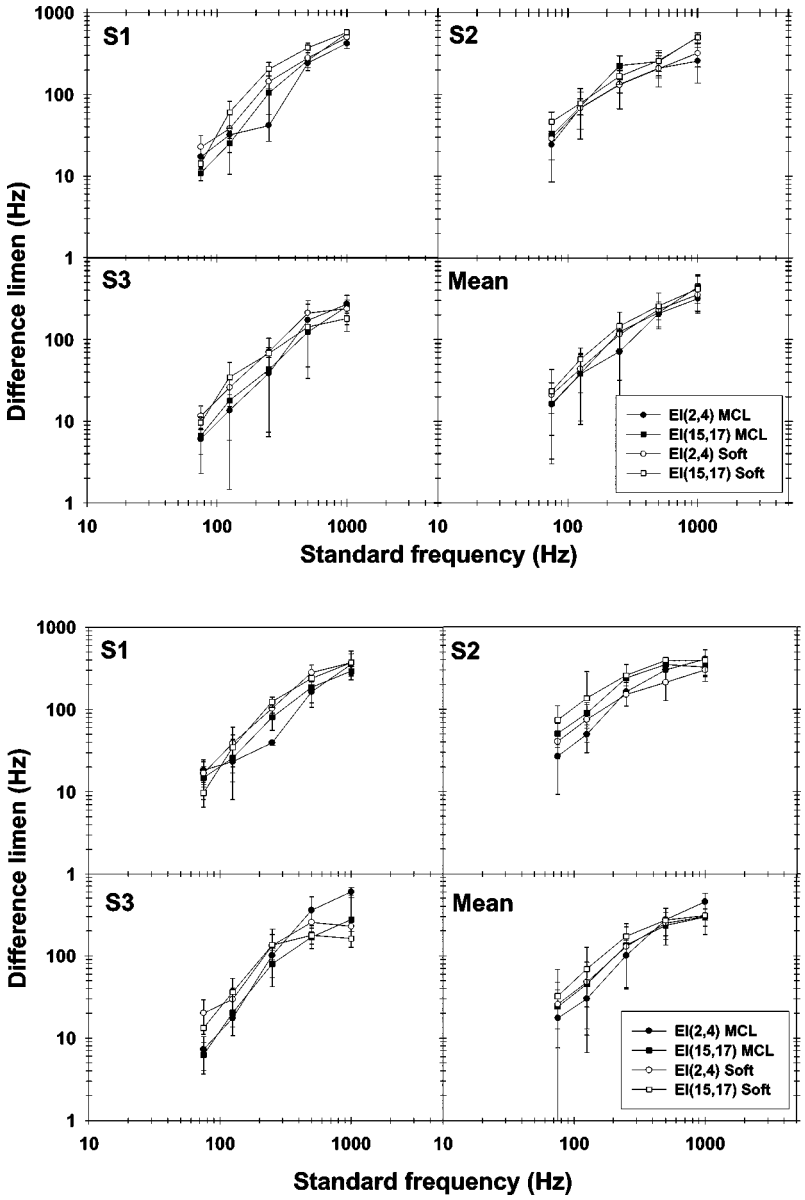


FIG. 2. Difference limens for frequency sweeps in three cochlear-implant subjects. The upper panel represents the data from the upward sweep and the lower panel represents the data from the downward sweep. Each graph shows the data for a single subject or the mean data (bottom-right graph in each panel). The *x* axis is the standard frequency and the *y* axis is frequency modulation difference limen. Circles represent the data collected from the basal electrode while squares represent the data collected from the apical electrode. The open symbols represent the data at the soft level while the filled symbols represent data at MCL.

ward sweep (upper four panels) and the downward sweep (lower four panels). The individual data are represented according to the subject number (S#) and the mean data are shown in the bottom-right panel. The data from the apical electrodes are represented by squares while those from the basal electrodes are represented by circles. The data for the MCL are represented by filled symbols while those from the soft level are represented by open symbols.

In general, the difference limen is proportional to the standard frequency regardless of the sweep type, electrode position, and loudness level. A four-way ANOVA with a within-subject design and repeated measures confirms a significant main effect for the standard frequency [$F(4,8) = 36.4$, $p < 0.01$] but not for the other variables ($p > 0.05$). To compare the FM sweep detection between acoustic and electric hearing, Fig. 3 shows the grand averaged data across all conditions in cochlear-implant subjects (filled circles) as well as the acoustic data averaged over the upward and downward sweep types from three normal-hearing subjects (open circles).[2] The grand averaged difference limen in electric hearing ranged from 22 Hz at 75-Hz standard frequency to 361 Hz at 1000-Hz standard frequency, corresponding to Weber fractions (the ratio of difference limen to the standard frequency) of 0.29 and 0.36, respectively.

In contrast to the rising function in electric hearing, normal-hearing listeners show a relatively flat function with a much smaller difference limen between 5 and 10 Hz at all standard frequencies tested. Except for the lowest standard frequency (75 Hz), there is no overlap in the difference limen functions between acoustic and electric data. The higher the standard frequency, the poorer the performance in electric hearing as compared to acoustic hearing.

B. Sinusoidal frequency modulation

Figure 4 shows the individual and mean difference limens for sinusoidal FM as a function of standard frequency in cochlear-implant subjects. Since neither stimulation level nor electrode site produced any significant effects, the data were averaged over both levels and electrodes.

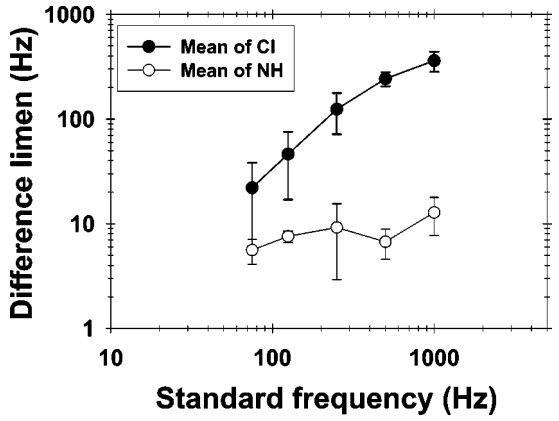


FIG. 3. The mean of difference limens for frequency sweeps for both normal-hearing (filled circles) and cochlear-implant subjects (open circles). Error bars represent plus and minus one standard error.

The obtained data were analyzed in two ways to deal with the missing data at higher modulation rates. First, the data from modulation rates at 5, 10, and 20 Hz were examined using a within-subject design and a four-way repeated-measures ANOVA (stimulation level×electrode×standard frequency×FM rate). Similar to the FM sweep result, standard frequency was the only significant main factor: the difference limen increased monotonically from about 10 Hz at the 75-Hz standard to about 100 Hz at the 1000-Hz standard [$F(4,8)=21.4$, $p<0.05$]. Second, a three-way ANOVA was used to examine the main effects of electrode, loudness, and modulation rate at different standard frequencies. No significant main effect was found except for the greater difference limen at higher modulation rates than at the lower modulation rates. The difference limen was 294.3 Hz for the 160-Hz modulation rate at the 500-Hz standard [$F(5,10)=12.0$, $p<0.05$], 400.0 and 549.4 Hz for the 160- and 320-Hz rates at the 1000-Hz standard [$F(6,6)=14.4$, $p<0.05$], respectively.

To facilitate comparison, the averaged data from three normal-hearing listeners are shown on the bottom-right panel of Fig. 4. Again, in contrast to the cochlear-implant data, the averaged normal-hearing data show a relatively flat function

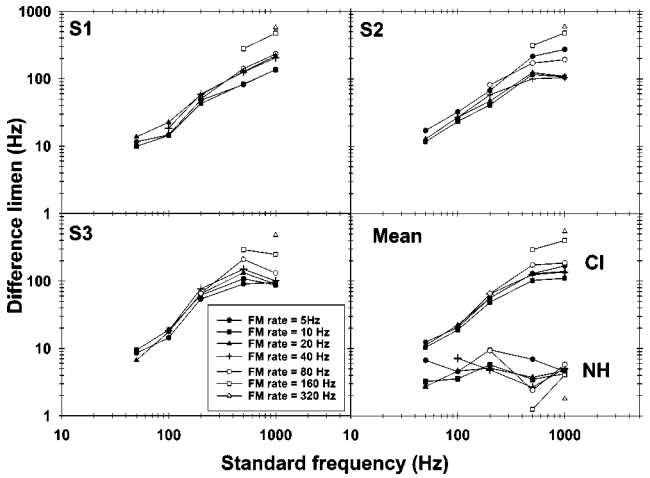


FIG. 4. Difference limens of sinusoidal frequency modulation in three cochlear-implant subjects. The bottom-right panel represents the mean value in both cochlear-implant (CI) and normal-hearing (NH) subjects. Different symbols represent different modulation rates.

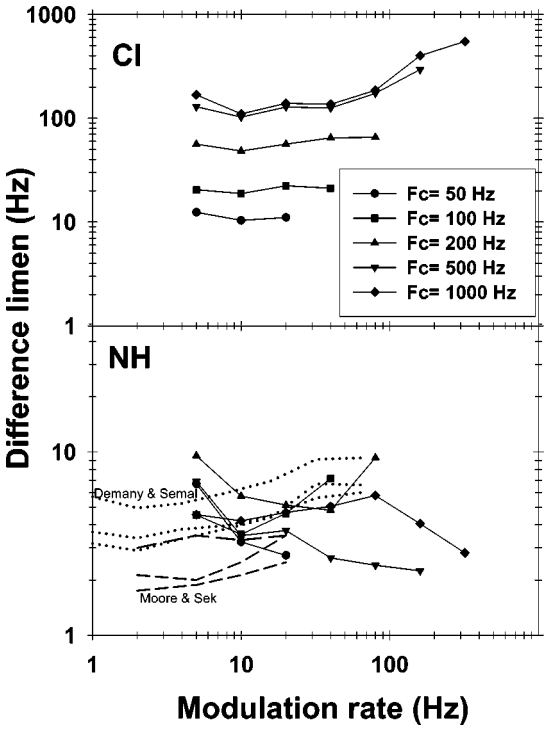


FIG. 5. Comparison of difference limens between normal-hearing and cochlear-implant subjects for sinusoidal frequency modulation. Difference limens are plotted as a function of modulation rate. Different symbols represent different standard frequencies from 50 to 1000 Hz. Previous data from Demany and Semal (1989) are presented as three dotted lines and those from Moore and Sek (1996) are presented as three dashed lines. Each of the three lines represents a different standard frequency (1000, 500, and 250 Hz for lines from top to bottom, respectively).

with no significant effects of either standard frequency or modulation rate ($p>0.05$). If anything, there was a trend for higher modulation rates (160 and 320 Hz) to produce smaller difference limens than the lower modulation rates [$F(2,4)=9.6$, $p=0.06$].

IV. DISCUSSION

A. Comparison with previous studies

Figure 5 replots the sinusoidal FM data (Fig. 4) as a function of modulation rate in cochlear-implant subjects (upper panel) and compares the difference limens between the present and previous studies in normal-hearing subjects (lower panel). Previous data from Demany and Semal (1989) are presented as three dotted lines and those from Moore and Sek (1996) are presented as three dashed lines. Each of the three lines represents a different standard frequency (1000, 500, and 250 Hz for lines from top to bottom, respectively). Different standard frequencies are represented by different symbols in the present study.

Given that only three cochlear-implant subjects were tested and all were good users, the current data may not represent the performance of the broader cochlear-implant population. However, several general trends could be gleaned from the limited data. First, for standard frequencies of 200 Hz and higher, the difference limen for sinusoidal FM is one to two orders of magnitude poorer in electric hearing than acoustic hearing. Second, the difference limen mono-

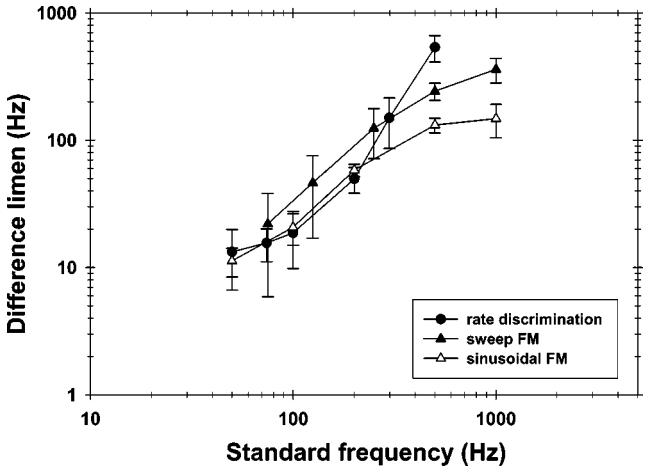


FIG. 6. Comparison of difference limens between rate discrimination (filled circles), frequency sweep (filled triangles), and sinusoidal frequency modulation (unfilled triangles) in cochlear-implant subjects. The data for the frequency sweep represent the average of both sweep types, levels, electrodes, and all subjects. The data for the sinusoidal frequency modulation represent the average across levels electrodes modulation rates, and subjects. The data for the rate discrimination represent the average of four subjects on both electrodes from a previous study (Zeng, 2002).

tonically increases with the standard frequency in electric hearing. In acoustic hearing, although the present data are in the same range as the previous data, this monotonic relationship is not as orderly as it was in the previous data. One possible reason is that relatively high ratios of modulation rates to standard frequencies were used in the current study and this may have allowed subjects to rely more strongly on place cues to differentiate the stimuli. Third, the difference limen is relatively independent of the modulation rate in both acoustic and electric hearing, with the difference limen showing an increasing trend at the highest modulation rates (160 and 320 Hz) in electric hearing and a decreasing trend at these rates in acoustic hearing.

Because only the temporal cue is available to cochlear-implant subjects, we interpret the present result as providing evidence for a significant role of the place cue in FM detection for normal-hearing listeners. The difference limens between acoustic and electric hearing are close in value only at low standard frequencies. As the standard frequency increases, FM difference limens in cochlear-implant users are elevated significantly. The opposing trend at high modulation rates (160 and 320 Hz) between acoustic and electric hearing further suggests that the temporal cue is most effective at low modulation rates ($<$80 Hz). Together, these data suggest that the temporal mechanism operates only for low standard frequencies at low modulation rates.

Figure 6 compares cochlear-implant subjects' pulse-rate difference limens from a previous study [filled circles; from Zeng (2002)] with difference limens for sinusoidal (unfilled triangles) and sweep (filled triangles) FM in the present study. The data represent a grand average across subjects, levels, and electrodes. Data could not be collected for rate discrimination at 1000 Hz as subjects reported no pitch difference between stimuli. Difference limens were generally similar for standard frequencies less than 300 Hz for all three tasks, but diverged at higher standard frequencies. Cochlear-implant subjects typically could not detect a pitch difference based on rate above 300–500 Hz, but appeared to be able to detect FM differences, with sinusoidal modulation producing the smallest difference limens at these high frequencies. The ability to detect sinusoidal frequency modulation with high carrier frequency and low modulation rate may be related to previously observed temporal pitch changes between the pitch of the carrier and the pitch of the modulator in sinusoidal amplitude modulations (McKay *et al.*, 1995; McKay and Carlyon, 1999). Further study is needed to determine whether FM and AM related pitch changes in electric hearing have the same underlying mechanisms.

There are at least two reasons that sinusoidal FM produced the smallest difference limen. First, for a constant difference limen, the frequency-varying rate in sinusoidal FM is greater than that in frequency sweeps. For instance, the difference limen of 229 Hz at 500 Hz for a 300-ms upward sweep would produce a frequency-varying rate of 763 Hz per second ($\Delta F/\Delta t = 229\,\text{Hz}/0.3\,\text{s} = 763\,\text{Hz/s}$). On the other hand, the difference limen was 129 Hz for the 300-ms sinusoidal FM with 500-Hz standard frequency and 5-Hz modulation rate. The frequency-varying rate in one cycle can be approximated by the peak-to-valley frequency difference divided by half of the period, equaling to $\Delta F/\Delta t = 129/0.1 = 1290\,\text{Hz/s}$. Second, the sinusoidal FM can be viewed as a series of frequency downward and upward sweeps (see Fig. 1). It is possible that "multiple looks" (Viemeister and Wakefield, 1991) at frequency changes produced smaller difference limens for sinusoidal FM than the "single look" for frequency sweeps.

B. Mechanisms

The present results may shed light on two unresolved issues in dynamic frequency encoding. One issue is the relative contributions of spectral and temporal cues to FM processing. The spectrum of sinusoidally frequency-modulated signal consists of the carrier frequency plus and minus all the integer multiples of the modulation rate (Chowning, 1973). At low modulation rates (5–80 Hz), a listener can process temporally varying pitch by following changes in instantaneous frequencies. This mechanism is primarily temporally based, as evidenced by the relatively close values in difference limens at these low modulation rates between acoustic and electric hearing. At higher modulation rates, FM detection is likely to be performed by detecting the timbre change induced by the additional sidebands. This is a spectrally based mechanism, requiring the ear to resolve sidebands in the frequency domain. Cochlear-implant subjects cannot perform this task via single-electrode stimulation, thus producing increasingly poorer performance at higher modulation rates.

The other issue is whether amplitude and frequency modulations are processed independently by the auditory system (Moore and Sek, 1992; Edwards and Viemeister, 1994b; Saberi and Hafter, 1995). This controversy has been difficult to resolve because amplitude modulation and frequency modulation interact at the excitation pattern level in acoustic hearing (Zwicker, 1952). A "pitch-sampling model" has been proposed to explain the FM processing, in which

FM detection is based upon successive samplings of a stimulus's pitch using a relatively short-term temporal window (Hartmann and Klein, 1980; Demany and Semal, 1986). Edwards and Viemeister (1994b) compared the discriminability of equally detectable amplitude- and frequency-modulated signals and suggested a second frequency-modulation encoding mechanism that tracks the instantaneous frequency. The present study provides direct evidence for such a mechanism given that cochlear-implant subjects cannot resolve the sidebands in frequency modulations. The present study also extends the boundary that such a temporally based cue may be available from 10 Hz or below to at least 80 and possibly 320 Hz (Moore and Sek, 1995, 1996). Because FM difference limens at these modulation frequencies are always poorer in electric hearing than in acoustic hearing, the temporal coding of FM is considered ''sluggish.''

Finally, inconsistent with the Schouten (1985) study, the present result shows no perceptual imbalance between upward and downward sweeps. This inconsistency may be explained by the findings of Kohlrausch and Sander (1995) and Carlyon and Datta (1997) that asymmetry in frequency modulation is due to the cochlear nonlinearity. In their studies, a complex tone with positive or negative Schroeder phase was used to mask a sinusoidal tone, and the result showed that the positive Schroeder-phase stimulus (downward sweep) produced much less masking than the negative Schroeder-phase stimulus (upward sweep). Interestingly, recent studies (Recio and Rhode, 2000; Summers *et al.*, 2003) presented direct evidence that the nonlinear cochlear filter produces more amplitude modulation with the positive Schroeder-phase stimulus than the negative Schroeder-phase stimulus, pointing again to a possible common mechanism between amplitude and frequency modulation. In cochlear-implant subjects, the cochlear nonlinearity is totally absent, producing no difference in FM detection between downward and upward sweeps.

C. Clinical application

The present result may also help improve speech-processing strategies for cochlear-implant subjects. Most contemporary cochlear implants have employed a fix-rate carrier strategy (Wilson *et al.*, 1991; McDermott *et al.*, 1992) and are continuing to increase the carrier rate in order to better represent the temporal envelopes and/or to restore stochastic responses in electrically stimulated auditory nerve fibers (Rubinstein *et al.*, 1999). The present study suggests that cochlear-implant subjects can benefit from a new speech processing strategy, in which, instead of using a constant, high-rate carrier, the carrier rate may be changed dynamically on each electrode according to the frequency change in the input acoustic signal (Lan *et al.*, 2004; Nie *et al.*, 2004b). The present result clearly demonstrates that cochlear-implant subjects can access the FM information at least for standard frequencies up to 1000 Hz and modulation rates up to 320 Hz. The encoding of frequency modulation with relatively lower pulse rates, in combination with amplitude modulation, could provide information regarding both the temporal fine structure and envelope in speech and music. Such a strategy may have the potential to improve pitch perception and other pitch related tasks including speaker identification, auditory object segregation, and tonal language perception (Xu and Pfingst, 2003; Kong *et al.*, 2004; Nie *et al.*, 2004a).

V. CONCLUSIONS

FM detection was assessed for upward frequency sweeps, downward frequency sweeps, and sinusoidal FM in three cochlear-implant and three normal-hearing subjects. The following conclusions could be made:

(1) Regardless of the FM type, the FM difference limen increased monotonically as a function of standard frequency in electric hearing. On the contrary, the difference limen was nearly constant as a function of standard frequency in acoustic hearing. Only at low standard frequencies (75 Hz) did the difference limen in electric hearing approach that in acoustic hearing. At high standard frequencies, the difference limen in electric hearing was one to two orders of magnitude larger than that in acoustic hearing.
(2) There was no significant difference in difference limens between upward and downward sweeps. Neither stimulation level nor electrode position was a significant factor in FM detection in electric hearing.
(3) The difference limen for sinusoidal FM detection was not significantly affected by modulation rate except at 160 and 320 Hz for detecting sinusoidal FM in electric hearing.
(4) The present result supports the existence of an independent temporally based FM coding mechanism, particularly at low standard frequencies and at low modulation rates.
(5) The present finding suggests that FM via varying pulse rate may be used to encode the temporal fine structure in cochlear implants to improve their performance in pitch perception and other pitch-related tasks such as speaker identification, auditory object segregation, and tonal language perception.

ACKNOWLEDGMENTS

We thank Neal Viemeister and two anonymous reviewers for their helpful comments on the manuscript. We also thank our cochlear-implant subjects for their time and dedication. This work was supported partially by NIH (2R01-DC02267).

[1]J. Wygonski and M. E. Robert, ''HEI Nucleus research interface specification,'' House Ear Institute (2001).
[2]No statistical significant difference was observed between the upward and downward sweeps in normal-hearing subjects [$F(1,2)=1.7$, $p>0.05$].

Arlinger, S. D., Jerlvall, L. B., Ahren, L. B., and Holmgren, E. C. (**1977**). ''Thresholds for linear frequency ramps of a continuous pure tone,'' Acta Oto-Laryngol. **83**, 317–327.
Bilger, R. C. (**1977**). ''Psychoacoustic evaluation of present prostheses,'' Ann. Otol. Rhinol. Laryngol. Suppl. **86**, 92–140.
Carlyon, R. P., and Datta, A. J. (**1997**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.
Chowning, J. (**1973**). ''The synthesis of complex audio spectra by means of frequency modulation,'' J. Audio Eng. Soc. **21**, 526–534.

Collins, M. J., and Cullen, Jr., J. K. (**1978**). ''Temporal integration of tone glides,'' J. Acoust. Soc. Am. **63**, 469–73.
Demany, L., and Semal, C. (**1986**). ''On the detection of amplitude modulation and frequency modulation at low modulation frequencies,'' Acustica **61**, 243–255.
Demany, L., and Semal, C. (**1989**). ''Detection thresholds for sinusoidal frequency modulation,'' J. Acoust. Soc. Am. **85**, 1295–1301.
Dooley, G. J., and Moore, B. C. J. (**1988**). ''Detection of linear frequency glides as function of frequency and duration,'' J. Acoust. Soc. Am. **84**, 2045–2057.
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**, 1–39.
Edwards, B. W., and Viemeister, N. F. (**1994a**). ''Psychoacoustic equivalence of frequency modulation and quasi-frequency modulation,'' J. Acoust. Soc. Am. **95**, 1510–1513.
Edwards, B. W., and Viemeister, N. F. (**1994b**). ''Frequency modulation versus amplitude modulation discrimination: evidence for a second frequency modulation encoding mechanism,'' J. Acoust. Soc. Am. **96**, 733–740.
Emmerich, D. S., Ellermeier, W., and Butensky, B. (**1989**). ''A reexamination of the frequency discrimination of random-amplitude tones, and a test of Henning's modified energy-detector model,'' J. Acoust. Soc. Am. **85**, 1653–1659.
Fearn, R., Carter, P., and Wolfe, J. (**1999**). ''The perception of pitch by users of cochlear implants: possible significance for rate and place theories of pitch,'' Acoust. Austral. **27**, 41–43.
Goldstein, J. L., and Srulovicz, P. (**1977**). ''Auditory-nerve spike intervals as an adequate basis for aural frequency measurement,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 337–347.
Grant, K. W. (**1987**). ''Frequency modulation detection by normally hearing and profoundly hearing-impaired listeners,'' J. Speech Hear. Res. **30**, 558–563.
Hartmann, W. M., and Hnath, G. M. (**1982**). ''Detection of mixed modulation,'' Acustica **50**, 297–312.
Hartmann, W. M., and Klein, M. A. (**1980**). ''Theory of frequency modulation detection for low modulation frequencies,'' J. Acoust. Soc. Am. **67**, 935–946.
Henning, G. B. (**1967**). ''A model for auditory discrimination and detection,'' J. Acoust. Soc. Am. **42**, 1325–1334.
Hochmair-Desoyer, I. J., Hochmair, E. S., Burian, K., and Stiglbrunner, H. K. (**1983**). ''Percepts from the Vienna cochlear prosthesis,'' Ann. N.Y. Acad. Sci. **405**, 295–306.
Kohlrausch, A., and Sander, A. (**1995**). ''Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets,'' J. Acoust. Soc. Am. **97**, 1817–1829.
Kong, Y.-Y., Vongphoe, M., and Zeng, F.-G. (**2004**). ''Independent contributions of amplitude and frequency modulations to auditory perception. II. Melody, tone, and speaker identification,'' 26th ARO Midwinter Research Meeting, Daytona Beach, FL.
Lan, N., Nie, K. B., Gao, S. K., and Zeng, F.-G. (**2004**). ''A novel speech-processing strategy incorporating tonal information for cochlear implants,'' IEEE Trans. Biomed. Eng. **51**, 752–760.
Loughin, P., and Tacer, B. (**1996**). ''On the amplitude-and-frequency modulation decomposition of signals,'' J. Acoust. Soc. Am. **100**, 1594–1601.
Madden, J. P., and Fire, K. M. (**1996**). ''Detection and discrimination of gliding tones as a function of frequency transition and center frequency,'' J. Acoust. Soc. Am. **100**, 3754–3760.
McDermott, H. J., McKay, C. M., and Vandali, A. E. (**1992**). ''A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant,'' J. Acoust. Soc. Am. **91**, 3367–3371.
McKay, C. M., and Carlyon, R. P. (**1999**). ''Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains,'' J. Acoust. Soc. Am. **105**, 347–357.
McKay, C. M., McDermott, H. J., and Clark, G. M. (**1995**). ''Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: the effect of modulation depth,'' J. Acoust. Soc. Am. **97**, 1777–1785.
Moore, B. C., and Glasberg, B. R. (**1989**). ''Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation,'' J. Acoust. Soc. Am. **86**, 1722–1732.
Moore, B. C., and Sek, A. (**1992**). ''Detection of combined frequency and amplitude modulation,'' J. Acoust. Soc. Am. **92**, 3119–3131.
Moore, B. C., and Sek, A. (**1995**). ''Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation),'' J. Acoust. Soc. Am. **97**, 2468–2478.
Moore, B. C., and Sek, A. (**1996**). ''Detection of frequency modulation at low modulation rates: evidence for a mechanism based on phase locking,'' J. Acoust. Soc. Am. **100**, 2320–2331.
Nabelek, I., and Hirsh, I. J. (**1969**). ''On the discrimination of frequency transitions,'' J. Acoust. Soc. Am. **45**, 1510–1519.
Nie, K., Stickney, G., and Zeng, F.-G. (**2004a**). ''Independent contributions of amplitude and frequency modulations to auditory perception. I. Consonant, vowel, and sentence recognition,'' 26th ARO Midwinter Research Meeting, Daytona Beach, FL.
Nie, K., Stickney, G., and Zeng, F.-G. (**2004b**). ''Encoding Frequency Modulation to Improve Cochlear Implant Performance in Noise,'' IEEE. Trans. Biomed. Eng. (in press).
Palmer, A. R., and Russell, I. J. (**1986**). ''Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells,'' Hear. Res. **24**, 1–15.
Pollack, I. (**1968**). ''Detection of rate of change of auditory frequency,'' J. Exp. Psychol. **77**, 535–541.
Recio, A., and Rhode, W. S. (**2000**). ''Basilar membrane responses to broadband stimuli,'' J. Acoust. Soc. Am. **108**, 2281–2298.
Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). ''Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey,'' J. Neurophysiol. **30**, 769–793.
Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999**). ''Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation,'' Hear. Res. **127**, 108–118.
Saberi, K., and Hafter, E. R. (**1995**). ''A common neural code for frequency- and amplitude-modulated sounds,'' Nature (London) **374**, 537–539.
Schouten, M. E. (**1985**). ''Identification and discrimination of sweep tones,'' Percept. Psychophys. **37**, 369–376.
Sek, A., and Moore, B. C. (**1995**). ''Frequency discrimination as a function of frequency, measured in several ways,'' J. Acoust. Soc. Am. **97**, 2479–2486.
Sergeant, R. L., and Harris, J. D. (**1962**). ''Sensitivity to unidirectional frequency modulation,'' J. Acoust. Soc. Am. **34**, 1625–1628.
Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.
Shannon, R. V., Adams, D. D., Ferrel, R. L., Palumbo, R. L., and Grandgenett, M. (**1990**). ''A computer interface for psychophysical and speech research with the Nucleus cochlear implant,'' J. Acoust. Soc. Am. **87**, 905–907.
Siebert, W. M. (**1970**). ''Frequency discrimination in the auditory system: Place or periodicity mechanisms,'' Proc. IEEE **58**, 723–730.
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.
Summers, V., de Boer, E., and Nuttall, A. L. (**2003**). ''Basilar-membrane responses to multicomponent (Schroeder-phase) signals: understanding intensity effects,'' J. Acoust. Soc. Am. **114**, 294–306.
Tong, Y. C., Clark, G. M., Blamey, P. J., Busby, P. A., and Dowell, R. C. (**1982**). ''Psychophysical studies for two multiple-channel cochlear implant patients,'' J. Acoust. Soc. Am. **71**, 153–160.
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.
Tsumura, T., Sone, T., and Nimura, T. (**1973**). ''Auditory detetion of frequency transition,'' J. Acoust. Soc. Am. **53**, 17–25.
Tyler, R. S., Wood, E. J., and Fernandes, M. (**1983**). ''Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **74**, 1190–1199.
Viemeister, N. F., and Wakefield, G. H. (**1991**). ''Temporal integration and multiple looks,'' J. Acoust. Soc. Am. **90**, 858–865.
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.
Xu, L., and Pfingst, B. E. (**2003**). ''Relative importance of temporal envelope and fine structure in lexical-tone perception,'' J. Acoust. Soc. Am. **114**, 3024–3027.
Zeng, F.-G. (**2002**). ''Temporal pitch in electric hearing,'' Hear. Res. **174**, 101–106.
Zeng, F.-G., and Galvin, J. J. (**1999**). ''Amplitude mapping and phoneme

recognition in cochlear implant listeners,'' Ear Hear. **20**, 60–73.
Zeng, F.-G., and Shannon, R. V. (**1999**). ''Psychophysical laws revealed by electric hearing,'' NeuroReport **10**, 1931–1935.
Zeng, F.-G., Grant, G., Niparko, J., Galvin, J. J., Shannon, R. V., Opie, J., and Segel, P. (**2002**). ''Speech dynamic range and its effects on cochlear implant performance,'' J. Acoust. Soc. Am. **111**, 377–386.
Zwicker, E. (**1952**). ''Die Grenzen der Horbarkeit der Amplitudemodulation und der Frequenzmodulation eines Tones,'' Acustica **2**, 125–133.
Zwicker, E. (**1956**). ''Die elementaren Grundlagen zur Bestimmung der Informationskapazitat des Gehors,'' Acustica **6**, 356–381.

Masker-first advantage for cues in informational masking

Virginia M. Richards[a] and Rong Huang
Department of Psychology, Suite 302C, 3401 Walnut Street, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Gerald Kidd, Jr.
Communication Disorders and Hearing Research Center, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 2 February 2004; revised 28 June 2004; accepted 29 June 2004)

The detectability of a pure-tone signal may be reduced by adding a small number of randomly drawn masker tones remote from the signal frequency, an effect attributed to informational masking. A pretrial cue consisting of either the upcoming signal or masker releases informational masking, but a pretrial cue of the signal-plus-masker stimulus does not. In these experiments the relative potency of pre- and posttrial cues in releasing informational masking was examined. In separate conditions the masker-alone and signal-plus-masker stimuli were cues. The results indicated a masker-first advantage, i.e., sensitivity was superior when a masker cue preceded a yes/no trial interval compared to (a) when a signal-plus-masker preceded the trial, and (b) when either cue type followed the yes/no trial interval. A masker-first advantage was also obtained when the results from a two-interval forced-choice same/different task were examined. In contrast, a masker-first advantage was not obtained when the frequency of the signal to be detected was random. For detection tasks using random multi-tone maskers there may be differences in processing efficiency depending on the order in which stimuli are presented. The ''masker-first advantage'' may depend, in part, on observers maintaining their attention at the signal frequency. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1784433]

PACS numbers: 43.66.Mk, 43.66.Dc [NFV] Pages: 2278–2288

I. INTRODUCTION

Observers' ability to detect a tone of fixed frequency added to a spectrally sparse multi-tone masker is degraded when the frequencies of the masker components are drawn at random on each presentation. Randomizing the frequencies of the masker components appears to introduce informational masking (e.g., Watson, 1987). Recent experiments reported by Richards and Neff (2004) showed that substantial release from informational masking was achieved when the detection trial was preceded by a pretrial cue. In those experiments, the pretrial cue was an exact preview of either the signal or the masker, depending on which type of cue was tested. For yes/no trials (a reminder task; e.g., Macmillan and Creelman, 1991), Richards and Neff found that a preview of the signal led to an average release in masking (relative to no cue) of 14.4 dB and the preview of the masker led to an average release in masking of 13.9 dB. Presumably, the cue reduced uncertainty in the task and thereby provided a release from informational masking.

In light of the results of Richards and Neff (2004), it is surprising that Richards and Huang (2003) reported no release from informational masking when a preview of the signal-plus-masker preceded a yes/no trial. For the six observers, the average threshold with a signal-plus-masker pretrial cue was approximately 3 dB *higher* than when there was no cue at all. In contrast, a masker cue led to an 8-dB reduction in threshold compared to no cue. This result casts doubt on the interpretation that pretrial cues reduce informational masking because the cues reduce stimulus uncertainty—were that so, it is difficult to imagine why a preview of the masker afforded a reduction in uncertainty while a preview of the signal-plus-masker did not. The purpose of this study was to attempt to better understand this cuing asymmetry and to learn more about the role of cuing in general in providing a release from informational masking.

The four experiments presented below report a variety of conditions in which the asymmetric cuing effect described above was explored. To some degree the exploration was sequential; the results of experiment I contributed to the development of experiment II, etc. (although the observers did not necessarily run the experiments sequentially). To perhaps a larger degree, though, the exploration was derived from basic questions as to what the underlying mechanism(s) might be.

Experiment I examined whether the cuing asymmetry remained when the cue followed, rather than preceded, the yes/no detection trial. The maskers were drawn at random prior to each two-interval sequence. Depending on whether the cue was presented in the first or second interval, the cue either preceded or followed the trial interval. Two types of cues, a copy of the masker and a copy of the signal-plus-masker, were tested. In the second experiment, we attempted to determine whether the cuing asymmetry reflected a heightened sensitivity to the addition of a signal tone compared to the removal of a signal tone (e.g., Bregman, 2003). Here the terms ''addition'' and ''removal'' refer to the sequential aspects of the two intervals of each trial. Observers

[a]Electronic mail: richards@cattell.psych.upenn.edu

indicated whether a sound was composed of N or $N+L$ tones, where N was six and L's of one and two were tested. In separate conditions, a preview of the N or $N+L$ stimulus was provided prior to each yes/no trial interval. If the cuing asymmetry depended on differential sensitivity to increments versus decrements in the number of components, our expectation was that sensitivity would be superior for the N-component compared to the $N+L$-component pretrial cue.

In the third experiment, the question was whether the cuing asymmetry remained when there was uncertainty with regard to cue type and to the temporal position of a cue within the trial sequence. The four conditions of experiment I were repeated, except that they were intermixed within a block of trials. Extrapolating from experiment I, on any one trial the cue might be in the first or second interval (pre- and posttrial cues of experiment I), and might be a copy of the masker or the signal-plus-masker. Under these circumstances, the detection task was equivalent to a two-interval, forced-choice (2IFC) same/different task. Comparing the results of this experiment to those of experiment I provided a rough estimate of the ''level of processing'' from which the cuing asymmetry was driven. To the degree that the cuing asymmetry remained in the face of randomization of the type of cue and the temporal position of the cue interval relative to the yes/no trial interval, that result would suggest the underlying mechanism is relatively low level, and possibly obligatory. On the other hand, if randomizing cue type and temporal position disrupted the cuing asymmetry, that result would suggest the asymmetry depended, at least in part, on observers' knowledge of the trial structures.

In the fourth experiment several conditions derived from experiment I were tested. These conditions examined whether the cuing asymmetry remained even when the signal frequency was randomly chosen, when the masker was fixed across all trials, and when the two stimulus intervals were presented to different ears.

II. EXPERIMENT I: EFFECT OF CUE TYPE AND TEMPORAL POSITION OF CUE

In experiment I the effect of the type of cue and the temporal position of the cue relative to the yes/no trial interval were jointly examined in an informational masking task. The signal to be detected was a 1000-Hz tone, and the masker was a six-component complex whose components had equal level but randomly drawn phases and frequencies. The task was to indicate whether the signal was present and d' was the dependent variable. Four conditions were tested depending on the type of cue and its position relative to the yes/no trial. The terms PreMCue, PreSMCue, PostMCue, and PostSMCue are used to distinguish the four conditions. ''Pre'' and ''Post'' specify the position of the cue relative to the detection trial. ''M'' and ''SM'' specify the type of cue, masker or signal-plus-masker.

Data collection was blocked. As a result, there was no uncertainty as to the cue type or the cue's temporal position relative to the detection trial.

A. Methods

1. Observers

Results from a total of 14 normal-hearing observers are reported. A fifteenth observer participated in this experiment and some of the experiments reported below, but her data were very erratic and therefore were not included. The observers range in age from 18 to 32. All had thresholds in quiet of 15 dB HL or better for octave frequencies from 250 to 8000 Hz as determined by audiometric testing. Except for Obs 5, the second author, all were paid for participation.

2. Stimuli and procedures

Observers were instructed to press one key if the two stimuli had the same sounds and another key if the two stimuli had different sounds. Feedback appropriate to those instructions was provided after every trial.[1]

Before each two-interval stimulus sequence, the phases of the masker components were drawn at random from a uniform distribution with a range of 2π rad. Additionally, the frequencies of the masker components were independently drawn from a uniform distribution on a logarithmic scale ranging from 200 to 5000 Hz with the exception that the component frequencies were not allowed to fall within a ''protected'' region of $\pm 12\%$ surrounding the 1000-Hz signal frequency. The purpose of the protected region was to reduce the amount of energetic masking that occurred (i.e., masking which is assumed to reflect interactions of the masker and signal at the auditory periphery).

The masker components were presented at a level of 50 dB SPL per component. The signal was played synchronously with the masker. The total stimulus duration was 100 ms including 5-ms cosine-squared onsets and offsets. The two observation intervals were separated by 350 ms. Richards and Neff (2004) reported little effect on thresholds of the time between a pretrial masker cue and the subsequent trial. The 350-ms ISI is in the middle of the range of values they studied.

The digitally generated stimuli were presented using two channels of a 16-bit DAC played at a sample rate of 20 kHz. The signal and maskers were then low-pass filtered at 7 kHz using matched filters (Stewart VBF 10M), separately attenuated, summed and presented diotically using Sennheiser HD410SL headphones. Observers were tested individually in a double-walled sound booth.

Observers 1–4 participated only in experiment I. The remaining observers participated in at least one of the other experiments reported below. Of those, Obs 5, 10, and 11 ran this experiment first. Before starting the experiment, observers practiced for 2–3 h. In addition, at least 300 practice trials preceded data collection when an observer began a new condition.

Each condition was tested using two signal levels (''high'' and ''low'') that were typically chosen during the initial 2–3 h of practice. When an observer had previously run another experiment (e.g., Obs 8, 9, 12, 13, and 14), the signal level used in that experiment was the high signal level in this experiment. The aim was to choose signal levels that, averaged across conditions, yielded percent correct scores

TABLE I. The low and high signal levels tested for experiment I are listed for the individual observers. Levels are in dB SPL.

	Low	High
Obs 1	36	45
Obs 2	33	51
Obs 3	36	45
Obs 4	30	51
Obs 5	30	42
Obs 6	36	45
Obs 7	39	48
Obs 8	36	51
Obs 9	36	51
Obs 10	36	48
Obs 11	33	45
Obs 12	28	39
Obs 13	38	44
Obs 14	40	48

near 70 and 90. Due to differences in sensitivity for different conditions, individual observers sometimes approached floor and ceiling levels of performance in one or another of the conditions. When choosing signal levels, the latter was more carefully controlled against than the former. The signal levels tested are listed in Table I for each observer.

Data collection in each condition and at each signal level was blocked. For Obs 1–5 the precue conditions were run before the postcue conditions. Within these subdivisions (pre-versus posttrial cues) the order in which the different signal levels and cue types were tested was random. For Obs 6–11 the four conditions (pre- and posttrial cues, masker and signal-plus-masker cues) and two signal levels were tested in random order. For Obs 12–14 the conditions were run in random order, but the high signal levels were run before the low signal levels. Trials were presented in either 50- (Obs 1–11) or 60- (Obs 12–14) trial sets, and participants were encouraged to take a break after finishing five sets. A minimum of 500 trials were used to estimate d'.

The estimate of d' was calculated using a standard equation for yes/no trials, $d'=z_{Hit}-z_{FA}$ (Macmillan and Creelman, 1991). That is, regardless of whether the cue preceded or followed the yes/no interval, d' depended only on the observer's responses relative to the presence or absence of the signal in the yes/no trial. For example, if the cue followed the trial, the hit rate is the rate at which the observer correctly detected that the first interval contained a signal-plus-masker stimulus.[1] For percent correct scores below 55,[2] d's were set to zero.

B. Results and discussion

Figure 1 shows the values of d' averaged across observers separately for the low and high signal levels (which yield lower and higher d's). Points plotted to the left indicate values of d' when the cue preceded the yes/no trial and points plotted to the right indicate values of d' when the cue followed the yes/no trial. Squares are for the masker-cue conditions and circles are for the signal-plus-masker-cue conditions. Error bars are the standard errors of the mean across 14 observers. Figure 2 shows the results averaged across signal levels and then across observers. In all other respects the figure is as Fig. 1, except that the ordinate has a smaller range of values. The appendix lists the individual results for each condition of experiment I, allowing a comparison between the summary and individual results.

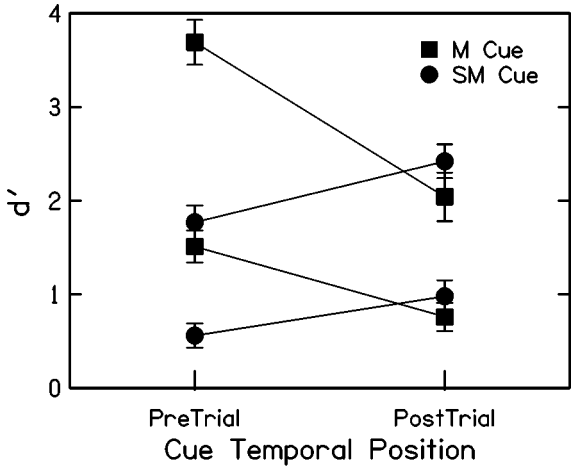


FIG. 1. Values of d' averaged across observers are plotted for low and high signal levels (yielding lower and higher d's) for pre- (left) and post- (right) trial cues. Squares are for the masker-cue conditions (MCue) and circles are for the signal-plus-masker-cue conditions (SMCue). Error bars are the standard errors of the mean across 14 observers.

The results plotted in Figs. 1 and 2, are consistent with the findings of Richards and Huang (2003). For pretrial cue conditions, d's were higher when the cue provided a preview of the subsequent masker stimulus (squares) than when it provided a preview of the subsequent signal-plus-masker stimulus (circles). This result held for both signal levels tested (Fig. 1). Averaged across signal levels, for all 14 observers the maximum d' was achieved in the PreMCue condition (Fig. 2, left, squares). For 10 of the 14 observers the minimum d' occurred in the PreSMCue condition (Fig. 2, left, circles).

Considering the postcue conditions, averaged d's tended not to depend strongly on the type of cue that followed the detection trial. On average the d's estimated in the PostSMCue condition were only 0.3 units higher than the d's estimated in the PostMCue condition. A coarse summary of the averaged results shown in Fig. 2 is that d's

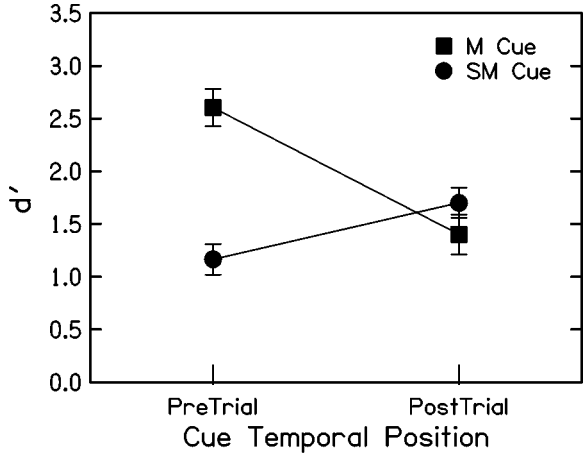


FIG. 2. The d' values averaged across levels and observers are plotted as in Fig. 1.

were higher for the PreMCue condition and relatively similar in the other three conditions. For that reason, the data shown in Figs. 1 and 2 will be referred to as demonstrating a ''masker-first advantage.''

An ANOVA applied to the data of experiment I revealed that all potentially significant features, both main effects (signal level, cue type and temporal position of cue) and interactions, were statistically significant at a p value less than 0.02. Floor effects may play an important role in generating significant interactions. As an example, d' scores for 5 of the 14 observers were coded as zero in the PreSMCue condition for the low signal level. Therefore, changing from low to high signal levels would not be expected to produce parallel lines (see Fig. 1). This floor effect makes it likely that significant interactions would be obtained when signal level was a factor.

To summarize, the current experiment demonstrates a masker-first advantage. This term incorporates three features concerning stimulus cues in informational masking: (a) A pretrial masker cue provides a larger release from informational masking than does a pretrial signal-plus-masker cue, (b) the advantage depends on the temporal order of the cue and trial, and (c) much of the difference in values of d' for pre- and posttrial cues is due to the unusually high d's obtained when a masker cue preceded an informational masking trial.

III. EXPERIMENT II: EFFECT OF REMOVING LEVEL DIFFERENCES AND DETECTION OF CHANGES IN NUMBER OF COMPONENTS

The two experiments described in this section will be referred to as experiments IIa and IIb. In experiment IIa the conditions were identical to the conditions of experiment I, except that the addition of the signal did not lead to a consistent increment in level relative to masker-alone trials. This was achieved in two ways. First the overall levels of the masker and signal-plus-masker stimuli were equated. Second, the levels were randomly perturbed using draws from a uniform distribution with a 5-dB range. This experiment was designed to determine whether the increment in level associated with the change from a masker to a signal-plus-masker stimulus (cue followed by a signal trial) contributed to the masker-first advantage obtained in experiment I.

In experiment IIb a task similar to that described by Neff and Green (1987) was used. Observers indicated whether a complex sound was composed of N versus $N+L$ tonal components. The frequencies of the $N+L$ tones were drawn at random and the complex sounds had equal level regardless of the number of components. In addition, in separate conditions a pretrial cue of either N or $N+L$ tones was provided. Accordingly, each trial had two intervals, a cue interval followed by a yes/no trial interval. When the pretrial cue had N tones, the same N tones were presented on the subsequent trial interval—either with or without the additional L tones. When the pretrial cue had $N+L$ tones, the subsequent yes/no trial interval was composed of either those same $N+L$ tones, or only N of the tones.

If one considers the N-tone complex as the masker and the $N+L$-tone complex as the signal-plus-masker, then this experiment reproduces the pretrial-cue conditions of experiment I. That is, the pretrial cue can be thought of as a preview of either the masker-alone or signal-plus-masker stimulus. As a result, this experiment explores the question of whether observers are more sensitive to increments than decrements in the number of components. A fundamental difference between this experiment and experiment I is that here the added signal had a frequency (or frequencies if $L=2$) that was random. As a result, unlike experiment I, observers could not direct their attention to a signal frequency that was constant across trials.

A. Methods

The methods used for experiment IIa were nearly identical to those of experiment I. The signal levels tested were the same as those tested in experiment I. For one observer the mean overall stimulus level was equal to the masker stimulus, 58 dB SPL, while for the other two observers the mean stimulus levels had the slightly higher level that was equal to the signal-plus-masker stimulus. In addition, the overall level was chosen at random from a uniform distribution with a 5-dB range using a 0.1-dB gradation. Each estimated d' was based on responses to 600 trials.

Observers 12–14 participated in this experiment. For observers 13 and 14 data were collected at the same time as the data collected in experiment I. Observer 12 ran this experiment after completing experiment III. As in experiment I the feedback depended on whether the two intervals contained the same or different stimuli, but the d's were estimated assuming a yes/no trial structure.

For experiment IIb the stimuli, composed of N or $N+L$ tones, had median levels of 60 dB SPL. In addition, a 5-dB rove of the overall level was applied to the cue and yes/no trial intervals using a 0.1-dB gradation. Consistent with the methods adopted by Neff and Green (1987), the stimulus durations were 150 ms rather than the 100 ms tested in the other experiments reported here.

Prior to each two-interval stimulus sequence, the frequencies of the $N+L$ tones were chosen at random between 200 and 5000 Hz using a uniform distribution on a logarithmic scale. N of those tones were arbitrarily designated as the ''masker'' stimulus. L's of one and two were tested, with the value of L being chosen at random prior to each two-interval presentation sequence. Observers 6–11 participated in this experiment. Observers 6–9 ran this experiment before experiment I, while Obs 10 and 11 ran this experiment after experiment I. The feedback used was appropriate for a same/different task, but the d's were estimated assuming a yes/no trial structure. Responses to at least 360 trials were used to estimate d'.

B. Results and discussion

The results for experiment IIa are listed in top portion of Table II and the summary data are shown in Fig. 3 (filled symbols). Corresponding results from experiment I are also tabulated (lower portion of Table II) and plotted (unfilled symbols in Fig. 3). For Table II the low and high signal levels are indicated to the left and right, respectively, and the results are listed separately for the different cue positions

TABLE II. The d's for the individual observers are compared for experiment IIa (level difference removed) and experiment I. The masker (M) or signal-plus-masker (SM) cue either preceded (Pre) or followed (Post) the yes/no trial. Averages and standard errors of the mean are also shown for each experiment.

	Low signal level				High signal level			
	PreM	PostM	PreSM	PostSM	PreM	PostM	PreSM	PostSM
				Experiment IIa				
Obs12	1.3	0	0	0	2.1	0.6	0	1.4
Obs13	2.3	1.0	0	0.9	3.7	1.0	0.9	1.7
Obs14	1.3	1.5	1.1	0.9	2.1	2.3	1.4	2.1
AVG	1.6	0.8	0.4	0.6	2.7	1.3	0.8	1.7
SEM	0.3	0.5	0.4	0.3	0.5	0.5	0.4	0.2
				Experiment I				
Obs12	0.5	0	0	0.3	2.4	1.2	0.6	1.7
Obs13	1.7	0.9	0.7	2.0	3.2	2.0	1.6	2.3
Obs14	1.4	1.3	1.2	1.8	2.9	2.5	2.3	2.0
AVG	1.2	0.7	0.6	1.4	2.9	1.9	1.5	2.0
SEM	0.4	0.4	0.3	0.5	0.2	0.4	0.5	0.2

(pre-versus posttrial) and the different cue types [masker (M) or signal-plus-masker (SM)]. In Fig. 3 the abscissa indicates the position of the cue relative to the yes/no trial and the ordinate is d'. Squares are for the masker-cue conditions and circles for the signal-plus-masker-cue conditions. Comparing the results from experiments IIa and I, removing level differences had no effect on the values of d' for conditions in which the masker was the cue. When the signal-plus-masker was the cue, d''s fell an average of 0.5 d' units. This reflects lower d''s in the PostSMCue condition for the low signal level and in the PreSMCue conditions for the high signal level (see Table II). The consequence of removing useable level differences between the masker and signal-plus-masker stimuli was to *enhance* the efficacy of the masker cue relative to the signal-plus-masker cue.

Figure 4 displays the averaged results for experiment IIb. The ordinate shows d' and the abscissa indicates the number of added tones that constitute the signal, $L=1$ (left) or 2 (right). Squares show the averaged d''s for the condition in which the pretrial cue was the N-component stimulus and circles show the averaged d's when the pretrial cue was the $N+L$-component stimulus. Although d' increased as L increased from one to two, there was no effect of the number of components in the pretrial cue stimulus. Thus, changes in the number of components *per se* did not drive the masker-first advantage obtained in experiment I.

To summarize, the results of experiment IIa indicated that level equalization did not adversely affect the masker-first advantage, and may have enhanced the effect because the signal-plus-masker cue became less beneficial. The results of experiment IIb failed to reveal a "fewer tones first" advantage parallel to the masker-first advantage obtained in experiment I. Given the findings in experiment IIa, the difference in the results of experiments I and IIb are unlikely to depend on the fact that useable level differences were removed in experiment IIb. Rather, the fact that a masker-first advantage was obtained in experiments I and IIa but not IIb may be driven by other differences in stimuli, including

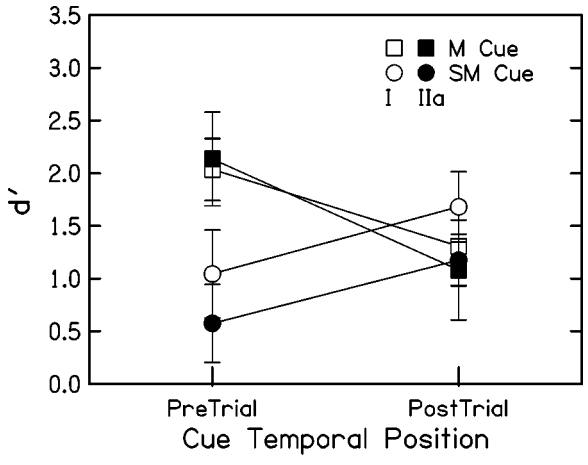


FIG. 3. The d' values averaged across level and observers are plotted as a function of whether the cue preceded (left) or followed (right) the trial. Squares are for the masker-cue conditions and circles are for the signal-plus-masker-cue conditions. The filled symbols are for experiment IIa (no level differences) and the unfilled are for experiment I. Error bars are the standard errors of the mean across three observers.

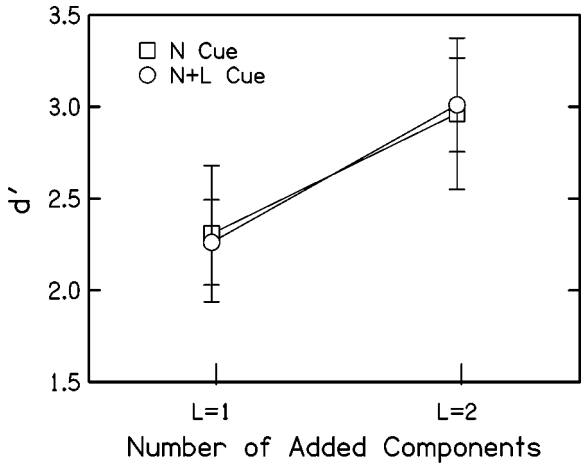


FIG. 4. The d' values averaged across observers are plotted for experiment IIb. $N=6$ and L's of one and two were tested (abscisa). The pretrial cue either previewed the N-tone or the $N+L$-tone complex (squares and circles, respectively). Error bars indicate the standard error of the mean across three observers.

whether or not the signal frequency was randomized. The masker-first advantage may have depended in part on observers having directed their attention toward a known signal frequency (experiments I and IIa) and accordingly disappeared when signal frequency was uncertain (experiment IIb).

IV. EXPERIMENT III: DETECTION OF AN ADDED TONE IN A SAME/DIFFERENT TASK

The aim of experiment III was to explore whether the masker-first advantage reflected observers' reliance on a different detection strategy when the masker was reliably present in the first interval than when it was not. This issue was examined by intermixing trials drawn from all four conditions tested in experiment I: PreMCue, PreSMCue, PostMCue, and PostSMCue. Intermixing conditions encourages the observer to adopt a single strategy for all trial types encountered. As a result, if the masker-first advantage reflected different strategies in the PreMCue condition than in the other conditions, intermixing conditions could reduce or eliminate the advantage. It is important to appreciate, however, that generalized uncertainty effects may lead to the same prediction. An increase in uncertainty, if treated as an added source of variance,[3] would have a larger detrimental effect in the PreMCue condition where d's are highest and a smaller detrimental effect in the PreSMCue condition where the d's are lowest. Referring to Fig. 2, increasing uncertainty would be expected to both reduce d's and reduce the magnitude of the interaction. If increased uncertainty does not diminish the masker-first advantage, it is unlikely that the advantage reflects a difference in strategy for the PreMCue conditions compared to the other conditions studied in experiment 1.

Note that intermixing trials from the four conditions of experiment I leads to a same/different procedure. The logic is as follows. Denoting the masker stimulus using an M, and the signal-plus-masker stimulus using an SM, the trials of experiment I may be identified as follows. For the PreMCue condition, the two possible trial sequences are ⟨M M⟩ and ⟨M SM⟩. As another example, trials from the PostSMCue condition are ⟨M SM⟩ and ⟨SM SM⟩. Together, the four conditions of experiment I include four trial types: ⟨M M⟩, ⟨M SM⟩, ⟨SM M⟩, and ⟨SM SM⟩. When these different trial structures are intermixed and chosen at random a same/different procedure results.

The analysis of the data obtained in this experiment was devised in an effort to estimate the detrimental effect of trial-type randomization. Towards that end, responses to different trial types were extracted from the data set and used to estimate values of d' for comparison with those obtained in experiment I. For example, extracting responses to the ⟨M, M⟩ and ⟨M, SM⟩ trials in the current intermixed-trials condition provides data that may be compared to those obtained from the blocked-trials design used in the PreMCue condition of experiment I. These extracted trials form a ''virtual'' PreMCue condition. ''Virtual'' PreSMCue, PostMCue, and PostSMCue conditions were formed in a similar manner. After extracting the trials appropriate to each of these virtual conditions, values of d' were determined in the same manner as in experiment I. Difference in d's estimated from the blocked (experiment I) and virtual (current experiment) conditions indicated the detrimental effect of intermixing trial types.

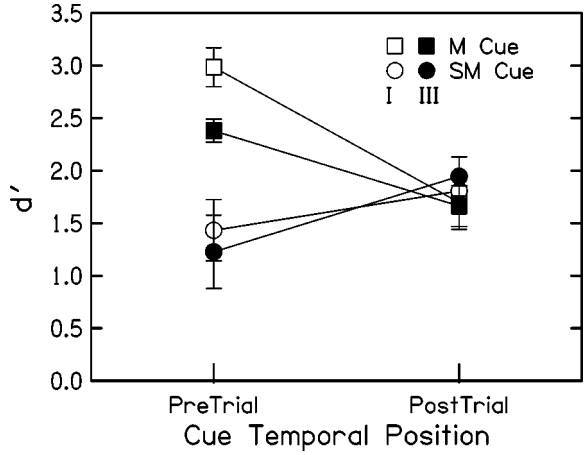


FIG. 5. The d' values averaged across observers are compared for the same/different task (filled) and experiment I (unfilled). Squares are for the masker-cue-conditions and circles are for the signal-plus-masker-cue conditions. In other respects the figure is as Fig. 2.

A. Methods

The methods used for this experiment were similar to those of experiment I. The fundamental difference, as indicated above, was that the trials from the different conditions were intermixed rather than blocked. The signal level tested in this experiment was the ''high'' signal level of experiment I. The maskers were drawn at random prior to each trial and the same masker was used in the two observation intervals. Observers were instructed to indicate whether the two intervals contained the same versus different sounds, and feedback appropriate to those instructions was provided after each trial. Observers 5 and 11–14 participated in this experiment. All of the observers, except Obs 12, had completed at least one other experiment before beginning this experiment.

One aspect of the ''virtual'' analysis was that responses to the same trials were used to form two virtual conditions. Consider, for example, the trials in which there was a signal-plus-masker in the first interval and a masker alone in the second interval. Using the notation introduced above, this would be a ⟨SM M⟩ trial. Responses to these trials contributed to two virtual conditions; the virtual PreSMCue condition (a no-signal trial because there was no signal in the second, yes/no trial interval) and the virtual PostMCue condition (a signal trial because there was a signal in the first, yes/no trial interval). The d's extracted for different virtual conditions, therefore, are not all statistically independent from one another. Each virtual d' was based on 600 trials which were derived from a total of 1200 trials.

B. Results and discussion

The filled symbols in Fig. 5 show the results of the same/different experiment averaged across observers. The leftmost points are for the virtual precue conditions and the rightmost points are for the virtual postcue conditions. The squares are for the virtual masker-cue conditions and the

circles are for the virtual signal-plus-masker-cue conditions. Bear in mind that these "virtual" d's were derived from trials extracted from the larger data set. The unfilled symbols in Fig. 5 show the results from experiment I for the five observers who were tested in this experiment. The pattern of results for four of the five observers corresponds closely to the averaged patterns shown in Fig. 5. For Obs 14 there was, for all intents and purposes, no difference between the four virtual conditions in the current same/different experiment.

The averaged data indicate that intermixing trial types reduced the values of d' more in the PreMCue condition than the other conditions. It might be that this reduction reflects a floor effect in the other conditions, where d's were already lower. This seems unlikely given that the obtained d's were above chance levels of performance, plus the fact that a modest reduction in d' was obtained for two of the three remaining conditions. Even noting that the largest detrimental effect of intermixing conditions occurred when the masker stimulus was in the first interval, d's in the virtual PreMCue condition remained higher than in the other conditions. Thus, intermixing trial types reduced, but did not abolish, the masker-first advantage. An ANOVA applied to the virtual conditions of this experiment did not reveal statistically significant main effects at the 0.05 level ($p>0.05$ for cue type, $p>0.2$ for temporal position of the cue). The interaction term, however, was significant [$F(1,4)=10.1, p<0.05$]. The pattern of results shown in Fig. 5 reflects the following pattern in the raw data: performance, as percent correct, was generally good when the masker was in the first interval (regardless of what followed) and generally poor when the signal-plus-masker was followed by a masker stimulus.

These results suggest that observers might take special advantage of the masker being played in the first interval when the masker was a cue, i.e., when the first interval was reliably the masker stimulus. Under those conditions, a unique perceptual cue might be incorporated into the observers' detection decision. An alternative account is that uncertainty more strongly affected the PreMCue condition than the other conditions.[3] Whatever the explanation, the masker-first advantage remained robust for four of the five observers even when the trial types were intermixed. It would appear that only Obs 14 was likely to have substantively altered his detection strategy among the different conditions of experiment I.

V. EXPERIMENT IV: EFFECT OF RANDOM SIGNAL FREQUENCY, FIXED MASKERS, AND DICHOTIC PRESENTATION

Experiment IV is composed of three experiments, which will be referred to as experiment IVa, IVb, and IVc. All of the experiments used the same/different task described in experiment III. And, as in experiment III, the responses to different trial types were extracted to form virtual conditions. These experiments further explored conditions in which the masker-first advantage might be maintained/abolished.

In experiment IVa the effect of randomly assigning the signal frequency prior to each trial on the masker-first advantage was revisited. We were interested in examining signal frequency randomization in more detail because it was the only manipulation in any of the experiments described above that abolished the masker-first advantage.

In experiment IVb stimulus uncertainty was minimized. In this experiment, the signal and masker frequencies were fixed for all trials. This experiment was designed to help us better understand the somewhat ambiguous role of varying uncertainty in the earlier experiments. In experiment IVc the stimuli were presented dichotically. The first interval of the same-different task was presented to the left ear and the second interval was presented to the right ear. This experiment was included in order to test whether the masker-first advantage might be due to a peripheral effect, such as neural adaptation.

A. Methods

The methods are similar to those of experiment III. Deviations from those methods are detailed separately for each experiment below. Observers 5 and 12–14 participated in these experiments. All observers had completed experiments I, IIa and III before beginning this experiment.

In experiment IVa the signal frequency was randomly chosen from a range of 200–5000 Hz on each trial. The random masker was constrained such that no component fell within $\pm 12\%$ of the randomly drawn signal frequency. This process was followed even on masker-alone trials; that is, a signal frequency was chosen and masker components were excluded from falling near that frequency but the signal was not presented. The signal level tested in this condition was the high signal level of experiment I. As a result, unlike experiment IIb for $L=1$, the level of the signal was not equal to the levels of the masker components which we hypothesized might have been a factor in abolishing the masker-first advantage.

In experiment IVb the six tones that comprised the masker were fixed in frequency across all trials. The component frequencies were 252, 380, 795, 1259, 2200, and 3973 Hz. The phases of the individual components were randomly drawn prior to each trial. This particular masker was chosen based on results obtained in a separate experiment that indicated this masker provided substantial masking for most subjects. The signal, when present, had fixed level and frequency (1000 Hz), but a randomly drawn phase. Compared to conditions in which masker frequencies were randomly drawn, somewhat lower signal levels were required for most observers in order to maintain average percent correct values of approximately 75. For Obs 5, 12, 13, and 14 signal levels of 33, 39 (equal to the high signal level tested in experiment I), 40, and 42 dB SPL, respectively, were tested. Due to the change in signal level, the results from this experiment cannot be directly compared to the results of the other experiments. Instead, the goal was to determine whether the pattern of d's estimated for the virtual conditions was similar to those obtained from experiments I and IIa.

In experiment IVc the stimulus in the first interval was played to the left ear and the stimulus in the second interval was played to the right ear. Note that because a same/different task was used, the dichotic separation is not based on stimulus type, i.e., each ear was as likely to receive a

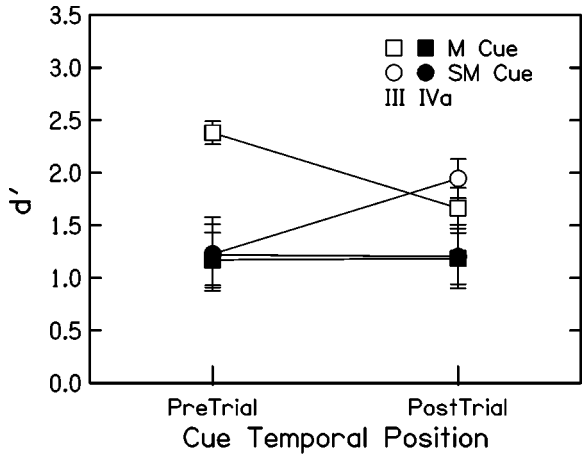


FIG. 6. The d' values averaged across observers for experiment IVa in which the signal frequency was drawn at random (filled) are compared to the results of experiment III (unfilled) for the same observers. In other regards the plot is as Fig. 5.

masker as a signal-plus-masker stimulus. To the degree that a masker-first advantage was obtained in this condition, it would seem unlikely that the effect was associated only with monaural processing.

B. Results and discussion

Figure 6 shows the averaged virtual d's from experiment IVa (filled symbols) in which the signal frequency was chosen at random prior to each trial. The averaged results of experiment III for the same observers are displayed using unfilled symbols. Conditions for virtual pre- and posttrial cue conditions are plotted left and right, respectively. Squares are for the trials in which the masker was the virtual cue and circles are for the trials in which the signal-plus-masker was the virtual cue. Error bars are the standard errors of the mean across four observers.

In experiment IVa, shown with filled symbols, there was no effect of cue type (masker versus signal-plus-masker) and additionally no effect of the temporal position of the virtual cue relative to the virtual yes/no trial. The individual data similarly did not indicate a strong effect of cue type or the ordering of the virtual cue and the virtual trial. Also, not surprisingly given the fact that the signal frequency was random, the virtual d's were lower in this experiment than in experiment III, which used the same signal levels.

These results are consistent with the findings of experiment IIb and demonstrate that even when level cues were available, and when the spectral components were not of equal level (the signal was not the same level as the masker components), signal frequency randomization abolished the masker-first advantage.

Figure 7 shows the results of experiment IVb in which the masker was composed of fixed-frequency tones across all trials. Because the pattern of d's in the virtual conditions was different for different observers, the results for individual observers are plotted in separate panels. The bottom panel plots the group mean results and error bars show the

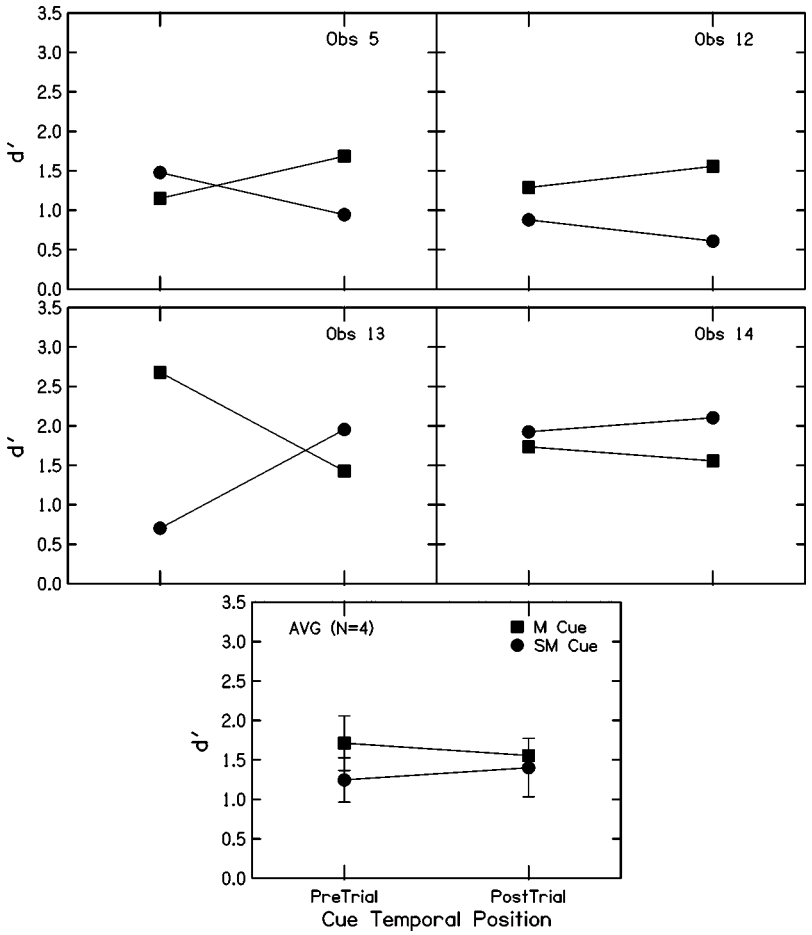


FIG. 7. For each observer, and for the average across observers (bottom panel), virtual d's are plotted for experiment IVb in which both the signal and the masker frequencies were fixed across all trails. In other regards the plot is as Fig. 5.

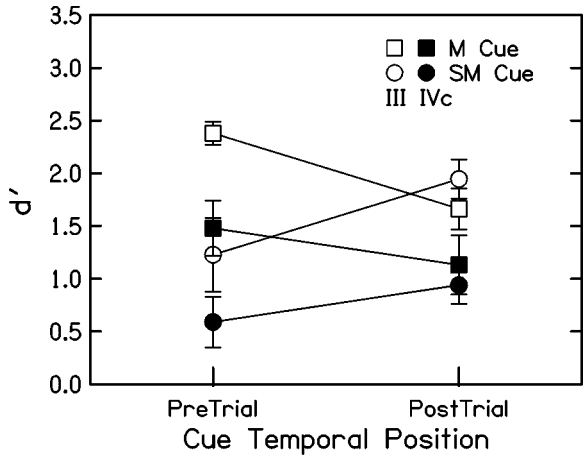


FIG. 8. The d' values averaged across observers for experiment IVc in which the first interval was presented to left ear and the second interval was presented to the right ear (filled symbols). The unfilled symbols are from experiment III for the same observers. In other respects the figure is as Fig. 5.

standard errors of the mean. Recall that in order to maintain comparable d's the signal levels tested in this experiment were generally lower than those tested in the previous experiments. As a result, a comparison of the magnitude of virtual d's across experiments is not meaningful. For the averaged data there was a small advantage for virtual masker pretrial cue compared to the signal-plus-masker pretrial cue. As would be expected based on the results shown in Fig. 7, an ANOVA failed to reveal significant main effects of virtual cue type (masker versus signal-plus-masker) and virtual cue order (pre versus post relative to the virtual yes/no trial). Nor was the interaction term significant. Thus, the masker-first advantage was not reliably present under conditions of minimal uncertainty.

Figure 8 shows the results for experiment IVc in which the stimulus in the first interval was played to the left ear and the stimulus in the second interval was played to the right ear (filled symbols). For the sake of comparison, the averaged results for the same observers in experiment III are shown using unfilled symbols. Error bars are the standard errors of the mean across four observers. For the dichotic experiment (filled symbols) the results for one observer (Obs 14) differed from the pattern apparent in the averaged data; as for the averaged data his d's were generally higher for the masker cue conditions, but were higher in the virtual PostMCue than the virtual PreMCue condition (i.e., an interaction in the opposite direction, with larger differences in d''s when the virtual cue followed the trial). The averaged data reasonably reflect the results of the other three observers.

The averaged data plotted in Fig. 8 clearly show a loss in sensitivity associated with presenting the sequential stimuli dichotically (filled symbols) compared to diotically (unfilled symbols). An ANOVA applied to the dichotic data set indicated a main effect of cue type (masker versus signal-plus-masker), but did not reveal significant effect of temporal position of the cue ($p>0.99$) nor interaction ($p>0.2$). It is unlikely that the absence of an interaction term reflected a floor effect—the d's were measurable for all of the virtual conditions. Rather, it seems that the interaction was not strong enough to overcome individual differences. It is difficult to envision a consistent explanation for a significant difference in cue type without an interaction; thus we suggest the masker-first advantage remained, but was weakened. To the degree that this speculation is accurate, we would suggest that the masker-first advantage does not depend only on peripheral/monotic processing.

To summarize the results of this final sequence of measurements: randomizing the signal frequency abolished the masker-first advantage; a weakened masker-first advantage appeared to remain when sequential stimuli were presented to different ears; and little can be made of the reduction in uncertainty associated with the use of a single masker across many trials—individual differences defy a clear interpretation.

VI. SUMMARY AND GENERAL DISCUSSION

These experiments addressed an asymmetry in the potency of different types of cues in providing a release from informational masking. Informational masking was introduced by randomly choosing a spectrally sparse masker prior to each trial. The results of experiment I corroborated those of Richards and Huang (2003) by demonstrating that a pretrial masker cue provided a greater release from informational masking than did a pretrial signal-plus-masker cue. Averaging across two signal levels and 14 observers, the difference in d' was 1.45. Moreover, the results of experiment I indicated that when a yes/no trial was reliably followed by a posttrial cue, the d's depended only slightly on whether the posttrial cue was the masker or the signal-plus-masker. Again using averaged results, the d' was 0.3 units smaller for the masker than for the signal-plus-masker posttrial cue. These findings suggested that the relative effectiveness of cues in informational masking tasks does not depend just on reductions in uncertainty.

The ''masker-first advantage'' demonstrated in experiment I was largest when the cue was consistent across trials, e.g., when data were collected using a blocked design. Nonetheless the effect was obtained, at least for four of the five observers tested, when there was no consistent ''cue,'' i.e., for a same/different task. This result suggested that the masker-first advantage did not depend only on differences in strategies adopted when different cue types were tested using blocked designs. On the other hand, no masker-first advantage was obtained when the signal frequency was randomly drawn prior to each trial. This result implied that the phenomenon is at least partly due to processing that occurs beyond the auditory periphery, i.e., it is sensitive to the incorporation of *a priori* knowledge about a fixed signal frequency or at least about the perceptual ramifications of such a signal's presence. That a modest masker-first advantage appeared to be present when dichotic presentations were tested is also consistent with the suggestion that peripheral processing was not critical to the masker-first advantage.

The masker-first advantage is an indication of enhanced sensitivity when a yes/no detection trial is preceded by a preview of the masker. This might reflect sensory adaptation—the frequency channels excited by the masker components might be adapted, thereby enhancing the relative strength of a signal tone of another frequency. A related ex-

planation is one in which psychophysical enhancement (Viemeister, 1980; Viemeister and Bacon, 1982) plays a role—the initial masker stimulus reduced the suppression acting on neighboring frequencies, and in particular the signal frequency. However, at least three factors suggest that adaptation effects were not the sole mechanism behind the masker-first advantage. First, the masker-first advantage was absent when the frequency of the signal tone was chosen at random. Second, because the masker was spectrally sparse with components having frequencies that were chosen at random across trials, the signal frequency was the most likely of all frequencies to be present in the stimulus. This suggests that, if anything, adaptation would be strongest at the signal frequency than at masker frequencies. Indeed, it seems more likely that adaptation at the signal frequency might contribute to the relatively small impact of the signal-plus-masker pretrial cue. Third, Viemeister (1980; see also Summerfield *et al.*, 1984) found that psychophysical enhancement did not withstand dichotic presentations. The current findings suggested that the masker-first advantage remained, albeit attenuated, when the stimuli were presented dichotically. Importantly, in this study, the signal was not always in the ear opposite the masker, making comparisons with Viemeister's experiment, which ensured a separation of masker and signal, tenuous.

Next consider two strategies that have been proposed by Durlach and colleagues to account for informational masking, at least as it occurs in some conditions similar to those tested here (Durlach *et al.*, 2003). For one strategy, the ''Listener Max'' strategy, it was hypothesized that observers used the first stimulus of a trial to form an acceptance filter, allowing the energy at the signal frequency to pass, or be maximized, on a subsequent interval. For a second strategy, the ''Listener Min'' strategy, it was hypothesized that observers used the first stimulus of a trial to form a rejection filter so that masker energy was minimized in subsequent intervals. Because a signal cue was not tested in the present experiments, only the Listener Min model is considered here. Imagine that, upon hearing a pretrial cue, the listener formed a rejection filter based on information available from the cue. If the cue is a preview of the masker, a rejection filter matching the cue could aid the observer in detecting the signal. For a signal-plus-masker cue, however, there is little basis for forming a useful rejection filter. As a result, the subsequent signal would be more readily detected when the masker preceded the trial than when the signal-plus-masker preceded the trial. However, this model also predicts that randomizing the signal frequency should have no effect on d' when the masker cue preceded the trial. We know from the data presented above that this prediction is incorrect. Thus, it seems unlikely that a rejection filter model can succeed in accounting for the masker-first advantage without additional stages, such as a mechanism by which observers attend to a known signal frequency, etc.

Finally, consider the effect a masker-first advantage might have on informational masking studies in general. At a minimum the results of the current experiments suggest caution in interpreting data obtained when a pretrial cue is used. Because a ''virtual'' masker-first advantage was observed when a same/different procedure was tested, it seemed possible that biases may also occur in informational masking studies that use 2IFC detection procedure. The parallel stimulus configuration is one in which the same masker is presented across intervals but random maskers are used on different trials. This assumes, as is usually the case, that the signal is as likely to be in the first as in the second interval. The logic is as follows. When the masker is in the first interval, and the signal-plus-masker is in the second interval observers ought to be very accurate in choosing the second interval, thereby leading to a large number of ''interval 2'' responses. When the signal-plus-masker is in the first interval and the masker-alone is in the second interval, observers ought to be less accurate, and thus respond ''interval 1'' at a relatively lower rate (assuming, of course, that they pay attention to the response feedback and incorporate that into their decision strategy). The end result would be that more ''signal in interval 2'' responses would be generated. We examined data on hand from two such 2IFC experiments.[4] One data set, taken from Tang and Richards (2003), included results from 12 observers. Of the twelve, the data for four of the observers indicated no response bias, i.e., the observers responded ''interval 1'' and ''interval 2'' at statistically indistinguishable rates. Of the remaining eight observers, one reliably overresponded ''signal in interval 2'' and seven reliably over-responded ''signal in interval 1'' (based on at least 490 trials per observer, $p<0.05$). In a second unpublished data set, four of the observers from the current study (Obs 5, 12–14) subsequently completed a 2IFC experiment in which the maskers were drawn randomly on every trial but were the same across intervals of a trial. All four exhibited statistically significant response biases; they overresponded ''signal in interval 1'' (at least 360 trials per observer, $p<0.05$). These retrospective analyses indicated that many observers were biased in their responses and the biases tended to be in the same direction across observers. The probability that an observer exhibited a response bias may have depended on past experience (in this case, whether they had participated in the current experiments). More importantly, in light of the current discussion, the *direction* of the response bias was not in the expected direction; observers were *not* more likely to respond ''interval 2'' than ''interval 1.'' Even so, the presence of consistent response biases suggests that analyses of experiments similar to those tested here should be scrutinized for potential sequential effects and/or consistent response biases.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. RO1 DC 02012 and R01 DC04545 from the National Institutes of Health. Associate Editor Dr. Neal Viemeister and two anonymous reviewers provided extraordinarily helpful comments on an earlier draft of this manuscript.

APPENDIX: INDIVIDUAL RESULTS FOR EXPERIMENT I.

TABLE III. The results for experiment I are listed for different signal levels and conditions. The entries are d's. The final two rows show the averages and standard errors of the mean across observers.

	Low signal level				High signal level			
	PreM	PostM	PreSM	PostSM	PreM	PostM	PreSM	PostSM
Obs1	1.35	0.63	1.21	1.59	4.21	1.30	1.33	2.94
Obs2	1.32	0.00	0.00	0.71	4.91	2.44	1.36	3.54
Obs3	1.24	0.78	0.36	0.79	2.92	1.88	1.51	1.54
Obs4	1.44	0.00	0.00	0.00	5.28	0.70	1.50	2.72
Obs5	1.36	0.39	0.00	0.67	2.86	1.72	1.68	1.47
Obs6	1.94	1.59	1.18	1.52	3.65	3.71	2.47	3.30
Obs7	1.29	1.08	0.83	0.43	3.01	2.66	2.21	2.97
Obs8	1.33	1.56	1.19	0.84	4.52	4.09	3.28	2.30
Obs9	1.74	0.80	0.55	1.37	3.45	1.99	1.71	2.57
Obs10	3.38	1.24	0.71	1.63	4.85	1.36	2.27	2.95
Obs11	1.14	0.38	0.00	0.00	3.53	1.08	1.05	1.57
Obs12	0.54	0.00	0.00	0.34	2.42	1.23	0.57	1.72
Obs13	1.74	0.87	0.66	1.97	3.22	1.96	1.58	2.29
Obs14	1.38	1.29	1.17	1.78	2.91	2.49	2.28	1.99
AVG	1.51	0.76	0.56	0.98	3.69	2.04	1.77	2.42
SEM	0.17	0.15	0.13	0.17	0.24	0.26	0.18	0.18

[1]From the observer's perspective, the two intervals might be treated as (1) a cue interval and a yes/no trial, and (2) two intervals containing sounds that are the same or different. If observers adopted the latter interpretation, but feedback was appropriate to the former interpretation, response incompatibilities would be introduced when the conditions change. In our initial work with the PreMCue and PreSMCue conditions, signal-present feedback was used. When faced with higher thresholds in the PreSMCue condition compared to the PreMCue condition, observers were instructed as to the response contingencies, etc. Ultimately, we repeated the experiment using a same/different response structure. The resulting thresholds indicated no effect of feedback contingency. Given the large number of conditions tested in the current experiments, a same/different response/feedback contingency was used. The trials were treated as cued yes/no trials, and the estimate of d' was based on formulas appropriate for yes/no trials.

[2]The minimal number of trials contributing to a single d' is 500. Based on a binomial distribution, this leads to 95% confidence limits about $p=0.5$ of approximately $\pm 5\%$. We chose 55% as an upper limit on guessing, and set d's to zero if the percent correct scores were less than 55. This affected a total of 14 d' scores, nine of which were in experiment I (low signal level). Only one of the scores fell below 45% correct—Obs 2 in the PreSMCue condition of experiment I.

[3]Here typical signal-detection theory assumptions are made, including that uncertainty adds the same independent variance to the internal signal and no-signal distributions regardless of condition and that due to the fixed signal level the differences between the means of the internal signal and no-signal distributions do not depend on condition. Under these assumptions, because the total variance in the task is smallest in the PreMCue condition (largest d'), the effect of uncertainty is expected to be largest in that condition. Likewise, because the d' is smallest in the PreSMCue condition, d''s are expected to shift the least in that condition.

[4]The trials came from experiments in which the signal was altered using an adaptive algorithm. After removing trials within two steps of the lowest and highest signal levels encountered, the proportions of "1" and "2" responses were determined for the trials in which the signal level was in the lower $\frac{1}{3}$ of the remaining signal levels. The percent correct scores for those trials were approximately 70% correct. High signal levels were not included because high percent correct scores are incompatible with response biases.

Bregman, A. S. (**2003**). "Issues in the use of acoustic cues for auditory scene analysis," J. Acoust. Soc. Am. **113**, 2231(A).

Durlach, N. I., Mason, C. R., Kidd, Jr., G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking (L)," J. Acoust. Soc. Am. **113**, 2984–2987.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge U. P., Cambridge).

Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 408–415.

Richards, V. M., and Huang, R. (**2003**). "Criterion placement in staircase procedures," J. Acoust. Soc. Am. **113**, 2287(A).

Richards, V. M., and Neff, D. L. (**2004**). "Cuing effects for informational masking," J. Acoust. Soc. Am. **115**, 289–300.

Summerfield, A. Q., Haggard, M. P., Foster, J. R., and Gray, S. (**1984**). "Perceiving vowels from uniform spectral: Phonetic exploration of an auditory aftereffect," Percept. Psychophys. **35**, 203–213.

Tang, Z., and Richards, V. M. (**2003**). "Estimation of a linear model in an informational masking study," J. Acoust. Soc. Am. **114**, 361–367.

Viemeister, N. F. (**1980**). "Adaptation of masking," in *Psychophysical, Physiological, and Behavioral Studies in Hearing*, edited by G. van den Brink and F. A. Bilson (Delft U. P., Delft, The Netherlands), pp. 190–199.

Viemeister, N. F., and Bacon, S. P. (**1982**). "Forward masking by enhanced components in harmonic complexes," J. Acoust. Soc. Am. **71**, 1502–1507.

Watson, C. S. (**1987**). "Uncertainty, informational masking, and the capacity of immediate auditory memory," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ).

The effect of onset asynchrony on profile analysis by normal-hearing and hearing-impaired listeners

Jennifer J. Lentz[a)]
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

Marjorie R. Leek and Michelle R. Molis
Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC 20307

(Received 17 September 2003; revised 30 June 2004; accepted 9 July 2004)

The effect of onset asynchrony on discrimination of spectral shape was evaluated for hearing-impaired and normal-hearing listeners. Stimuli were the sum of four tones equally spaced on a logarithmic frequency scale. The standard stimulus had tones of equal amplitude, and the signal stimulus had two adjacent components increased in level, and the other two components decreased in level. Thresholds for discrimination between the standard and signal stimuli were measured as a function of an onset asynchrony among the components of 0, 50, and 200 ms. Hearing-impaired and normal-hearing listeners had similar thresholds when the stimulus components were widely spaced in frequency, but hearing-impaired listeners had much higher thresholds for narrowly spaced components. Excitation pattern analyses indicated that listeners may use spectral peaks in the stimulus rather than the change in excitation across the full stimulus bandwidth for spectral shape discrimination tasks. Increasing temporal asynchrony of components resulted in increased thresholds for both groups of listeners to a greater extent in the wide span than the narrow span. Reduced effects of onset asynchrony in the narrow span suggest that spectral resolvability of components plays an important role in the processing of onset asynchrony across frequency. [DOI: 10.1121/1.1787125]

PACS numbers: 43.66.Sr, 43.66.Fe, 43.66.Jh [GK] Pages: 2289–2297

I. INTRODUCTION

Listeners with hearing loss are known to have a great deal of difficulty communicating in noisy environments. When multiple sounds are present, such as speech and noise, the ability to determine which frequency components belong to each sound source can be of particular importance for effective communication. Hearing loss may detrimentally affect this process, resulting in an impaired ability to understand speech in noise. Because temporal asynchrony is an important cue to accurate grouping of frequency components (Darwin and Carlyon, 1995), a deficit in temporal processing in hearing-impaired listeners may lead to abnormal grouping of frequency components, or the opposite, segregation of frequency components. In the current study, a spectral shape discrimination paradigm is used to determine whether component onset asynchrony is equally disruptive to listeners with normal hearing and listeners with hearing loss.

Spectral shape discrimination, an important precursor to speech perception, has been studied extensively in normal-hearing listeners. Green and his colleagues developed a controlled paradigm, called profile analysis, to study the ability to discriminate between sounds with different spectral shapes (Spiegel *et al.*, 1981; Green, 1988). A typical profile-analysis experiment employs stimuli consisting of the sum of tones, and listeners detect a change in the level of one or more of the component tones. When the number of stimulus components is small, sensitivity improves with increasing numbers of tonal components. Once individual components begin to interact within a single auditory filter, further increases in the number of stimulus components can cause thresholds to rise (Bernstein and Green, 1987). The frequency selectivity of the auditory system dictates the number of components at which thresholds will begin to rise.

Listeners with hearing loss, who have reduced frequency selectivity compared to normal-hearing listeners (Glasberg and Moore, 1986; Leek and Summers, 1993b), also maintain some of the ability to discriminate between sounds with different spectral shapes. Lentz and Leek (2002) asked normal-hearing and hearing-impaired listeners to discriminate between a standard stimulus consisting of five widely spaced, equal-amplitude components and a signal stimulus in which the central component of the standard was increased in level. Measured thresholds were similar for normal-hearing and hearing-impaired listeners. In a later paper, Lentz and Leek (2003) used a slightly different stimulus construction, in which the standard stimulus was the sum of six widely spaced, equal-amplitude components, and a signal stimulus was generated by increasing the level of half of the standard components and decreasing the level of the other half. Hearing-impaired listeners again had thresholds comparable to normal-hearing listeners.

A different pattern of results for hearing-impaired listeners emerges, however, when stimuli are spectrally dense. Here, reduced frequency selectivity causes a ''smearing'' of the spectral representation of the stimulus (Leek *et al.*, 1987), leading to elevated thresholds compared to normal-hearing listeners having better frequency selectivity (Leek and Summers, 1996). Summers and Leek (1994) investigated the ability of listeners to detect a change in spectral shape

a)Electronic mail: jjlentz@indiana.edu

that produced a spectral ripple in a 121-component stimulus. The hearing-impaired listeners' reduction in frequency selectivity adequately accounted for elevated spectral shape discrimination thresholds. The poorer frequency selectivity apparently smeared the internal representation of the spectral ripple and made discrimination between the ''rippled'' signal stimulus and a standard with a ''flat'' spectral shape more difficult.

In addition to good frequency selectivity, the auditory system must also have an ability to compare and contrast the levels of individual frequency components within a stimulus. A requirement for good spectral shape discrimination is that frequency components should be ''grouped'' together in a single auditory object. This ''grouping'' allows for simultaneous comparison of component levels, a process that leads to lower thresholds than sequential comparisons of components' levels (Green *et al.*, 1983). One prerequisite, then, is that all information-bearing stimulus components should be temporally synchronous. Asynchrony of information-bearing components can substantially degrade spectral shape discrimination ability. In one experiment, Green and Dai (1992) had listeners detect an intensity increment in one tone (the signal component) of a multitonal stimulus (the reference components). They measured the lowest thresholds when the signal and reference components shared onsets and offsets. Thresholds increased by about 20 dB when the signal component onset led the reference components by 500 ms and shared offsets with the reference components. Hill and Bailey (1997) reported increases in threshold of 6-16 dB when the signal component led the reference components by 320 ms. Dai and Green (1992) included a more complex profile-analysis stimulus in which the spectral shape change was distributed across all stimulus components. Components were either all simultaneous, or, in an asynchronous condition, the offsets of the even-numbered components coincided with the onsets of the odd-numbered components. Thresholds were approximately 15 dB higher in the asynchronous condition. In all these cases, higher thresholds were reported when components were asynchronous.

Although it has been shown that temporally offsetting information-bearing components interferes with the spectral shape discrimination process for normal-hearing listeners, the effects of onset asynchrony on spectral shape discrimination have not been clearly determined in listeners with hearing loss. Cochlear changes present in the impaired auditory system may impoverish the internal spectral and temporal representation of a sound (Moore, 1995), potentially altering the effects of onset synchrony on the ability to detect changes in spectral shape. The effects of onset asynchrony may be different in listeners with hearing loss, who have been shown to have poorer temporal processing than listeners with normal hearing. Listeners with hearing loss experience a much slower-than-normal rate of recovery from forward masking (Glasberg *et al.*, 1987), and gap detection experiments indicate poorer temporal processing abilities when stimuli are presented at the same sound pressure level (Fitzgibbons and Wightman, 1982). These differences in temporal processing may alter the degree to which the impaired auditory system is affected by asynchrony of stimulus components.

It is not known whether the effects of onset asynchrony on spectral shape discrimination may differ between normal-hearing and hearing-impaired listeners, but Grose and Hall (1996a) have described the influence of onset asynchrony in a comodulation-masking-release (CMR) study. In Grose and Hall's experiment, listeners detected a tone added to a band of noise (the target band) and additional bands of noise were placed distant in frequency from the tone (the flanking bands). The flanking bands were modulated either coherently or incoherently with the target band. Thresholds were lowest when the distant bands were synchronous and the target and flanking bands were coherently modulated. Thresholds increased as the flanking bands became more asynchronous with the target band. The elevation in threshold with increasing onset asynchrony was similar for normal-hearing and hearing-impaired listeners. This CMR study suggests that asynchrony of components is processed similarly by normal-hearing and hearing-impaired listeners. It should be noted, however, that ''true'' CMR requires integration of envelopes across frequency, a process that may occur at a level higher in the auditory pathway than binaural integration (Schooneveldt and Moore, 1989). For tasks that require lower-level integration, such as profile analysis (Green and Kidd, 1983), the effects of onset asynchrony may differ. Grose and Hall (1996b) have shown that when listeners detect a temporal gap between two tones having disparate frequencies, deficiencies in spectro-temporal processing by hearing-impaired listeners are observed. Therefore, extrapolation of onset effects from CMR to profile analysis may not be straightforward.

The current experiment tests whether onset asynchrony has similar effects on the discrimination of spectral shape by normal-hearing and by hearing-impaired listeners. Stimuli consisted of the sum of tones spaced across a frequency range, and listeners detected a spectral shape change between signal and standard stimuli. In one set of conditions, spectral components were widely separated in frequency. These conditions were intended to extend the findings of Grose and Hall (1996a) to spectral shape discrimination in a profile-analysis task. In a second set of conditions, components were more narrowly spaced in frequency. These conditions were used to determine the role that spectral resolution has on onset asynchrony effects. Because hearing-impaired listeners have reduced frequency selectivity, the spectral shape to be detected was limited to one change in level across the spectrum (a ''one-step'' change) so that the internal spectral contrast available to hearing-impaired listeners would likely be reduced relative to that of normal-hearing listeners by similar amounts for all conditions tested. Any difference in performance due to differences in frequency selectivity between normal-hearing and hearing-impaired listeners should be similar across all stimulus configurations, and the effects of onset asynchrony could be evaluated independently of frequency selectivity. Differences in the amount of change in threshold with increasing onset asynchrony would indicate alterations in temporal processing and the effect that temporal processing has on the ability to group frequency compo-

TABLE I. Audiometric thresholds of test ears for hearing-impaired listeners (HI1–HI4; dB HL *re*: ANSI 1989).

			Frequency (Hz)						
Observer	Age	Test ear	250	500	1000	2000	3000	4000	8000
HI1	60	R	45	40	40	45	45	50	80
HI2	69	L	30	30	30	30	35	30	40
HI3	74	R	50	40	55	60	65	70	70
HI4	75	R	30	35	40	70	70	65	65

nents together. Three different frequency configurations also were tested to evaluate the relative effects of different frequency regions in hearing-impaired listeners.

II. METHODS

A. Observer characteristics

Four normal-hearing listeners (NH1–4), ranging in age from 30 to 56 years, with a mean of 41 years, and four hearing-impaired listeners (HI1–4), ranging from 60 to 75 years with a mean age of 69 years, participated. Normal-hearing listeners had pure-tone audiometric thresholds no greater than 20 dB HL (ANSI, 1989) between 250 and 8000 Hz. Hearing-impaired listeners were selected to have pure-tone audiometric thresholds greater than or equal to 30 dB HL at all octave test frequencies. Hearing losses were moderate and bilateral; the site of lesion was presumed to be cochlear based on air- and bone-conduction thresholds and normal immitance audiometry. For normal-hearing listeners, the right ear was tested for three listeners and the left ear for one listener, as she had a threshold of 30 dB HL at 6000 Hz in her right ear. In hearing-impaired listeners, if both ears had hearing loss appropriate for this study, the better ear was selected for testing. The audiometric configurations of the test ears for the hearing-impaired listeners are reported in Table I. All observers enrolled in the study voluntarily agreed to participate and gave written informed consent. All had previously participated in psychoacoustic studies.

B. Stimuli

The stimuli were the sum of four tones, spaced equally on a logarithmic frequency scale. Phases of each tone were selected randomly from a uniform distribution ranging from 0 to 2π rad. The standard stimulus was the sum of four equal-amplitude tones, and the signal stimulus had two ''up'' components, generated by increasing the level of two components of the standard, and two ''down'' components, generated by decreasing the level of two components of the standard. The up components were increased by the same decibel amount as the down components were decreased ($\Delta L_{\mathrm{up}} = \Delta L_{\mathrm{down}}$; ''balanced'' stimuli as described by Durlach *et al.*, 1986). The signal strength is described as the change in amplitude of the up components relative to the mean amplitude of the standard components (i.e., $20\log(\Delta A/A)$; sig *re*: stan, in dB).

Two frequency spans were tested. For the wide span, four component frequencies were selected from a possible six frequencies, equispaced between 200 and 4000 Hz on a logarithmic frequency scale (i.e., 200, 360, 660, 1210, 2200, and 4000 Hz). For the narrow frequency span, components were spaced between 1000 and 4200 Hz (1000, 1330, 1780, 2370, 3150, and 4200 Hz). The actual frequencies present for each stimulus depended on the increment shape tested, as described below.

Three increment shapes were tested, named according to which frequencies within a stimulus were incremented: high, low, and flank. For the high increment shape, the stimulus was made using the four highest frequencies of the six possible frequencies (e.g., 660, 1210, 2200, and 4000 Hz for the wide span). Within the stimulus, the up and down components were the two highest frequencies (2200 and 4000 Hz) and the two lowest frequencies (660 and 1210 Hz), respectively. For the low increment shape, the stimulus was the sum of the four lowest of the six possible frequencies (e.g., 200, 360, 660, and 1210 Hz for the wide span). The two lowest frequencies (200 and 360 Hz) were the up components, and the two highest frequencies (660 and 1210 Hz) were the down components. The flank stimulus was made using the four central frequencies of the original six (e.g., 360, 660, 1210, and 2200 Hz). Within the stimulus, the up components were the lowest and highest frequency components (360 and 2200 Hz), and the down components were the middle components (660 and 1210 Hz). Schematic power spectra of the signal stimuli for the six possible stimulus configurations are plotted in Fig. 1. It should be noted that the frequencies of the down components remained constant for all stimuli within each frequency span.

For all stimulus configurations, the down components began earlier in time than the up components by 0, 50, or 200 ms, generating an onset asynchrony between the down and up components, with all components sharing common offsets. Note that all standard and signal stimuli within a block of trials were generated with the same onset asynchrony. Thresholds for the just-detectable change in spectral

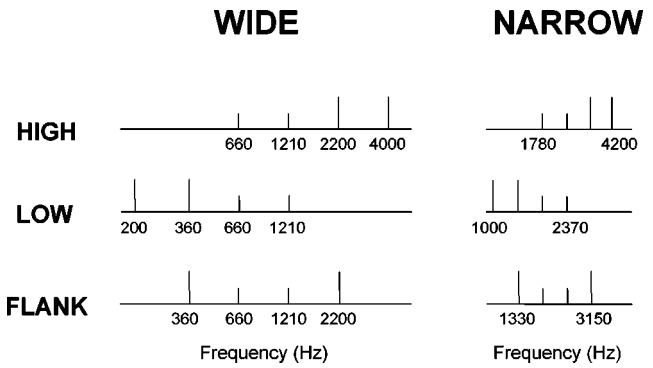


FIG. 1. Power spectra representing the signal stimuli for six different stimulus configurations are shown.

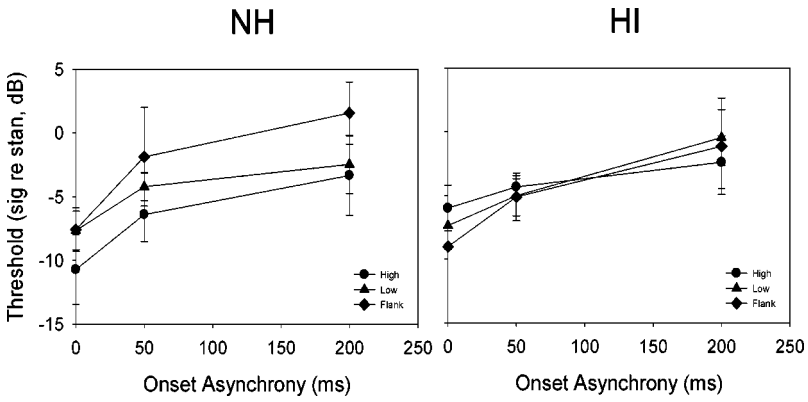


FIG. 2. Spectral shape discrimination thresholds are plotted as a function of onset asynchrony for the wide span. High, low, and flank increment shapes are indicated by the circles, triangles, and diamonds, respectively. Error bars represent standard errors of the mean across four listeners.

shape were measured as a function of the onset asynchrony between the up and down components for the six different stimulus configurations. To discourage the use of changes in intensity as a cue to the signal stimulus, overall stimulus levels were chosen randomly over a uniformly distributed ± 6-dB range with a 0.1-dB gradation.

The stimuli were generated digitally at a sampling rate of 24 414 Hz (4096×10^{-5} s) and played using a 24-bit digital-to-analog converter (Tucker-Davis Technologies; TDT RP2). The resulting stimuli were fed into a programmable attenuator (TDT PA4), a headphone buffer (HB5), and then into one insert earphone (ER-2). An antialiasing filter was not used, as the TDT digital-to-analog converter does not produce aliased stimuli below the sampling rate. The mean level per component was 85 dB SPL. The up components always had a duration of 200 ms, and the down components, which began prior to the up components, had durations of either 200, 250, or 400 ms, corresponding to onset asynchronies of 0, 50, and 200 ms. Thus, across all onset asynchronies, the down and up components overlapped in time for the final 200 ms of the stimulus. Stimuli had 20-ms cosine-squared rise/fall times.

C. Procedure

A standard/two-alternative forced-choice task was used to estimate thresholds, with trial-by-trial signal levels chosen according to a 3-down, 1-up adaptive tracking procedure to estimate 79%-correct discriminations (Levitt, 1971). Observers were seated in a sound-attenuating room and heard the three sounds separated by about 1 s. Interstimulus intervals (ISIs) differed slightly for the different onset asynchronies of the stimuli: the 0-, 50-, and 200-ms onset asynchronies had ISIs of 0.8, 0.95, and 1 s, respectively. The standard was always presented in the first interval, and either the signal or standard was presented in the second and third interval, determined randomly with equal likelihood. Listeners indicated which interval contained the signal stimulus via a button box. Correct answer feedback was provided to the listener following each trial.

At the beginning of every track, the signal strength was set to 10 or 15 dB above an estimate of the listener's final threshold. The initial step size of the tracking procedure was 4 dB, and after three reversals the step size was reduced to 2 dB. The track continued until a total of eight reversals of the direction of the track was obtained. The mean of the signal strengths at the last four reversal points was taken as threshold.

Data were collected using a pseudorandomized block design. There were 18 conditions (2 span types $\times$ 3 increment shapes $\times$ 3 onset asynchronies), and each was presented at least four times. First, one of the two span types (narrow or wide spacing) was selected randomly. For that span type, the three increment shapes were tested in random order, with all three onset asynchronies presented in random order for each shape before a different increment shape was tested. After these nine experimental conditions were tested, the same procedure was followed for the remaining span type (with a different random ordering of conditions). After all 18 conditions were tested, the order of conditions was again randomized three times until at least four threshold replicates for each condition were obtained.

Sometimes it was necessary to measure more than four thresholds because of evidence of learning effects. In order to assess the effects of initial learning, paired t-tests were carried out between the thresholds from the first and second replicates for a given condition. Using a conservative estimate of practice effects, a probability of an overall difference in thresholds between these two blocks less than 0.1 required that a fifth block of all conditions be run. Two of the normal-hearing listeners and three of the hearing-impaired listeners ran five blocks total for each condition. Data were the average of the last four threshold estimates for each of the 18 stimulus conditions.

III. RESULTS

Figure 2 plots profile-analysis thresholds obtained from the wide span as a function of onset asynchrony. The left panel displays data obtained from the normal-hearing listeners, and the right panel displays data obtained from the hearing-impaired listeners. Thresholds for hearing-impaired and normal-hearing listeners are similar at the 0-ms onset asynchrony, as anticipated. This suggests that neither the hearing loss nor the wide difference in age between the two groups of listeners resulted in reduction in the ability to discriminate spectral shapes in these conditions. Data for both groups of listeners show a progressive increase in threshold as the onset asynchrony increases. Across the different increment shapes, thresholds are between -6 and -11 dB in the synchronous case (onset asynchrony $= 0$ ms) and are elevated

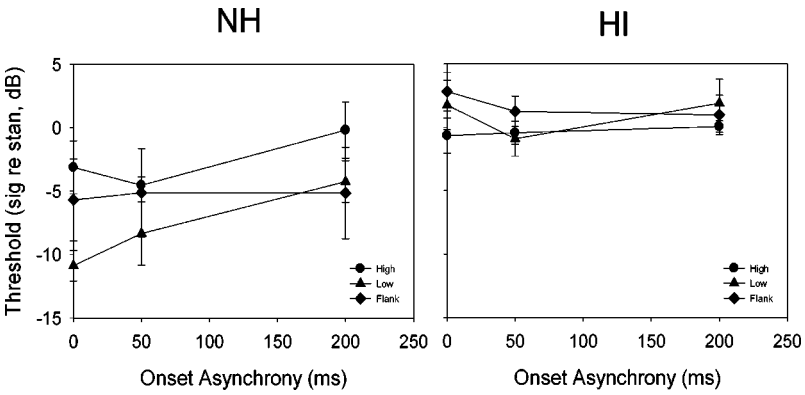


FIG. 3. As in Fig. 2, spectral shape discrimination thresholds are plotted as a function of onset asynchrony for the narrow span. Error bars represent standard errors of the mean across four listeners.

by 3–10 dB when the onset asynchrony is 200 ms. Collapsed across the different onset asynchronies and the increment shapes, normal-hearing and hearing-impaired listeners have very similar average thresholds of -4.7 and -4.5 dB, respectively. Additionally, thresholds vary little across the three increment shapes, and all show similar increases in threshold with increasing onset asynchrony. In the normal-hearing data, the flank conditions appear to lead to higher thresholds, but, as will be seen in the statistical report, there was no significant effect of increment shape.

Figure 3 plots mean profile-analysis thresholds obtained from the narrow frequency span. Again, left and right panels display data obtained from normal-hearing and hearing-impaired listeners, respectively. At the 0-ms onset asynchrony, hearing-impaired listeners have higher thresholds than normal-hearing listeners. Normal-hearing listeners show an increase in threshold with increasing onset asynchrony for the low increment shape and a much smaller increase in threshold for the high increment shape. The flank shape shows almost no effect of onset asynchrony on the thresholds. Thresholds for the hearing-impaired listeners do not change with increases in onset asynchrony, and remain constant at an average level of 0.8 dB for all increment shapes.

Average thresholds across subjects, onset asynchronies, and increment types remain similar for the normal-hearing listeners for the two span types (-4.7 dB for the wide span and -5.3 dB for the narrow span) and increase by 5.3 dB for the hearing-impaired listeners (-4.5 dB for the wide span and 0.8 dB for the narrow span). Thresholds of hearing-impaired listeners in the narrow frequency span are elevated with respect to the normal-hearing data and their own thresholds in the wide span. A repeated-measures ANOVA was carried out with frequency span, increment shape, and onset asynchrony as repeated measures and group membership as a between-subjects measure. Neither group membership, frequency span, nor increment shape produced a significant main effect ($p>0.07$), but onset asynchrony significantly increased threshold levels [$F(2,12)=15.8; p<0.001$]. Collapsed across observers, spans, and increment shapes, average thresholds are -5.4 dB at 0 ms, -3.7 dB at 50 ms, and -1.2 dB at 200 ms. There was a significant interaction of group membership and frequency span [$F(1,6)=6.5; p<0.05$], with larger differences due to frequency span for the hearing-impaired listeners than for those with normal hearing. A further interaction between frequency span and onset asynchrony [$F(2,12)=12.3; p<0.001$] is illustrated in Fig. 4, which shows the amount of threshold change between the synchronous condition and the 200-ms onset asynchrony condition collapsed across increment shape. The change in threshold with increasing onset asynchrony is largest in the wide span, with normal-hearing listeners showing changes in threshold of 7.2 dB, and hearing-impaired listeners having changes in threshold of 6.1 dB. The effect of onset asynchrony is greatly reduced for the narrow span. For normal-hearing listeners, the change in threshold is 3.4 dB, and for hearing-impaired listeners the change in threshold is -0.3 dB. The decrease in the effects of onset asynchrony is due both to higher thresholds for the 0-ms conditions and lower thresholds for the 200-ms conditions. This indicates that the reduced spectral contrast in the narrow condition compensates in some way for the detrimental influence of onset asynchrony on spectral shape discrimination for both groups of listeners.

IV. DISCUSSION

When stimulus components were widely spaced, normal-hearing and hearing-impaired listeners had similar

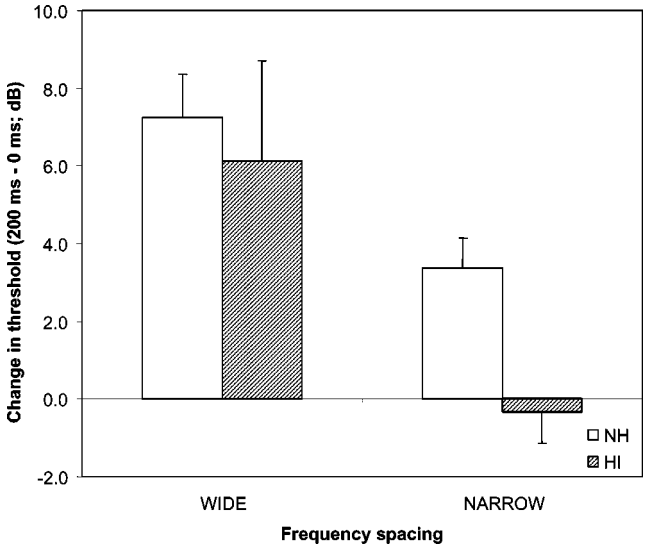


FIG. 4. The change in threshold between the 200-ms and 0-ms onset asynchrony, collapsed across increment types, is plotted for the wide and narrow frequency span. Normal-hearing and hearing-impaired data are shown as unfilled and hatched bars, respectively. Error bars indicate standard errors of the mean across the four listeners.

spectral shape discrimination thresholds, and the detrimental effects of onset asynchrony of stimulus components were similar for the two groups. However, narrowly spaced components led to a different result: for hearing-impaired listeners, spectral shape discrimination thresholds were much higher than in the wide span but this was not the case for the normal-hearing listeners. Also for the narrow span, onset asynchrony was not as detrimental to the spectral shape discrimination process for both groups of listeners. The closer spacing of components apparently leads to poorer spectral shape discrimination for hearing-impaired listeners as well as a reduction in the effects of onset asynchrony for both groups.

The thresholds for 0-ms onset asynchrony may serve to illuminate the role that frequency selectivity may have on spectral shape discrimination thresholds across groups and conditions. In the wide conditions at 0 ms, normal-hearing listeners have only slightly better thresholds than the hearing-impaired listeners (-8.7 dB vs -7.5 dB, respectively). In the narrow conditions, normal-hearing thresholds increase slightly to -6.6 dB, and hearing-impaired thresholds increase by a much greater amount to $+1.3$ dB. The small change in threshold for the normal-hearing listeners and the large change in threshold for the hearing-impaired listeners may be, in part, due to smearing of changes in the power spectrum.

To determine whether spectral smearing due to imperfect frequency resolution can predict the threshold differences, excitation patterns were constructed and the spectral contrast available to hearing-impaired listeners was compared to normal-hearing listeners. First, signal stimuli (generated at a signal level of -5 dB, 80 dB SPL per component) and standard stimuli were passed through 201 auditory filters (described by Glasberg and Moore, 1990 for normal-hearing listeners) with center frequencies ranging between 126 and 8000 Hz.[1] Second, the power at the output of each auditory filter was computed. As suggested by Glasberg and Moore, excitation levels below the minimum audible pressure (MAP) were set to the corresponding MAP value. A second set of excitation patterns was generated to simulate a flat hearing loss of 40 dB HL. Broader auditory filters for the hearing-impaired listeners were simulated by scaling the p values (the parameters that control the slopes of the filter skirts) of the roex auditory filters by 0.5, producing filters with double the normal bandwidths. For these patterns, excitation levels below MAP thresholds+40 dB were set to the corresponding MAP+40 dB value. A measure of spectral contrast was determined by summing the squared difference between signal and standard excitation patterns for filter center frequencies ranging between 0.8 times the lowest stimulus-component frequency and 1.2 times the highest component frequency.[2]

For all conditions tested, spectral contrast was reduced for the hearing-impaired listeners. For the narrow conditions, the average spectral contrast ratios (HI contrast/NH contrast) collapsed across increment type was 0.70. This metric predicts that hearing-impaired listeners should have higher spectral shape discrimination thresholds in the narrow span than normal-hearing listeners, a result observed in the data (1.3 dB for HI vs -7.5 dB for NH at 0 ms). In the wide span, averaged across all increment types, the ratio of HI spectral contrast to NH spectral contrast is 0.66. This spectral contrast ratio is similar to the spectral contrast ratio in the narrow span and also leads to the prediction that hearing-impaired listeners should have higher spectral shape discrimination thresholds than the normal-hearing listeners. However, hearing-impaired and normal-hearing listeners have similar thresholds for the 0-ms (synchronous) onset asynchrony (-7.5 and -8.7 dB, respectively), suggesting that either the reduction in available spectral contrast does not greatly influence the thresholds in the wide span (i.e., that even a reduced contrast for the hearing-impaired leaves sufficient information to support the discrimination) or that the spectral contrast metric does not adequately describe the spectral cues used by the listeners. All hearing-impaired listeners had audiometric thresholds greater than 30 dB HL at all stimulus frequencies, so no frequency components were processed by regions of normal hearing. Yet, listeners with hearing loss still performed as well as normal-hearing listeners on the spectral shape discrimination conditions with widely spaced components.

Both listener groups also have greater spectral contrast in the wide frequency span when compared to the narrow frequency span. The greater spectral contrast in the wide span is largely due to more filters cueing the difference between standard and signal stimuli. Normal-hearing listeners had ratios (narrow contrast/wide contrast) of 0.54, on average. Hearing-impaired listeners also had an average ratio of 0.57, suggesting that all listeners should have higher thresholds in the narrow span compared to the wide span for the 0-ms conditions. While hearing-impaired listeners do have higher thresholds in the narrow versus the wide conditions (-7.5 dB vs 1.3 dB, respectively), normal-hearing listeners have only a modest increase in threshold between the wide and narrow spans at 0 ms (-8.7 dB vs -6.6 dB).

The result that the spectral contrast metric does not adequately describe performance may be related to results presented by Lentz *et al.* (1999) in which auditory filter bandwidths were fit to notched-noise and profile-analysis data using a model assuming integration of all information-bearing channels (Durlach *et al.*, 1986). Their results indicated that the same auditory filter bandwidths could not jointly account for notched-noise and profile-analysis data obtained in the same listeners. The auditory filter bandwidths needed to predict the profile-analysis data were at least two times narrower than those needed to predict the notched-noise data. In one interpretation of the profile-analysis results, an increase in the number of channels contributing to the discrimination decision could account for the decrease in threshold with increasing number of components (for small numbers of components). This interpretation is contingent on an auditory system that is capable of using only peaks in the power spectrum to make discrimination decisions. The auditory system has been shown to be more sensitive to changes in spectral peaks than in valleys (Moore *et al.*, 1989). The use of spectral peaks and not valleys also forms the basis for certain models of vowel perception (Klatt, 1982; Sidwell and Summerfield, 1985). A healthy auditory system with an ac-

tive suppressive mechanism may have the capability of enhancing the spectral contrast present in stimuli (Leek and Summers, 1993a), thus making the peaks more perceptually salient. This could account for the narrow auditory filters obtained in Lentz *et al.*'s profile-analysis experiment and also would account for the failure of the spectral contrast metric used here to describe the performance of the two groups of listeners across conditions.

Relating this idea to the current study, the number of spectral peaks in the excitation pattern remains the same between the wide and the narrow conditions for the normal-hearing data. Because each spectral component largely excites a single auditory filter, the excitation pattern has four distinct peaks—one for each of the spectral components. The hearing-impaired listeners, who may have auditory filters 2–3 times wider than normal (Leek and Summers, 1993b), have an excitation pattern that has broad spectral peaks in the wide conditions, but in the narrow frequency span, the excitation pattern is greatly smoothed. The distinct spectral peaks that are prominent in the normal ear's excitation pattern (and in the impaired ear's excitation pattern in the wide span) are gone. Thus, the standard and signal stimuli have excitation patterns without broad spectral peaks. If the auditory system uses the change in excitation only at the spectral peaks (and not the whole excitation pattern) to accomplish spectral shape discrimination, the ability of the impaired auditory system to do this is compromised by the smoothing of those features.

An alternative explanation is related to experiments employed by Doherty and Lutfi (1996, 1999) and Lentz and Leek (2003). Their tasks tested the abilities of normal-hearing and hearing-impaired listeners to discriminate a change in level of one or more tones of a multitone stimulus and measured the relative importance of each stimulus component to the discrimination decision. Doherty and Lutfi (1996, 1999) showed that hearing-impaired listeners had a different pattern of weights across frequency than normal-hearing listeners for intensity-discrimination tasks. Lentz and Leek (2003) also showed small differences in the weights adopted by hearing-impaired listeners for profile-analysis tasks. It is not clear whether overlap of excitation between stimulus components would lead to substantially different weights adopted by hearing-impaired listeners. These differences could lead to a decrease in performance and may impact thresholds to a greater degree in the narrow span because reduced frequency selectivity could magnify inefficient weighting schemes.

The high thresholds obtained by the hearing-impaired listeners in the narrow-span conditions indicate that listeners could have been using intensity cues to make the discrimination—a possibility that is meant to be defeated by the application of a roving level. If listeners based their decisions solely on the change in intensity of a *single* component from signal to standard, it is predicted that using a 3-down, 1-up tracking procedure with a 12-dB range of rove would yield thresholds of -3.9 dB (sig *re*: stan, dB; see Green, 1988). Thresholds at 0 ms obtained from the normal-hearing listeners are all below this limit, and therefore it can be concluded that listeners were not using the change in level of a single component to discriminate between signal and standard stimuli. However, thresholds for hearing-impaired listeners for the narrow span at an onset disparity of 0 ms exceed the level detection limit that would be predicted if listeners used only intensity cues. Thus, the change in processing that led to the increase in thresholds may have forced hearing-impaired listeners to use changing level as a cue to discriminate stimuli. Note that other intensity-based decisions are possible, such as using the change in level of two components or using the change in level of the entire stimulus. The analysis provided above indicates a minimum limit for the use of intensity-based cues. Other intensity-based strategies could lead to an even higher predicted limit.

Along with the comparison of absolute thresholds between the two groups of listeners, it is also of interest to note how profile-analysis thresholds changed with differences in onset asynchrony, evident in Figs. 2 and 3. In the wide span, onset asynchrony had a similar effect on thresholds for both groups of listeners. The threshold changes between 0- and 200-ms onset asynchronies (7.2 dB for normal-hearing listeners and 6.1 dB for listeners with hearing loss) are on the order of those reported by Hill and Bailey (1997). Green and Dai (1992), however, reported somewhat larger effects of onset asynchrony, although for different stimulus configurations than used here and in Hill and Bailey's study. Indications that onset asynchrony has little effect on profile analysis for sparsely spaced stimuli is consistent with the findings of Grose and Hall (1996a), who reported that thresholds for a CMR task increased by similar amounts as the onset asynchrony between signal and flanking bands increased. Even though CMR requires integration at a higher level in the auditory system than profile analysis, onset asynchrony affects performance on both tasks similarly.

Both groups of listeners indicate a somewhat different pattern of data in the narrow span. The change in threshold with increasing onset asynchrony decreased for both groups of listeners (only 3.4 dB for normal-hearing listeners and -0.3dB for hearing-impaired listeners; see Fig. 4). Normal-hearing listeners do not have thresholds above the level detection limit of -3.9dB, and therefore they could not have been using level cues to discriminate between the sounds. The smaller effect of onset asynchrony, then, must be related to the fact that the stimulus components are more narrowly spaced and may reflect within-channel processing of the onset asynchrony. Auditory grouping may require that onset differences be processed in independent channels. If two components with different onsets are processed by the same channel, the auditory system may group those components together despite the onset difference. Therefore, onset asynchrony will be less detrimental to the spectral shape discrimination process as two components processed by the same channel are grouped together instead of having separate auditory representations. Hearing-impaired listeners, who have reduced frequency selectivity, are even less susceptible to the effects of onset asynchrony for this type of task. The result that hearing-impaired listeners had reduced effects of onset asynchrony in the narrow span and not the wide span suggests that temporal resolution differences are less of a factor in the processing of onset asynchrony than reduced fre-

quency selectivity. It is likely that within-channel smearing of onset differences would pose a problem for listeners with hearing loss on tasks in which onset differences are helpful in the segregation of sounds, such as listening to speech in background noise.

V. SUMMARY AND CONCLUSIONS

The effect of onset asynchrony in profile-analysis tasks was evaluated for hearing-impaired and normal-hearing listeners. For components spanning a wide frequency range, thresholds were similar for normal-hearing and hearing-impaired listeners at the 0-ms onset asynchrony, in agreement with past data that suggest hearing loss has little effect on the ability to discriminate sounds with different spectral shapes. A different pattern of data emerged for the narrow frequency span, in which the hearing-impaired thresholds were elevated by more than 5 dB relative to the wide span, while the normal-hearing thresholds were similar for the wide and narrow spans. These results suggest that hearing loss can compromise the ability to compare and contrast levels across different frequency bands when stimulus components are not independently represented internally. The amount of spectral contrast available to listeners with hearing loss when compared with normal-hearing listeners was reduced by similar amounts in all conditions, and it would be predicted that hearing-impaired listeners would have higher spectral shape discrimination thresholds than normal-hearing listeners in all conditions. The result that hearing impairment led to higher thresholds in the narrow span and not the wide span suggests that smearing of the spectral peaks in the stimulus may affect the abilities of hearing-impaired listeners to process changes in spectral shape.

Both groups of listeners were affected negatively by asynchronous information-bearing components. For the wide frequency span, thresholds increased by similar amounts for normal-hearing and hearing-impaired listeners with increasing onset asynchrony. For the narrow frequency span, thresholds were not greatly affected by onset asynchrony for either group of listeners. This result is likely due to within-channel interactions of stimulus components having different onsets and cannot be accounted for by temporal processing differences between the two groups.

ACKNOWLEDGMENTS

This work was supported by Grants DC 005835 and DC 00626 from the National Institute on Deafness and Communication Disorders. The Walter Reed Army Medical Center Clinical Investigation Committee and Human Use Committee approved this research, which was carried out under Work Unit No. 2591-99. The authors thank Lina Kubli for assistance with data collection, and two anonymous reviewers for their helpful comments on an earlier version of the manuscript. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

[1]Software used to generate the excitation patterns in Matlab was provided by Jens Appell, Carl-von-Ossietzky-Universitat, Germany.

[2]The authors thank an anonymous reviewer for suggesting this analysis to quantify spectral contrast available across conditions and groups.

ANSI (**1989**). ANSI S3.6-1989. ''Specification for audiometers,'' (American National Standards Institute, New York).

Bernstein, L. R., and Green, D. M. (**1987**). ''The profile-analysis bandwidth,'' J. Acoust. Soc. Am. **81**, 1888–1895.

Dai, H., and Green, D. M. (**1992**). ''Auditory intensity perception: Successive versus simultaneous across-channel discriminations,'' J. Acoust. Soc. Am. **91**, 2845–2854.

Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory Grouping,'' in *Hearing*, edited by B.C.J. Moore (Academic, New York), pp. 387–424.

Doherty, K. A., and Lutfi, R. A. (**1996**). ''Spectral weights for overall level discrimination in listeners with sensorineural hearing loss,'' J. Acoust. Soc. Am. **99**, 1053–1058.

Doherty, K. A., and Lutfi, R. A. (**1999**). ''Level discrimination of single tones in a multitone complex by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **105**, 1831–1840.

Durlach, N. I., Braida, L. D., and Ito, Y. (**1986**). ''Towards a model for discrimination of broadband sounds,'' J. Acoust. Soc. Am. **80**, 63–72.

Fitzgibbons, P. J., and Wightman, F. L. (**1982**). ''Gap detection in normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **72**, 761–765.

Glasberg, B. R., and Moore, B. C. J. (**1986**). ''Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,'' J. Acoust. Soc. Am. **79**, 1020–1033.

Glasberg, B. R., Moore, B. C. J., and Bacon, S. P. (**1987**). ''Gap detection and masking in hearing-impaired and normal-hearing subjects,'' J. Acoust. Soc. Am. **81**, 1546–1556.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Research **47**, 103–138.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University Press, New York).

Green, D. M., and Dai, H. (**1992**). ''Temporal Relations in Profile Comparisons,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, New York), pp. 471–477.

Green, D. M., and Kidd, G. Jr., (**1983**). ''Further studies of auditory profile analysis,'' J. Acoust. Soc. Am. **73**, 1260–1265.

Green, D. M., Kidd, G. Jr., and Picardi, M. C. (**1983**). ''Successive versus simultaneous comparison in auditory intensity discrimination,'' J. Acoust. Soc. Am. **73**, 639–643.

Grose, J. H., and Hall, J.W. III, (**1996a**). ''Cochlear hearing loss and the processing of modulation: Effects of temporal asynchrony,'' J. Acoust. Soc. Am. **100**, 519–527.

Grose, J. H., and Hall, J.W. III, (**1996b**). ''Perceptual organization of sequential stimuli in listeners with cochlear hearing loss,'' J. Speech Hear. Res. **39**, 1149–1158.

Hill, N. I., and Bailey, P. J. (**1997**). ''Profile-analysis with an asynchronous target: Evidence for auditory grouping,'' J. Acoust. Soc. Am. **102**, 477–481.

Klatt, D. (**1982**). ''Prediction of perceived phonetic distance from critical-band spectra: A first step,'' IEEE, ICASSP, 1278–1281.

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.

Leek, M. R., and Summers, V. (**1993a**). ''The effect of temporal waveform shape on spectral discrimination by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **94**, 2074–2082.

Leek, M. R., and Summers, V. (**1993b**). ''Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise,'' J. Acoust. Soc. Am. **94**, 3127–3137.

Leek, M. R., and Summers, V. (**1996**). ''Reduced frequency selectivity and the preservation of spectral contrast in noise,'' J. Acoust. Soc. Am. **100**, 1796–1806.

Lentz, J. J., and Leek, M. R. (**2002**). ''Decision strategies of hearing-impaired listeners in spectral shape discrimination,'' J. Acoust. Soc. Am. **111**, 1389–1398.

Lentz, J. J., and Leek, M. R. (**2003**). ''Spectral shape discrimination by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **113**, 1604–1616.

Lentz, J. J., Richards, V. M., and Matiasek, M. R. (**1999**). ''Different auditory filter bandwidth estimates based on profile-analysis, notched-noise, and hybrid tasks,'' J. Acoust. Soc. Am. **106**, 2779–2792.

Levitt, H. (**1971**). ''Transformed up–down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.
Moore, B. C. J., Oldfield, S. R., and Dooley, G. (**1989**) ''Detection and discrimination of spectral peaks and notches at 1 and 8 kHz,'' J. Acoust. Soc. Am. **85**, 820–836.
Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford University Press, New York).
Schooneveldt, G. P., and Moore, B. C. J. (**1989**). ''Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking bands,'' J. Acoust. Soc. Am. **85**, 262–272.
Sidwell, A., and Summerfield, Q. (**1985**). ''The effect of enhanced spectral contrast on the internal representation of vowel-shaped noise,'' J. Acoust. Soc. Am. **78**, 495–506.
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.
Summers, V., and Leek, M. R. (**1994**). ''The internal representation of spectral contrast in hearing-impaired listeners,'' J. Acoust. Soc. Am. **95**, 3518–3528.

Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants

Tim Green,[a] Andrew Faulkner, and Stuart Rosen
Department of Phonetics and Linguistics, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, United Kingdom

(Received 10 January 2004; revised 21 May 2004; accepted 24 June 2004)

The limited spectral resolution of cochlear implant systems means that voice pitch perception depends on weak temporal envelope cues. Enhancement of such cues was investigated in implant users and in acoustic simulations. Subjects labeled the pitch movement of processed synthetic diphthongal glides. In standard processing, noise carriers (simulations) or pulse trains (implant users) were modulated by 400 Hz low-pass envelopes. In modified processing, carriers were modulated by two components: (1) Slow-rate (<32 Hz) envelope modulations, conveying dynamic spectral shape changes crucial for speech; (2) a simplified waveform (e.g., a sawtooth) matching the periodicity of the input diphthong. In both normal listeners and implant users performance was better with modified processing, though temporal envelope cues were less effective with higher $F0$. Factors contributing to the advantage for modified processing may include increased modulation depth and use of a modulation waveform featuring a rapid onset in each period, resulting in a clearer representation of $F0$ in the neural firing pattern. Eliminating slow-rate spectral dynamics, so that within-channel amplitude changes solely reflected $F0$, showed that dynamic spectral variation obscured temporal pitch cues. Though significant, advantages for modified processing were small, suggesting that the potential for developing strategies delivering enhanced pitch perception is limited. [DOI: 10.1121/1.1785611]

PACS numbers: 43.66.Ts, 43.71.Bp, 43.66.Hg [KWG] Pages: 2298–2310

I. INTRODUCTION

A major factor in improving cochlear implant performance has been the development of speech processing strategies such as continuous interleaved sampling (CIS) (Wilson *et al.*, 1991), and related spectral peak-picking schemes such as SPEAK (Seligman and McDermott, 1995) and ACE (Vandali *et al.*, 2000). These widely used strategies, which present spectro-temporal information to the implanted array in the form of fixed-rate pulse carriers modulated by low-pass filtered amplitude envelopes extracted from a limited number of frequency bands, have achieved impressive speech intelligibility results. Nonetheless, there are important aspects of speech perception that have received little attention, both in the design of speech processors and in the methods typically used to assess them. One such aspect is intonation, as conveyed by voice pitch variation. Although there is evidence that the availability of voice pitch information has little effect on performance in simple measures of speech perception such as vowel and consonant recognition (Faulkner, Rosen, and Smith, 2000), there is no doubt that intonation makes important contributions to speech perception. In addition to being a major component of prosody, providing cues to linguistic features such as question-statement contrasts and other features such as the speaker's gender and identity, intonation is widely held to play an important role in early language development (e.g., Jusczyk, 1997). This is of particular significance given the increasing prevalence of implantation in very young children. Voice pitch information will, of course, be still more important in the perception of tonal languages, in which it conveys lexical meaning.

CIS and similar processing strategies eliminate important cues to voice pitch that are typically available to normally hearing listeners. The limited spectral resolution means that the lower harmonics of speech which provide the principal pitch cues for normal listeners are not resolved. Therefore, pitch perception in CIS processed speech will largely depend upon deriving temporal pitch cues from modulations of the amplitude envelope at the voice fundamental frequency ($F0$). The extent to which this is possible depends upon several factors. $F0$ must be passed by the envelope smoothing filter and the pulse rate must be high enough to represent modulations at $F0$. Both physiological and psychophysical evidence suggest that accurate representation of the modulating envelope requires a carrier pulse rate at least 4 to 5 times the frequency of the modulation (McKay, McDermott, and Clark, 1994; Wilson, 1997). Several widely used implant systems use a pulse rate of less than 1 kHz which is insufficient to cover much of the voice pitch range.

In addition to these constraints there are limits on the ability of the auditory system to perceptually encode temporal amplitude modulation. In normally hearing listeners the ability to use temporal cues to derive a pitch percept from amplitude modulated noise is limited to frequencies below around 300 Hz (Burns and Viemeister, 1976, 1981; Pollack, 1969) while thresholds for the discrimination of modulation rate increase substantially with modulation rate (Grant, Summers, and Leek, 1998; Hanna, 1992). Similarly, in implant

[a]Electronic mail: tim@phon.ucl.ac.uk

users the ability to detect amplitude modulations in pulses applied to a single channel typically declines rapidly for modulation frequencies above 100–150 Hz (Busby, Tong, and Clark, 1993; Cazals *et al.*, 1994; Donaldson and Viemeister, 2002; Shannon, 1992). Thus, while some pitch information derived from temporal modulations corresponding to $F0$ is likely to be available to implant users, intonation perception would be expected to be severely limited in comparison to normal hearing.

There has been little direct investigation of voice pitch perception in users of CIS and similar strategies. However there is some recent research that, consistent with the above reasoning, demonstrates that implanted children have great difficulty in identifying Cantonese lexical tones (Ciocca *et al.*, 2002; Lee *et al.*, 2002). The majority of the subjects in these two studies were users of the SPEAK strategy which, because of its low pulse rate (around 300 Hz) would be expected to be particularly poor at conveying voice pitch information. However, Ciocca *et al.* (2002) reported that users of the ACE strategy, employing pulse rates of 900 or 1200 Hz, were also poor at discriminating different tones, although they did not explicitly compare performance across the different strategies. Poor performance at these higher rates implicates limitations in the ability to derive pitch percepts from modulations in pulse amplitude corresponding to $F0$. Of course, it is possible that post-lingually deafened adult implantees, with their greater language experience and more advanced cognitive skills may be able to utilize temporal pitch cues more effectively. In a preliminary report Wei *et al.* (2001) did find substantially better than chance tone identification in implanted post-lingually deafened adult speakers of Mandarin Chinese. However, in contrast to Cantonese, factors other than the F0 contour, such as amplitude contour and duration, provide reliable cues to tone identification in Mandarin (Fu and Zeng, 2000; Whalen and Xu, 1992).

A number of studies have provided evidence concerning the possible pitch cues available with CIS processing by using noise-excited vocoder acoustic simulations with varying numbers of channels and envelope filter cutoff frequencies. While such simulations cannot fully emulate the pulsatile stimulation of CIS processing, they do provide similar limitations on spectral and temporal detail and would, therefore, be expected to provide a reasonably close approximation to the performance of implant users. Fu *et al.* (1998) found that identification of lexical tonal patterns in Mandarin Chinese was significantly better with an envelope filter cutoff frequency of 500 Hz compared to 50 Hz, indicating that listeners were able to utilize the temporal voice pitch cues that were available when the envelope filter covered the $F0$ range.

Xu, Tsai, and Pfingst (2002) carried out a more extensive investigation, using acoustic simulations to assess the effects of a wider range of channel numbers and filter cutoff frequencies on the identification of both actual Mandarin tones and FM sweeps designed to mimic such tones. The latter allowed duration and amplitude cues to be eliminated, resulting in substantially poorer performance. When such cues were present, performance benefited from increases in both the filter cutoff frequency and the number of channels. In the absence of such cues, however, while performance improved substantially as the low-pass cutoff frequency of the envelope filter was increased so as to include $F0$, the number of channels generally had little effect. With a low-pass cutoff frequency of 512 Hz mean recognition of the four different tonal patterns was around 60 percent for tones in the typical male $F0$ range and around 50 percent for tones with typical female $F0$ values. For stimuli in the higher $F0$ range performance was substantially improved with very high numbers of channels, i.e., 30–40. This improvement can be attributed to the increased spectral resolution and the larger spacing of the higher harmonics with higher $F0$ resulting in the presence of spectral pitch cues. Typically, however, cochlear implant users only have six to eight effective channels available (e.g., Fishman, Shannon, and Slattery, 1997; Friesen *et al.*, 2001) so that such cues are not present.

However, Faulkner, Rosen, and Smith (2000) identified certain circumstances in which spectral cues to pitch were apparently available even with a small number of channels. When subjects were required to identify the direction of pitch change of noise-excited vocoder processed sawtooth waveforms, performance was reasonably good when there were just four channels and the envelope filter cutoff frequency was as low as 32 Hz, eliminating temporal cues to pitch. It was suggested that the regular movement of harmonics between analysis bands of the processor as the $F0$ of the sawtooth waveforms changed provided a form of spectral cue in this case. Green, Faulkner, and Rosen (2002), used four-band and single-band noise-excited vocoders in combination with envelope filter cutoff frequencies of 32 and 400 Hz to further examine both this form of spectral cue and temporal cues to voice pitch. They found that such spectral cues were no longer effective when stimuli consisted of synthetic diphthongal vowel glides rather than sawtooth waveforms. This suggests that the spectral envelope changes resulting from formant movement obscure the possible spectral cues to $F0$. In addition it was found that the utility of temporal pitch cues declined with increasing $F0$ and was severely limited for fundamental frequencies above 200 Hz.

The results of these simulation studies are consistent with the limited data available from implant users in suggesting considerable limitations on the extent to which CIS and similar strategies are able to convey intonation. Given the limitations on spectral and temporal detail in current implant systems, attempts to improve voice pitch perception have focused on enhancing the temporal pitch cues that are available from envelope modulations corresponding to $F0$. One such attempt was reported by Geurts and Wouters (2001). They implemented a version of CIS processing with three major differences to the standard approach. Firstly, phase distortion in the analysis filters was eliminated, so that the maxima of $F0$-related modulations in each channel of the pulses delivered to the electrode array coincided in time. Secondly, half-wave rather than full-wave rectification was used in envelope extraction in order to provide a more accurate reflection of the phase-locked behavior of auditory nerve fibers. This was expected to result in increased modulation depth in the envelope of lower-frequency channels and is consistent with a simulation study by Rosen (1989) which

showed better perception of voicing with half-wave rather than full-wave rectification. Thirdly, envelope filtering was modified by subtracting an attenuated version of the output of a 50 Hz low-pass filter from that of the standard 400 Hz low-pass filter, resulting in a larger modulation depth for $F0$-related fluctuations. However, they did not find significant improvements in the ability of four users of the LAURA cochlear implant to discriminate changes in the $F0$ of synthesized monophthongs with the new strategy compared to standard CIS processing.

The current study investigates an alternative approach to the enhancement of temporal cues to voice pitch. This involves a decomposition of the envelope into two separate components. One consists of slow-rate information conveying the dynamic changes in spectral shape that are crucial for speech, while the second presents $F0$-related information in the form of a simplified synthesized waveform. Such an approach has a number of potential advantages. Firstly it allows the elimination of temporal complexities in the modulation envelope, such as modulation peaks corresponding to the periods of harmonics of $F0$. Because the responses of the auditory nerve to electrical stimulation are much more strongly synchronized to the input than in acoustic stimulation, such complexities make it less likely that the period of $F0$ will be clearly represented in the neural firing pattern. Secondly, it allows the same modulation envelope to be applied to all channels. Geurts and Wouters (2001) showed that discrimination of the modulation frequency of sinusoidally amplitude modulated pulse trains was better when three adjacent channels were stimulated concurrently compared to when any one channel was stimulated alone. Applying the same $F0$-related modulation across all channels would, therefore, be expected to be beneficial. Thirdly, the shape of the modulating waveform can be optimized in order to maximize the salience of temporal pitch cues.

The timing of neural firing in response to modulation is likely to depend on the overall level of the envelope as well as the location in time of the peak of the modulating waveform. Thus, for speech signals, which have substantial variations in level, it seems likely that a ''temporally sharpened'' modulation envelope, with a rapid onset in each period, will lead to more consistent inter-pulse intervals in the neural firing pattern, and therefore, to enhanced voice pitch perception. It may well be the case that optimal transmission of pitch information would be attained by presenting just a single pulse in each modulation period. This kind of approach has been implemented in speech processing strategies in which the rate of stimulation is controlled by $F0$ (e.g., Fourcin *et al.*, 1979; Tong *et al.*, 1980). However, such strategies have generally been found to produce poorer performance on standard measures of speech perception than CIS-like strategies (e.g., McKay and McDermott, 1993; McKay *et al.*, 1992; Skinner *et al.*, 1999), presumably because of poorer transmission of spectral information.[1] The current approach is based upon conveying $F0$-related modulation with a sawtooth shaped waveform, which it is hoped will enhance the salience of temporal pitch cues while still allowing adequate encoding of the low-rate envelope information necessary for speech perception.

In a real world application such an approach would require the explicit extraction of $F0$ from the speech input. There are existing algorithms (e.g., Walliker and Howard, 1990) which perform this task at acceptable levels in moderately noisy conditions (Bosman and Smoorenburg, 1997), although in the current research only synthesized speech stimuli were used, making $F0$-extraction unnecessary. Initially, the influence of the shape of temporal modulation on pitch cues was examined using noise-excited vocoder simulations. A modified processing scheme was then developed and its effectiveness in transmitting voice pitch information compared with that of a standard CIS strategy in both simulations and implant users.

II. ACOUSTIC SIMULATION EXPERIMENTS

A. Signal processing

Two experiments were carried out using acoustic simulations. In experiment 1 three different versions of noise-excited vocoder processing were compared. In each case signals were first passed through a bank of fourth-order elliptic analysis filters to divide the spectrum into eight frequency bands. Cutoff frequencies, which were based on the Clarion S-Series Pulsatile Table with extended low- and high-frequency settings (Clarion Device Fitting Manual, 2001), were 250, 500, 730, 1015, 1450, 2000, 2600, 3800, and 6800 Hz. Note that because the lowest cut-off frequency is 250 Hz, the possibility of directly encoding $F0$ in the spectral domain is effectively eliminated. Amplitude envelopes were extracted from each band by half-wave rectification and low-pass filtering (second-order Butterworth) with a cutoff frequency of 32 Hz, so that $F0$-related fluctuations were eliminated. These envelopes then modulated independent noise carriers, prior to output filtering that matched the initial analysis filtering and a final summation across channels. The three processing conditions differed in the extent to which the noise carrier was manipulated in order to represent $F0$ information. In one condition, designated *None*, the carrier in each band was simply unmodulated noise, as in standard noise-excited vocoder simulations of implant processing. This condition was included to measure the contribution to performance of purely spectral cues that might arise from the shifting of harmonics through analysis bands (Faulkner, Rosen, and Smith, 2000). Although Green, Faulkner, and Rosen (2002) found that such cues were eliminated by the presence of speech-like spectral variation, their simulation used only four frequency bands. It is conceivable that with a greater number of channels such spectral cues may still have some utility.

The two other conditions were designed to assess the efficacy of temporal pitch cues provided by simplified modulation patterns, and the effects of temporally sharpening the $F0$-related modulation waveform. Temporal pitch cues were introduced by modulating the noise carrier in each channel with a waveform with identical periodicity to that of the input vowel, but with a simplified waveform shape, eliminating the temporal complexities in the $F0$-related modulations typically present in the envelopes extracted in CIS processing. The modulation depth was always 100 percent. In one

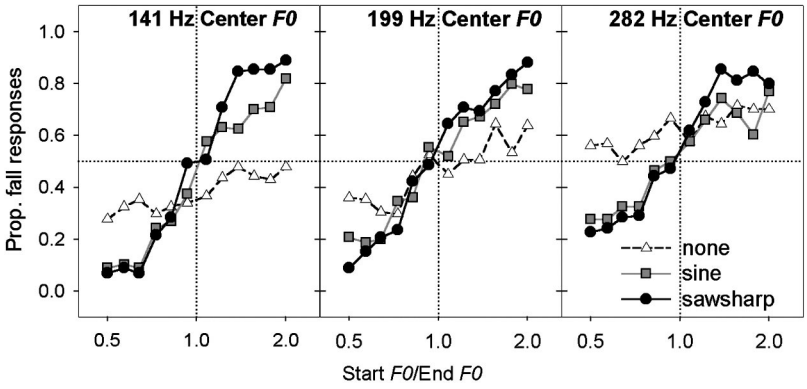


FIG. 1. Proportion of ''fall'' responses averaged across normally hearing listeners as a function of start-to-end frequency ratio for each processing condition and $F0$ range of experiment 1.

case, designated *Sine*, the modulating waveform was sinusoidal. In the other (*Sawsharp*) it was a modified sawtooth waveform in which the fall from peak to zero occurred in the first half of the period, with the second half of the period at zero. Pilot experiments in which a single listener labeled the direction of pitch change of diphthongal glides had suggested that this additional temporal sharpening increased the salience of temporal pitch cues relative to a standard sawtooth shape.

Two processing conditions were compared in experiment 2—the *Sawsharp* condition from experiment 1 and a simulation of standard CIS processing (*CIS*). The latter was identical to the *None* condition of experiment 1 with the exception that the cut-off frequency of the envelope smoothing filter was 400 Hz, a value typically used in CIS processing strategies and one that allows temporal envelope cues to be derived from modulations in the voice pitch range.

B. Stimuli

Stimuli were generated off-line using MATLAB. As in Green, Faulkner, and Rosen (2002) they consisted of processed versions of the diphthongs /aʊ/, /eɪ/, /aɪ/, and /ɔɪ/ which had been created using an implementation of the KLSYN88 Klatt synthesizer in cascade mode with a 20 kHz sample rate and parameters specified every 5 ms. Formant frequency trajectories were based on recordings of each diphthong embedded in a /cVc/ context spoken by a male Southern British English speaker. $F0$ changed logarithmically over the 500 ms duration.

In experiment 1 the ratio of start to end frequencies varied in six equal logarithmic steps from 1:0.5 to 1:0.93. There were three $F0$ ranges, with the center $F0$ (i.e., the geometric mean of the start and end $F0$s) of each glide being 141, 199, and 282 Hz respectively. The same formant values were used irrespective of the $F0$ range. For each diphthong, each ratio and each $F0$ range there was one ascending and one descending glide, giving a total of 144 different glides.

In experiment 2 a smaller set of stimuli was used. Three, rather than six, ratios of start to end $F0$ were used: 1:0.5, 1:0.66, 1:0.87, and there were just two different $F0$ ranges, with center frequencies of 113 and 226 Hz, approximately corresponding to the voice pitch ranges of adult male and female speakers. This resulted in a total of 48 different unprocessed glides.

C. Subjects and procedure

A total of seven normally hearing listeners aged between 21 and 40 years, including the first author (S5), participated, six in each experiment. Stimuli were presented binaurally through Sennheiser HD 414 headphones at a comfortable listening level (peak of 85–90 dB SPL measured over an 80 ms window). On each trial subjects heard a single glide and were required to identify it as either ''rising'' or ''falling'' in pitch. They responded via computer mouse by clicking on an image of either a rising or falling line. Before each block of trials subjects were able to listen to a selection of the glides to be presented in that block, visually labeled as rising or falling.

In experiment 1 processing condition was varied across blocks of trials consisting of the 144 glides presented in random order. The order in which blocks of trials were presented was random with the constraint that each set of three blocks contained one block with each type of processing. No feedback was provided.

In experiment 2 blocks consisted of 48 trials. Each block contained, in random order, two presentations of each of the 24 glides for one center $F0$ and one processing condition. Different $F0$ ranges were presented in separate blocks in an attempt to minimize a shift in response bias with center $F0$ that was evident for some subjects in experiment 1. In addition, visual feedback was presented after each trial. In both experiments subjects completed each different block of trials seven times, with the first treated as practice.

D. Results and discussion

1. Experiment 1—Temporal pitch cues with simplified F0 modulation

Mean psychometric functions for the proportion of ''fall'' responses as a function of the ratio of start to end frequencies for each $F0$ range and each processing condition are shown in Fig. 1. In the *None* condition there is a clear change in bias across the different $F0$ ranges, with the probability of glides being labeled as ''falling'' increasing as center $F0$ increases. However, for each center $F0$ the functions in the *None* condition are nearly flat, showing that listeners were unable to reliably discriminate the direction of pitch change with even an octave change in $F0$ over the course of the glide. Thus, the increased number of channels did not alter the finding of Green, Faulkner, and Rosen (2002) that

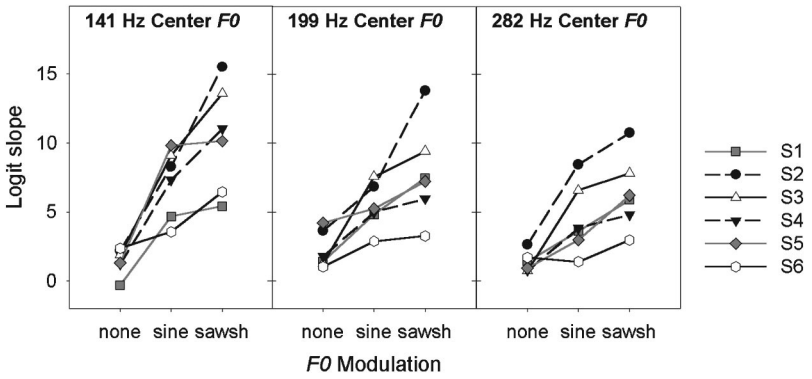


FIG. 2. Slopes of logistic regressions of the proportion of ''fall'' responses as a function of the log (base 10) of start-to-end frequency ratio for each normally hearing listener in each processing condition and $F0$ range of experiment 1. Higher slope values indicate better discrimination of the direction of pitch change.

spectral pitch cues are effectively eliminated by the time-varying spectral envelope typical of speech. For the other two conditions, pitch discrimination performance, as indicated by the steepness of the psychometric function, clearly declines with increasing center $F0$, but is still above chance even for the highest $F0$ range. For each center $F0$ the psychometric function is steeper, and therefore, pitch discrimination is better, in the *Sawsharp* rather than the *Sine* condition.

A logistic regression was carried out on the proportion of ''fall'' responses averaged across diphthongs as a function of the log (base 10) of the start-to-end frequency ratio for each processing condition and center $F0$ for each subject.[2] The regression slope estimates, displayed in Fig. 2, show that the pattern of performance across center $F0$ and processing condition apparent in the mean data was consistent across individual listeners. The intercept values obtained indicated that for subjects S1, S4, and S6 there was a consistent change in bias with center $F0$, such that ''falling'' responses became more likely with increasing $F0$ in all three processing conditions. For the other three subjects there was no clear pattern of bias across the different processing conditions and $F0$ ranges. Slope estimates for the *Sine* and *Sawsharp* conditions, in which temporal pitch cues were available, were analyzed using a two-way repeated-measures analysis of variance (ANOVA) with factors of processing condition and center $F0$. Here and elsewhere in this study the reported F tests used Huynh–Feldt epsilon correction factors. Both the effects of processing condition $[F(1,5)=17.85,\ p=0.008]$ and center $F0$ $[F(2,10)=14.53,\ p=0.010]$ were significant, but the interaction was not $[F(2,10)<1]$.

In addition to confirming the absence of reliable spectral cues to voice pitch in the presence of variations in spectral envelope typical of speech, Experiment 1 has shown that, when $F0$ information in the amplitude envelope is presented in a clarified form, temporal cues to voice pitch are available in each of the $F0$ ranges tested here, though their utility declines as $F0$ increases. In the four-band simulation of CIS processing (400 Hz envelope filter) used by Green, Faulkner, and Rosen (2002), performance in a nearly identical glide labeling task declined much more steeply with increasing $F0$ and was effectively at chance levels for glides in the highest $F0$ range. However, differences in factors such as the number of frequency bands and the range covered by the analysis filters mean that it is not certain that this difference reflects an advantage for the simplified modulation used here relative to CIS processing.

The salience of temporal pitch cues clearly is affected by the pattern of the modulating waveform. The temporally sharpened waveform in the *Sawsharp* condition was more effective at conveying pitch cues than sinusoidal modulation. This supports the idea that a rapid onset at the start of each modulation period results in a neural firing pattern which more accurately represents $F0$. The rate of decay of the modulating waveform shape from its peak value and the

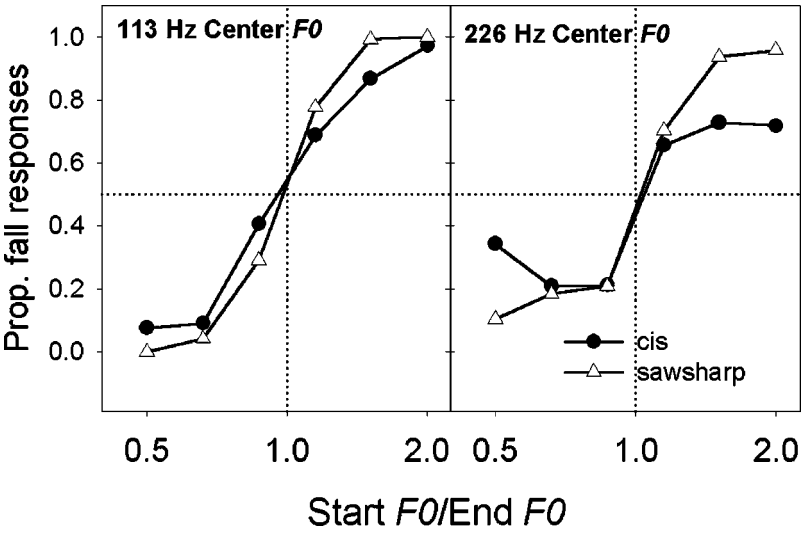


FIG. 3. Proportion of ''fall'' responses averaged across normally hearing listeners as a function of start-to-end frequency ratio for each processing condition and $F0$ range of experiment 2.

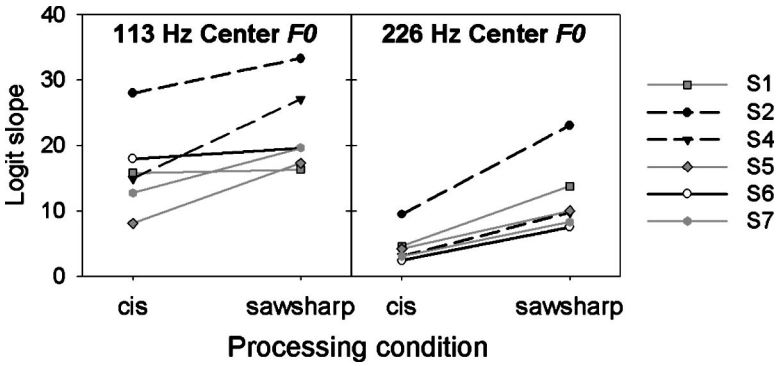


FIG. 4. Slopes of logistic regressions of the proportion of ''fall'' responses as a function of the log (base 10) of start-to-end frequency ratio for each normally hearing listener in each processing condition and $F0$ range of experiment 2.

length of time within each period that the waveform is at zero may also be important.

2. Experiment 2—Comparison of temporal pitch cues with simulated CIS processing versus simplified F0 modulation

Figure 3 shows the proportion of ''fall'' responses averaged across the six listeners and the four diphthongs as a function of the ratio between start and end frequencies for each $F0$ range and processing condition. Figure 4 shows slope estimates obtained from logistic regressions carried out as in Experiment 1. Intercept values were quite variable across listeners and conditions but in contrast to experiment 1 did not show any consistent pattern of bias. This difference is presumably attributable to the restriction of glides presented within a block to a single $F0$ range and the provision of feedback. The changes in procedure also appear to have resulted in increased discrimination of the direction of pitch change, since slope values with *Sawsharp* processing for all five listeners who took part in both experiments are higher for the 226 Hz center $F0$ glides in experiment 2 than with the 199 Hz center $F0$ glides in experiment 1. This is despite the fact that the increase in $F0$ would be expected to lead to poorer performance, though it is also possible that there may have been an overall practice effect.

In the lower $F0$ range mean glide labeling performance is somewhat better for the *Sawsharp* condition than the *CIS* condition. Four listeners show clearly higher slopes in the *Sawsharp* condition, but for the others there is little difference across processing conditions. In the higher $F0$ range, while performance overall is inferior to that in the 113 Hz center $F0$ condition, there is a clearer advantage for the modified processing strategy with all six listeners showing substantially higher slope estimates in the *Sawsharp* condition.

A repeated measures ANOVA showed significant effects of processing condition [$F(1,5)=44.69$, $p=0.001$] and center $F0$ [$F(1,5)=45.57$, $p=0.001$], while the interaction was not significant [$F(1,5)<1$]. Repeated measures *t*-tests comparing the two processing conditions separately for each $F0$ range showed significantly steeper slopes in the *Sawsharp* condition in both the higher [$t(5)=5.63$, $p=0.002$], and lower [$t(5)=3.27$, $p=0.02$] $F0$ ranges.

In summary, in noise-excited vocoder simulations, the modified processing method significantly increases the salience of temporal cues to voice pitch, relative to standard CIS processing. The reduction in temporal complexity in the modulation envelope, the fact that modulation depth was always 100 percent, applying the same modulation envelope to all channels, and the temporal sharpening of the modulation envelope probably all contribute to the increased effectiveness of temporal pitch cues relative to standard CIS processing.

III. TEMPORAL PITCH CUES IN IMPLANT USERS

Two glide labeling experiments were carried out with implant users. Experiment 3 was very similar to the acoustic simulation comparison between *CIS* and *Sawsharp* processing in experiment 2, but included an additional modified processing scheme in which the $F0$-related modulation was presented in the form of a normal, rather than sharpened, sawtooth. The inclusion of this additional condition (*Saw*)

TABLE I. Subject demographic information.

Subject	Age at onset	Years of profound deafness	Years of implant use	Etiology of deafness	Implant type	Strategy[a]	Mean threshold (Clinical Units)	Mean MCL (Clinical Units)
C1	68	2	3	otosclerosis	Enhanced Bipolar	SAS	81	288
C2	69	3	4	unknown	Enhanced Bipolar	MPS	97	397
C3	40	28	4	progressive	Enhanced Bipolar	MPS	60	633
C4	64	5	3	skull fracture	Hifocus	SAS	49	228
C5	51	10	2	unknown	Hifocus	MPS	40	138
C6	30	11	3	unknown	Hifocus	MPS	26	499
C7	45	17	3	unknown	Hifocus	MPS	85	174
C8	42	9	5	sensorineural	Enhanced Bipolar	CIS	97	589

[a]MPS is a variant of the CIS strategy in which pairs of channels are stimulated simultaneously.

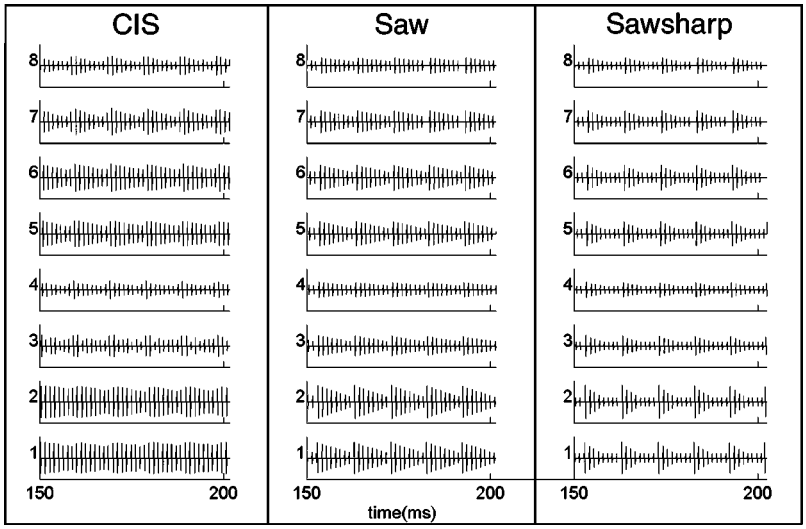


FIG. 5. Examples of the pulsatile output in each processing condition of experiment 3. The excerpts shown are from processed versions of the most steeply rising /eɪ/ glide with 113 Hz center $F0$ for subject C4.

was intended to provide information about some of the factors underlying possible benefits of the modified processing scheme. Both modified processing conditions feature a rapid onset at the start of each modulation period. Differences in performance across the two conditions would show that the salience of temporal pitch cues is also affected by the additional temporal sharpening of the modulation envelope in the *Sawsharp* condition.

Experiment 4 measured performance in conditions in which slow-rate spectral envelope changes were eliminated and within-channel variations in amplitude over the course of each stimulus were determined solely by the varying $F0$ of the vowel glides. This allowed an assessment of the extent to which such dynamic spectral variation affects the utility of temporal pitch cues. In addition to *CIS* and *Sawsharp* processing, experiment 4 also included a condition designated *Single*, in which stimuli consisted of a single pulse in each channel for each $F0$ period. Controlling the pulse rate by $F0$ in this way would be expected to provide the clearest possible representation of $F0$ in the neural response pattern, and therefore, to maximize the availability of temporal pitch cues.

A. Subjects and equipment

Eight post-lingually deafened adult users of the eight-channel Clarion 1.2 cochlear implant system took part. Summary information is contained in Table I. Two other subjects were excluded after pilot experiments showed that they were unable to perform above chance in the vowel glide labelling task in any condition. Both the excluded subjects had been profoundly deaf for over forty years prior to implantation, which is substantially longer than the other subjects and may be an important factor underlying their poor performance.

Experiments were controlled by a PC connected to a Clarion Research Interface (CRI) system (Wygonski *et al.*, 1999). The CRI system enables direct control over the stimulus patterns presented to the electrode array. It employs a DSP board (Motorola DSP56302) which is connected between the PC and a Clarion S-Series speech processor which in turn is connected to a standard headpiece. Regardless of the strategy and stimulus configuration normally used by subjects, stimulation always consisted of continuously-interleaved, monopolar, biphasic pulses with a duration of 76.9 μs per phase. The carrier rate was 812.5 pulses per second per electrode and electrodes were activated sequentially in apical-to-basal order.

Current levels delivered to electrodes were specified in clinical units. Nominally, for an ideal current source, 1 clini-

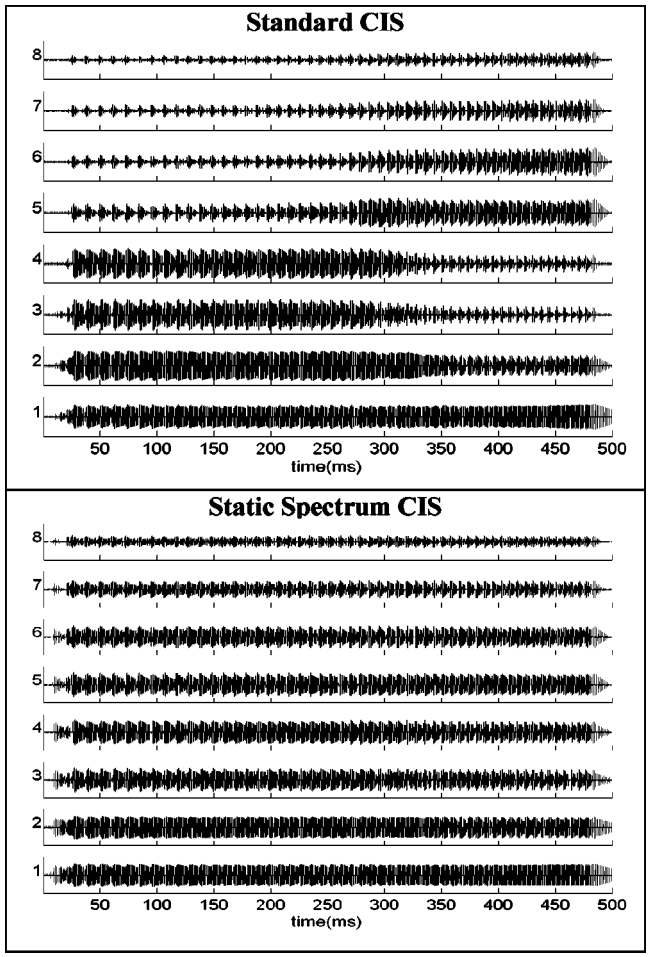


FIG. 6. Examples of the pulsatile output in CIS processing conditions in experiment 4. The stimuli shown are processed versions of the most steeply rising /oɪ/ glide with 113 Hz center $F0$ for subject C6.

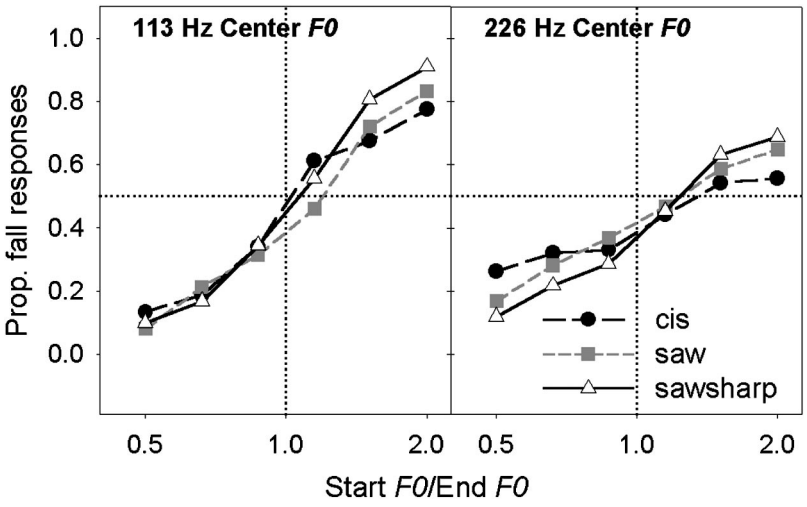


FIG. 7. Proportion of "fall" responses averaged across implant users as a function of start-to-end frequency ratio for each processing condition and $F0$ range of experiment 3.

cal unit corresponds to 1 μA. In practice the relationship between current levels and clinical unit values is somewhat nonlinear and dependent upon electrode impedance. Prior to experimentation, the CRI system was calibrated using a research intra-cochlear stimulator (RICS) connected to 15 kΩ resistors. Output current varied linearly with amplitude value specified in clinical units for values up to approximately 800. Although the point at which output current ceases to be a linear function of clinical unit value will vary to some extent according to the impedance of the stimulated electrode, the most comfortable current levels (MCLs) shown in Table I indicate that the system is unlikely to have been operating outside its linear range.

B. Speech processing

1. Experiment 3—Simplified F0 modulation compared to standard CIS processing

In both modified and standard CIS processing, the analysis filtering and envelope extraction stages were implemented as in the acoustic simulations, with one exception. Consistent with most commonly used CIS processing schemes, full-wave rather than half-wave rectification was used. Although half-wave rectification might be expected to increase the modulation depth of $F0$-related fluctuations (Geurts and Wouters, 2001; Rosen, 1989), pilot experiments which explicitly compared the two methods showed no effect on vowel glide labelling performance.

Amplitude envelopes were re-sampled to a rate of 6500 Hz, consistent with the overall pulse rate. In the *CIS* condition, the 400 Hz smoothed envelopes were then logarithmically compressed using a method similar to that of Geurts and Wouters (2001), and mapped to current values consistent with the dynamic range of the particular subject and channel. The envelope sample value that was mapped to the most comfortable current level (MCL) was the 99th percentile of the sample values for all channels and all stimuli in that processing condition. Sample values that were 50 dB or more below this value were mapped to the threshold current value (THR). Between these values sample values were compressed according to the following equation

$$y = 400 \log(x) + c, \tag{1}$$

where $y = 1000$ for the sample value (x) corresponding to MCL and $y = 0$ for the sample value corresponding to THR. The value of the constant c was chosen so as to avoid discontinuities at the transition points.

Modified processing was implemented by first calculating the starting point of each successive period of $F0$. Within each period the pulse train was amplitude modulated by the appropriate waveform shape. This pulse train provided the carrier for channel 1, while in other channels the carrier consisted of an appropriately delayed version of the pulses of each period in channel 1 up to the point at which the next period started. The modulation depth in each channel was 100 percent of that channel's dynamic range. Thus, in the *Saw* condition the amplitude of the carrier pulses declined from MCL to THR over the course of each period, while in the *Sawsharp* condition this decline occurred in the first half of each period, with all pulses in the second half of each period being at THR. The first pulse of each period was always delivered to channel 1, regardless of the channel on which the preceding pulse (i.e., the final pulse of the previous period) had occurred. Except when this preceding pulse was on channel 8, this meant a departure from the sequential order of stimulation. While the overall pulse rate was the same as in the *Standard CIS* condition, at the transition between successive periods the intervals between consecutive biphasic pulses could vary, being shorter than usual on the lower-frequency channels and higher than usual on the higher-frequency channels. These $F0$-modulated pulse carriers were then modulated by slow-rate amplitude envelopes with compression and mapping carried out as in *CIS* processing. Figure 5 shows example stimuli for each processing method.

2. Experiment 4—Temporal pitch cues in the absence of dynamic spectral variation

In the *Sawsharp* condition, slow-rate spectral dynamics were eliminated by calculating for each diphthong the average root-mean squared (RMS) level of the envelope sample values for each channel over the set of different start-to-end $F0$ ratio glides. These values were then multiplied by the 100 percent modulated $F0$-following sharpened sawtooth waveform prior to mapping between the appropriate thresh-

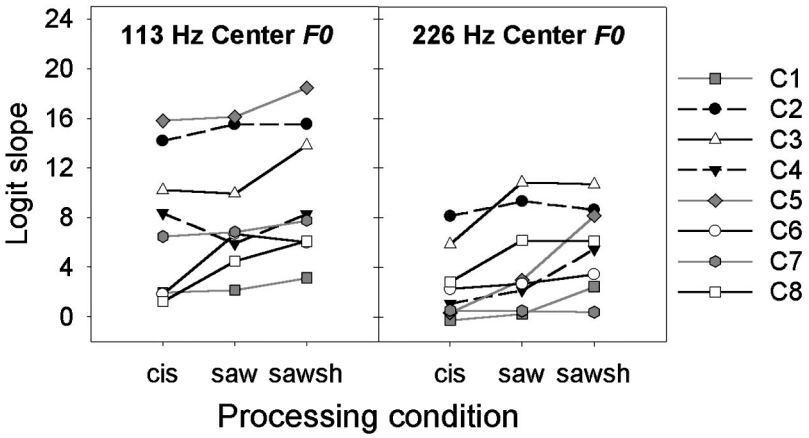


FIG. 8. Slopes of logistic regressions of the proportion of ''fall'' responses as a function of the log (base 10) of start-to-end frequency ratio for each implant user in each processing condition and $F0$ range of experiment 3.

old and MCL levels. The same average envelope sample values were used in the *Single* condition in which stimuli consisted of a single pulse in each channel for each $F0$ period.

In the *CIS* condition, slow-rate spectral variation was eliminated by modifying the envelope extraction stage of processing such that, in each channel, the waveform resulting from analysis filtering was divided, on a sample-by-sample basis, by a lowpass (30 Hz cutoff) envelope extracted from a full-wave rectified version of that waveform. The result, which contained only the higher-rate envelope modulations of the original waveform, was then full-wave rectified, lowpass filtered at 400 Hz, and adjusted in level to match the RMS value of the envelope samples for each channel in standard CIS processing before finally being mapped between the appropriate threshold and MCL levels. Any negative values resulting from filtering at 400 Hz were mapped to threshold level. As illustrated in Fig. 6, this resulted in pulsatile stimuli in which modulations in the $F0$ range were preserved, while slow-rate spectral changes were eliminated.

C. Stimuli and procedure

Both experiments with implant users used the same glide labeling task as in the acoustic simulations, with feedback always provided. Stimuli consisted of processed versions of the smaller set of vowel glides used in experiment 2. Blocks of trials again consisted of two presentations of each of the 24 glides for one processing condition and one center $F0$. In both experiments subjects completed a minimum of five blocks of trials for each combination of processing condition and center $F0$, with the first being treated as practice.

D. Results and discussion

1. Experiment 3—Comparison of temporal pitch cues with simplified F0 modulation versus standard CIS processing

Figure 7 shows the proportion of ''fall'' responses averaged across the eight implant users and four diphthongs as a function of the ratio of start to end frequencies for each $F0$ range and processing condition. Comparison with Fig. 3 shows that mean performance in both $F0$ ranges in both the *CIS* and *Sawsharp* conditions is substantially worse than in the acoustic simulation of the equivalent conditions in experiment 2. Even in the lower $F0$ range, mean performance in the *CIS* condition, while clearly above chance, is very limited. Even with an octave change in $F0$ over the course of the glide, the direction of pitch change is correctly identified on only around 80 percent of trials. Mean performance is improved in the *Saw* and *Sawsharp* conditions relative to *CIS*, though is still fairly poor.

Logistic regressions were carried out as in experiment 1. Intercept values showed a general trend for a bias towards ''rise'' responses that was fairly consistent across subjects and conditions. Regression slope estimates, displayed in Fig. 8, show a large degree of variation across subjects, with the better performers producing slope values within the range of values obtained from normally hearing listeners in experiment 2. A repeated measures ANOVA performed on the slope estimates showed significant effects of both processing condition [$F(2,14)=14.26$, $p<0.001$] and center $F0$ [$F(1,7)=8.47$, $p=0.020$], while the interaction was not significant [$F(2,14)<1$]. Bonferroni-corrected *post hoc* tests showed that slope values were significantly higher with *Sawsharp* than with *CIS* processing ($p=0.006$), while the difference between the *CIS* and *Saw* conditions just missed significance ($p=0.079$), as did that between the *Saw* and *Sawsharp* conditions ($p=0.074$).

Both the decline in glide labeling performance with increasing $F0$ and the significant advantage for *Sawsharp* over *CIS* processing are consistent with the results obtained with the acoustic simulation in experiment 2. In addition, the pattern of poorer mean performance by implant users compared to normally hearing listeners with noise-excited vocoder processing, but with the best implant users achieving scores within the range of those obtained in the simulations, is typical of comparisons between simulated and actual implant performance in speech recognition tasks (e.g., Friesen *et al.*, 2001; Fu, Shannon, and Wang, 1998). This supports the idea that the temporal pitch cues available from noise-excited vocoder simulations have a strong similarity to those available from CIS-like processing in implant systems. In both cases, the salience of temporal cues to pitch is enhanced by clarification of the $F0$-related modulation.

From inspection of Fig. 5 (see also Fig. 6) a particularly important difference between the modified and CIS processing schemes is likely to be the increased modulation depth of $F0$-related fluctuations in the lower frequency channels. The

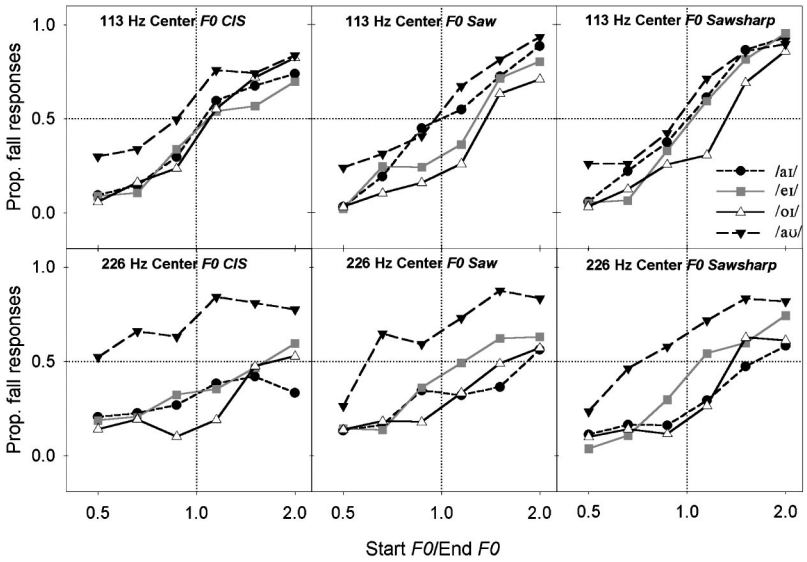


FIG. 9. Proportion of ''fall'' responses averaged across implant users as a function of start-to-end frequency ratio for individual diphthongs in each processing condition and $F0$ range of experiment 3.

synthetic vowel stimuli used here appear to provide a relatively simple $F0$ modulation pattern, with little temporal complexity in the form of, for example, secondary peaks within modulation periods. With natural speech the fact that the modified scheme simplifies the modulating waveform might provide an additional benefit relative to CIS processing.

The notion that the additional temporal sharpening of the modulation waveform provided in the *Sawsharp* compared to the *Saw* condition further enhances the salience of temporal pitch cues is not unambiguously supported. However, the difference between the two conditions was close to significance and for some subjects in one or both $F0$ ranges, slope values were considerably higher in the *Sawsharp* condition than in the *Saw* condition, while there were no instances of substantially higher slopes in the *Saw* condition. It is possible that different degrees of temporal sharpening of the modulation waveform shapes may be necessary to produce optimal performance for different implant users.

One factor that might be expected to influence individual variability in the salience of temporal pitch cues is the implantee's absolute dynamic range. Some previous evidence suggests that unlike in normal hearing, where sensitivity to amplitude modulation is determined largely by the relative modulation depth independent of the carrier level, in implant users a larger absolute modulation depth increases sensitivity to amplitude modulation and enhances discrimination of the modulation frequency (Geurts and Wouters, 2001; McKay, McDermott, and Clark, 1995; Shannon, 1992). Thus, subjects with a larger dynamic range might be expected to perform better in the glide labelling task. Pearson correlation coefficients were calculated for mean dynamic range across electrodes (in clinical units) and slope values in each of the six combinations of processing condition and center $F0$. While performance in the *Saw* condition when the center $F0$ was 226 Hz was significantly correlated with mean dynamic range ($r=0.765$, $p=0.027$), none of the other correlations was significant, so there is little evidence here that a larger dynamic range leads to greater salience of temporal pitch cues. However, it should be noted that the two subjects who were excluded from the study had mean ranges of around 100 clinical units, near the smallest value for any of the included subjects. In addition, subject C5, who had a small dynamic range but amongst the highest slope values, was a keen musician prior to becoming deaf which may have contributed to her relatively good performance.

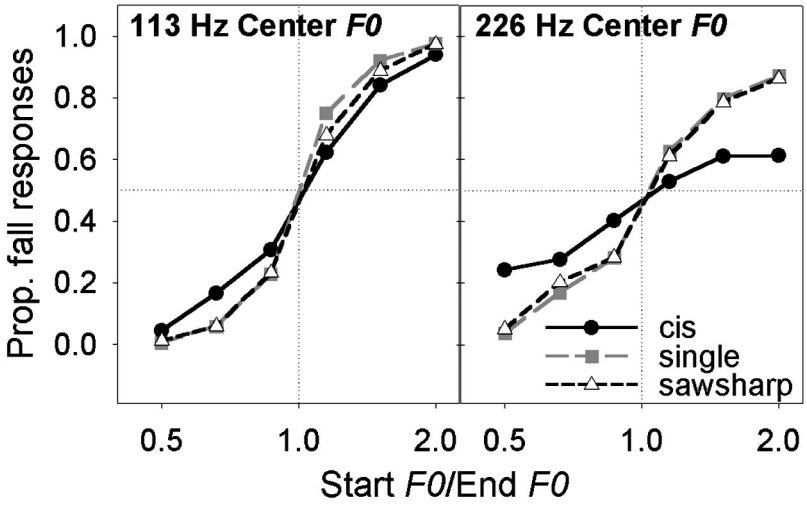


FIG. 10. Proportion of ''fall'' responses averaged across implant users as a function of start-to-end frequency ratio for each processing condition and $F0$ range of experiment 4, in which dynamic spectral variation was eliminated.

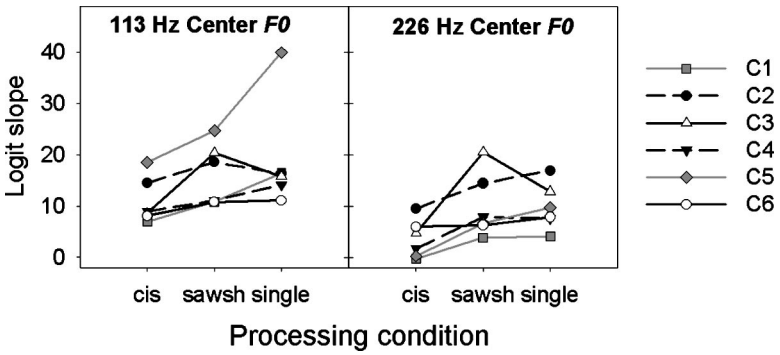


FIG. 11. Slopes of logistic regressions of the proportion of ''fall'' responses as a function of the log (base 10) of start-to-end frequency ratio for each implant user in each processing condition and F0 range of experiment 4, in which dynamic spectral variation was eliminated.

Figure 9 shows mean psychometric functions for each of the four diphthongs for each processing condition and center $F0$. It is noticeable that, particularly in the higher $F0$ range, responses are strongly influenced by vowel identity. For example, there was a tendency for /aʊ/ to be perceived as falling in pitch regardless of the actual direction of pitch change, while for the other three diphthongs there was a bias towards ''rise'' responses. All the formant frequencies of the /aʊ/ stimuli either remained constant or declined monotonically over the course of the glide, whereas the other diphthongs all featured a rising second formant frequency. This suggests that the shifts in the distribution of energy across analysis bands resulting from the dynamically changing spectral structure of the diphthongs can conflict with and obscure temporal cues to pitch derived from $F0$-related modulations in the amplitude envelope. A similar conflict between spectral and temporal cues was reported by Zeng (2002) on the basis of pitch magnitude estimates obtained in response to single-channel pulse trains of various rates presented at different electrode locations.

2. Experiment 4—Temporal pitch cues in the absence of dynamic spectral variation

Figure 10 shows mean glide labelling responses for six implant users in conditions in which dynamic spectral changes associated with changing formant structure were eliminated and modulations in amplitude within each channel were confined to the $F0$ range. Comparison with Fig. 7 shows that psychometric functions for the equivalent processing conditions are generally substantially steeper in the absence of slow-rate spectral variation. Logistic regression slope estimates are displayed in Fig. 11. An overall tendency for a bias towards ''rise'' responses was again apparent, though less strongly than in experiment 3.

Repeated measures ANOVA performed on slope values revealed a significant effect of processing condition [$F(2,10)=8.25$, $p=0.008$], while the effect of center $F0$ just missed significance [$F(1,5)=5.86$, $p=0.060$], and the interaction was not significant [$F(2,10)=1.08$, $p=0.365$]. According to Bonferroni-corrected *post hoc* tests, performance in the *CIS* condition was significantly poorer than that in both the *Sawsharp* ($p=0.044$), and *Single* ($p=0.034$) conditions, while the *Sawsharp* and *Single* conditions did not differ ($p=1.000$).

The trend towards a decline in performance for the higher $F0$ range and the advantage for *Sawsharp* processing over *CIS* are consistent with the previous results obtained both in the acoustic simulation in experiment 2 and with implant users in experiment 3. The similarity in performance in the *Sawsharp* and *Single* conditions suggests that the *Sawsharp* processing scheme provides close to the maximum possible temporal pitch information, at least for conditions in which across-channel differences in amplitude are encoded, as they must be to preserve speech spectral information.

To directly assess the influence of dynamic spectral variation on the effectiveness of temporal pitch cues, slope estimates from the *Sawsharp* and *CIS* conditions of experiments 3 and 4 were submitted to a three-way ANOVA. The presence of spectral variation had a significant effect [$F(1,5)=44.20$, $p=0.01$], while significant effects of processing condition [$F(1,5)=15.89$, $p=0.010$], and center $F0$ [$F(1,5)=8.77$, $p=0.030$], were again found, but there were no significant interactions. This confirms that the effectiveness of the temporal pitch cues derived from $F0$-related envelope modulations was hindered by the presence of dynamic spectral variation reflected in changes in the distribution of energy across channels.

IV. GENERAL DISCUSSION

A. Salience of temporal pitch cues

The modified processing scheme, in which the standard 400 Hz smoothed amplitude envelope was replaced by the product of a slow rate envelope and simplified $F0$-related modulation, consistently resulted in significantly better glide labeling performance than standard CIS processing, both in acoustic simulations and in implant listeners. This enhancement of the salience of temporal envelope cues to voice pitch presumably reflects a clearer representation of $F0$ in the neural firing pattern. Although there are considerable differences between the neural responses to electrical stimulation and those to noise-excited vocoder processed sound in normally hearing listeners, the consistent pattern of results across the experiments with acoustic simulations and with implant listeners suggests a considerable degree of similarity between the temporal pitch cues available in the two cases.

A number of factors may contribute to the better pitch perception obtained with the modified processing scheme, including the elimination of temporal complexities in the $F0$ range modulation pattern, increased modulation depth, the consistency in the modulation pattern across all channels, and the shape of the modulating waveform. The relative importance of these factors remains to be determined, though

these results strongly suggest that the shape of the $F0$ carrying waveform plays an important role. The sharpened sawtooth waveform was clearly superior to sinusoidal modulation in experiment 2, and appeared to be equally as effective as the single pulse per period modulation in experiment 4, which is likely to be optimal for the transmission of temporal pitch cues. It seems likely that a rapid onset of each period of the modulation envelope is particularly important in enhancing the salience of temporal cues.

In line with previous evidence regarding the detectability and discriminability of amplitude modulations of noise carriers in normally hearing listeners and of pulse trains in implant users, temporal voice pitch cues were less effective in the higher $F0$ range. This decline was evident in both modified and standard processing conditions and in both simulations and implant users. Benefits of modified processing were, though, apparent in both $F0$ ranges.

Another factor that clearly limits the utility of temporal pitch cues is the presence of dynamic slow-rate spectral variation such as the changes in spectral envelope caused by the changing formant structure of the diphthongal vowel glides. Green, Faulkner, and Rosen (2002) showed that for simple stimuli such as sawtooth waveforms, in which spectral shape is completely determined by $F0$, shifts in spectral envelope arising from harmonics of the input passing between analysis bands could be used to identify the direction of movement of $F0$ in the absence of temporal cues. Comparison of experiments 3 and 4 suggests that the presence of spectral shifts that are uncorrelated with $F0$ leads to a reduction in the effectiveness of temporal pitch cues for implant users. It is possible that this results from a conflict between temporal and spectral information at a relatively central processing level. However, there is some previous evidence that points to the independence of spectral and temporal information both in implant users and normally hearing listeners (Demany and Semal, 1993; Green, Faulkner, and Rosen, 2002; McKay, McDermott, and Carlyon, 2000; Tong *et al.*, 1983). An alternative possibility is that the better pitch perception in the absence of spectral envelope changes arises from the fact that there is a greater consistency in the modulation pattern within each channel. When dynamic spectral variation is present, as the level in a channel changes, so does the absolute modulation depth of the $F0$-related fluctuations. Such changes may hinder the effectiveness of temporal pitch cues.

With regard to the influence of spectral contributions to pitch information it is worth noting that the vowel glides used in the current studies were synthesized with the same formant frequency trajectories and bandwidths, regardless of variations in $F0$. This differs from natural speech in which there typically is some correlation between $F0$ and formant values. This could result in some useful cues to pitch in certain situations, for example, in distinguishing between male and female speakers. It seems unlikely, however, that tasks involving the identification of an intonation contour would be much affected by such correlations.

B. Implications for cochlear implant speech processing

Although the modified processing method used here clearly does provide significantly enhanced temporal cues to voice pitch, it must be acknowledged that the improvements in glide labeling relative to standard processing were not particularly large and performance with the modified processing scheme was still quite limited. The current scheme represented the exact periodicity of the input with what appears to be a close to optimal modulating waveform shape, and the maximum possible modulation depth while retaining across-channel level differences. It is, therefore, difficult to see how $F0$ could be more clearly represented in the pulses delivered to the electrode array. It is conceivable that somewhat larger advantages relative to standard processing strategies might be apparent after greater experience with the modified processing scheme, or with more natural speech stimuli, or in other pitch perception tasks. Since much of the important information related to voice pitch is conveyed by the pattern of pitch change rather than absolute pitch values, it is also possible that the advantages of the modified processing scheme could be maximized by lowering the rate of the simplified $F0$-related modulation so that it falls within a frequency range that gives better temporal encoding of pitch. Fourcin *et al.* (1984) showed such an approach to be effective in the context of a single electrode device providing $F0$ information as an aid to lip-reading. Nonetheless, on balance, it appears that the scope for improving pitch perception through manipulations of temporal envelope cues is limited.

Two major factors are likely to determine whether these limited improvements in pitch perception can be successfully incorporated into an improved everyday speech processing strategy. Firstly, the $F0$ extraction process will need to be sufficiently reliable. Even the best algorithms currently available tend to become unreliable in the presence of large amounts of noise so that one important consideration will be the need for a kind of gentle degradation, which might involve simply shutting down the $F0$ extraction process and presenting only the low-rate envelope information. Secondly, it is essential that the gains in pitch perception are not achieved at the expense of spectral speech information. Comparisons of performance with the modified and standard processing schemes in standard speech recognition tasks will be reported in the near future, in addition to tests of pitch perception involving natural speech stimuli and more realistic tasks.

ACKNOWLEDGMENTS

The authors thank all participants and are grateful to the cochlear implant team at the Royal National Throat Nose and Ear Hospital for help in recruiting implant users. This research was supported by the Royal National Institute for Deaf People (UK). Ken Grant, Fan-Gang Zeng, and an anonymous reviewer provided helpful comments on a previous version of this paper.

[1]There is some evidence that direct encoding of $F0$ in the carrier rate can be beneficial for tonal language perception. In a case report on a Chinese user

of an implant system in which pulse rate was controlled by $F0$, Xu, Dowell, and Clark (1987) reported good performance in tests of lexical tone perception. Also, in an acoustic simulation study using sinusoidal carriers, Lan *et al.* (2004) found better tone recognition when the frequency of the carriers in each band was dynamically varied as a function of $F0$, compared to when carrier frequencies were constant.

[2]Here, and also in subsequent experiments, a large majority of fits did not deviate significantly from the observed data, according to χ^2 tests with 10 degrees of freedom. The exceptions tended to occur when there were there were relatively substantial departures from monotonicity at larger $F0$ ratios, such as can be seen in Fig. 3 in the mean data for the CIS condition with 226 Hz center $F0$.

Bosman, A. J., and Smoorenburg, G. F. (**1997**). ''Evaluation of three pitch tracking algorithms at several signal-to-noise ratios,'' Acustica **83**, 567–571.

Burns, E. M., and Viemeister, N. F. (**1976**). ''Nonspectral pitch,'' J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). ''Played-again SAM: Further observations on the pitch of amplitude-modulated noise,'' J. Acoust. Soc. Am. **70**, 1655–1660.

Busby, P. A., Tong, Y. C., and Clark, G. M. (**1993**). ''The perception of temporal modulations by cochlear implant patients,'' J. Acoust. Soc. Am. **94**, 124–131.

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (**1994**). ''Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants,'' J. Acoust. Soc. Am. **96**, 2048–2054.

Ciocca, V., Francis, A. L., Aisha, R., and Wong, L. (**2002**). ''The perception of Cantonese lexical tones by early-deafened cochlear implantees,'' J. Acoust. Soc. Am. **111**, 2250–2256.

Demany, L., and Semal, C. (**1993**). ''Pitch versus brightness of timbre: Detecting combined shifts in fundamental and formant frequency,'' Music Percept. **11**, 1–14.

Donaldson, G., and Viemeister, N. (**2002**). ''TMTFs in cochlear implant users: The role of loudness clues,'' Assoc. Res. Otolaryngol. Abstr. **25**, 121.

Faulkner, A., Rosen, S., and Smith, C. (**2000**). ''Effects of the salience of pitch and periodicity information on the intelligibility of four-channel vocoded speech: Implications for cochlear implants,'' J. Acoust. Soc. Am. **108**, 1877–1887.

Fishman, K., Shannon, R., and Slattery, W. (**1997**). ''Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant,'' J. Speech Lang. Hear. Res. **40**, 1201–1215.

Fourcin, A., Douek, E., Moore, B., Abberton, E., Rosen, S., and Walliker, J. (**1984**). ''Speech pattern element stimulation in electrical hearing,'' Arch. Otolaryngol. **110**, 145–153.

Fourcin, A. J., Rosen, S. M., Moore, B. C. J., Douek, E. E., Clarke, G. P., Dodson, H., and Bannister, L. H. (**1979**). ''External electrical stimulation of the cochlea: Clinical, psychophysical, speech-perceptual and histological findings,'' Br. J. Audiol. **13**, 85–107.

Friesen, L., Shannon, R., Baskent, D., and Wang, X. (**2001**). ''Speech recognition in noise as a function of spectral channels: Comparison of acoustic hearing and cochlear implants,'' J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J., and Zeng, F.-G. (**2000**). ''Identification of temporal envelope cues in Chinese tone recognition,'' Asia Pac. J. Speech, Lang. Hearing **5**, 45–57.

Fu, Q.-J., Zeng, F.-G., Shannon, R. V., and Soli, S. D. (**1998**). ''Importance of tonal envelope cues in Chinese speech recognition,'' J. Acoust. Soc. Am. **104**, 505–510.

Fu, Q.-J., Shannon, R. V., and Wang, X. S. (**1998**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.

Geurts, L., and Wouters, J. (**2001**). ''Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants,'' J. Acoust. Soc. Am. **109**, 713–726.

Grant, K., Summers, V., and Leek, M. (**1998**). ''Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **104**, 1051–1060.

Green, T., Faulkner, A., and Rosen, S. (**2002**). ''Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants,'' J. Acoust. Soc. Am. **112**, 2155–2164.

Hanna, T. (**1992**). ''Discrimination and identification of modulation rate using a noise carrier,'' J. Acoust. Soc. Am. **91**, 2122–2128.

Jusczyk, P. (**1997**). *The Discovery of Spoken Language* (MIT, Cambridge, MA).

Lan, N., Nie, K. B., Gao, S. K., and Zeng, F.-G. (**2004**). ''A novel speech processing strategy incorporating tonal information for cochlear implants,'' IEEE Trans. Biomed. Eng. **51**, 752–760.

Lee, K. Y. S., van Hasselt, C. A., Chiu, S. N., and Cheung, D. M. C. (**2002**). ''Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children,'' Int. J. Pediatr. Otorhinolaryngol. **63**, 137–147.

McKay, C. M., and McDermott, H. J. (**1993**). ''Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP),'' Ear Hear. **14**, 350–367.

McKay, C. M., McDermott, H. J., and Carlyon, R. P. (**2000**). ''Place and temporal cues in pitch perception: are they truly independent?'' ARLO **1**, 25–30.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). ''Pitch percepts associated with amplitude-modulated current pulse trains by cochlear implantees,'' J. Acoust. Soc. Am. **96**, 2664–2673.

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1995**). ''Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: The effect of modulation depth,'' J. Acoust. Soc. Am. **97**, 1777–1785.

McKay, C. M., McDermott, H. J., Vandali, A. E., and Clark, G. M. (**1992**). ''A comparison of speech perception of cochlear implantees using the Spectral Maxima Sound Processor (SMSP) and the MSP (Multipeak) processor,'' Acta Oto-Laryngol. **112**, 752–761.

Pollack, I. (**1969**). ''Periodicity pitch for white noise—fact or artefact,'' J. Acoust. Soc. Am. **45**, 237–238.

Rosen, S. (**1989**). ''Temporal information in speech and its relevance for cochlear implants,'' in *Cochlear Implant: Acquisitions and Controversies*, edited by B. Fraysse and N. Cochard (Cochlear AG, Basel), pp. 3–26.

Seligman, P., and McDermott, H. (**1995**). ''Architecture of the Spectra 22 speech processor,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 139–141.

Shannon, R. (**1992**). ''Temporal modulation transfer functions in patients with cochlear implants,'' J. Acoust. Soc. Am. **91**, 2156–2164.

Skinner, M. W., Fourakis, M. S., Holden, T. A., Holden, L. K., and Demorest, M. E. (**1999**). ''Identification of speech by cochlear implant recipients with the Multipeak (MPEAK) and Spectral Peak (SPEAK) speech coding strategies,'' Ear Hear. **20**, 443–460.

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.

Tong, Y. C., Millar, J. B., Clark, G. M., Martin, L. F., Busby, P. A., and Patrick, J. F. (**1980**). ''Psychophysical and speech perception studies on two multiple channel cochlear implant patients,'' J. Laryngol. Otol. **94**, 1241–1256.

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.

Walliker, J. R., and Howard, I. S. (**1990**). ''Real-time portable multilayer perceptron voice fundamental-period extractor for hearing-aids and cochlear implants,'' Speech Commun. **9**, 63–72.

Wei, C.-G., Zeng, F.-G., Cao, K., and Wang, Z. (**2001**). ''Rate discrimination and tone recognition in Chinese cochlear-implant users,'' Assoc. Res. Otolaryngol. Abstr. **24**, 293.

Whalen, D. H., and Xu, Y. (**1992**). ''Information for Mandarin Tones in the amplitude contour and in brief segments,'' Phonetica **49**, 25–47.

Wilson, B. (**1997**). ''The future of cochlear implants,'' Br. J. Audiol. **31**, 205–225.

Wilson, B., Finley, C., Lawson, D., Wolford, R., Eddington, D., and Rabinowitz, W. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.

Wygonski, J. J., Lee, J., Faltys, M., Shannon, R. V., and Robert, M. (**1999**). ''Configurable speech strategy implementation using the Clarion research interface,'' 1999 Conference on Implantable Auditory Prostheses, Pacific Grove, CA.

Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). ''Features of the stimulation affecting tonal-speech perception: Implications for cochlear prostheses,'' J. Acoust. Soc. Am. **112**, 247–258.

Xu, S. A., Dowell, R. C., and Clark, G. M. (**1987**). ''Results for Chinese and English in a multichannel cochlear implant patient,'' Ann. Otol. Rhinol. Laryngol. Suppl. **96**, 126–127.

Zeng, F.-G. (**2002**). ''Temporal pitch in electric hearing,'' Hear. Res. **174**, 101–106.

Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task

Michael A. Stone[a] and Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

(Received 24 September 2003; revised 26 April 2004; accepted 30 June 2004)

Using a cochlear implant simulator, Stone and Moore [J. Acoust. Soc. Am. **114**, 1023–1034 (2003)] reported that wideband fast-acting compression led to poorer intelligibility than slow-acting compression in a competing speech task. Compression speed was varied by using different pairs of attack and release times. In the first experiment reported here, it is shown that attack times less than about 2 ms in a wideband compressor are deleterious to intelligibility. In experiment 2, fast wideband compression was applied to the target and background either before or after mixing. The former reduced the modulation depth of each signal but maintained the independence between the two signals, while the latter introduced ''comodulation.'' Using simulations with 6 and 11 channels, intelligibility was higher when compression was applied before mixing. In experiment 3, wideband compression was compared with multichannel compression; the latter led to reduced comodulation effects. For 6 channels, the position of the compressor, either wideband or within each channel, had no effect on intelligibility. For 11 channels, channel compression severely degraded intelligibility compared to wideband compression, presumably because of the greater reduction of across-channel contrasts. Overall, caution appears necessary in the use of fast-acting compression in cochlear implants, so as to preserve intelligibility.
[DOI: 10.1121/1.1784447]

PACS numbers: 43.66.Ts, 43.71.Gv, 43.66.Mk [DDO] Pages: 2311–2323

I. INTRODUCTION

Cochlear implants all contain some form of compression system to map the wide range of sound levels encountered in everyday life into the limited dynamic range (6–20 dB) available at each electrode (Clark *et al.*, 1990; Moore, 2003a; Zeng, 2004). There are many parameters that may be varied in the design of compression systems, among which are the attack and release times, which determine whether the compression should be described as ''fast'' (also called syllabic compression) or ''slow'' (also called automatic volume control). For reviews, see Dillon (1996) and Moore (1998).

We (Stone and Moore, 2003a) have previously described the use of a ''noise-vocoder'' simulation of a cochlear implant (Shannon *et al.*, 1995) to assess the effect of the speed of response of a wideband (single-channel) dynamic range compressor on speech intelligibility in a competing speech task. The compressors assessed were representative of those used in commercial cochlear implant systems. We found that fast-acting compression degraded speech intelligibility when compared to slow-acting compression: the slow-acting compression itself produced no significant degradation compared to no compression at all. We suggested that the reduction could be explained by two factors: (1) Fast compression introduces correlated fluctuations in amplitude in different frequency bands, which may promote perceptual fusion of the target and background sounds (Bregman *et al.*, 1985; Bregman, 1990; Hall and Grose, 1990; Carrell and Opie, 1992; Moore *et al.*, 1993; Darwin and Carlyon, 1995); (2) Fast compression reduces amplitude modulation depth and intensity contrasts; the importance of this effect has been the subject of much debate (Villchur, 1973; Plomp, 1988; 1994; Moore, 1990; 2003b; Dillon, 1996). For brevity, we refer to factor (1) as ''comodulation'' and factor (2) as ''modulation reduction.''

The comodulation introduced by fast-acting single-channel compression is actually of a somewhat unusual type. When two fluctuating signals (e.g., two talkers) with a similar overall level are mixed, the gain at any instant is determined mainly by the level of the signal that has the higher level. Peaks in one signal result in a reduction in gain (although the peak remains a peak following compression) that decreases the output level of the other signal. Hence, the comodulation of the wideband stimuli is out of phase; increases in level of one signal result in decreases in level of the other signal, and vice versa. While the perceptual grouping of two signals appears to be affected by the similarity in the envelopes of the signals, there is little evidence to support the idea that the relative phase of modulation of the signals is important, and there is some evidence to the contrary. For example, the perceptual segregation of pairs of vowels is not influenced by the relative phase of amplitude modulation of the vowels (Summerfield and Culling, 1992). Also, modulation detection interference, which is believed to result at least partly from perceptual grouping processes (Hall and Grose, 1991; Moore and Shailer, 1992), is not greatly affected by the phase of modulation of the interferer relative to that of the target (Moore and Shailer, 1992; Moore, 1992; Hall *et al.*, 1995). Thus, the type of comodulation introduced by a single-channel fast-acting compressor

[a]Electronic mail: mas19@cam.ac.uk

could decrease perceptual segregation even though it is out of phase.

When considering the effects of comodulation, another factor needs to be taken into account. The gain of a wideband compressor is controlled by the components of the input that are most intense, and these tend to be the lower-frequency components of speech signals. When the signal is split into frequency channels (by the auditory system or by an implant processor), the temporal fluctuations in channels passing the most intense components may be almost independent of those in remote (typically high-frequency) channels. Under these conditions, the target and background become comodulated *in phase* in the remote channels, because the same gain changes are imposed on the target and background. Thus, the pattern of comodulation is actually rather complex and varies depending on the nature, number, and center frequency of the channels involved, as well as on the statistical properties of the signal.

In our earlier experiments, we compared two systems with markedly different attack and release times, but with the same (high) compression ratio. Since both attack and release times were varied, we were not able to assess whether there were any specific effects associated with a fast attack time. However, there are theoretical reasons, discussed below, for thinking that such effects might occur. The first experiment reported here investigated the effect on intelligibility of varying the attack time of a wideband (single-channel) compressor, over the range 0.125 to 8 ms, while keeping other aspects of the compressor performance as similar as possible across conditions.

The data reported in our earlier experiments did not allow us to assess the relative importance of comodulation and modulation reduction. Experiments 2 and 3 were intended to clarify this issue. Experiment 2 compared the effect on intelligibility of wideband compression applied before or after mixing the target signal with the background interference. Compression before mixing produces modulation reduction without comodulation, while compression applied after mixing produces both modulation reduction and comodulation. The third experiment compared the intelligibility produced by use of a single, wideband compressor to that produced by a system employing multichannel independent compression, for a simulated implant with either 6 or 11 channels. The single wideband compressor produced both modulation reduction and comodulation, while the multichannel independent compression produced only limited comodulation; the comodulation occurred within but not across channels. If comodulation is an important effect, we would expect better performance with multichannel than with single-channel compression. However, other factors may influence the outcome, as discussed later.

II. EXPERIMENT 1: EFFECT OF ATTACK TIME ON INTELLIGIBILITY

A. Introduction and rationale

For an acoustic hearing-aid application, ANSI (1996) defines the attack time as the time taken for the output of the compressor to settle to within 2 dB of its steady state value after a 25-dB step increase in level at its input. In some systems, it is important to avoid high signal levels so as to avoid distortion (as in broadcasting or in hearing aids) and/or discomfort (as in hearing aids or cochlear implants). In such systems, it has been common to use fast attack times in the compressors. This ensures that the gain reacts rapidly to sudden increases in the input level. However, there are at least two reasons why a fast attack time might have detrimental effects; these are discussed below.

A fast change in gain is equivalent to a fast amplitude modulation of the signal being controlled. The fast modulation generates spectral sidebands around the frequency components in the original signal, which, in an acoustic aid, may momentarily mask other components present in the signal. In a cochlear-implant system, a fast-acting wideband compressor acting before the separation into channels may lead to spectral components within one or more channels that were not present in the original signal. The spread in frequency of the spectral sidebands is an inverse function of the attack time. The level of the sidebands increases with increasing compression ratio; high compression ratios produce large changes in gain. For input modulation rates slower than the reciprocal of the sum of the attack and release times, the gain changes can be almost as large as the signal modulation depth (Braida *et al.*, 1982; Stone and Moore, 1992). If the attack time is made longer, this reduces the level and spread of the sidebands, but the sidebands last for a longer time. However, with sufficiently long attack times, the spread becomes negligible. Hence, the deleterious effects of spectral spreading are likely to be greater for short attack times.

A second possible detrimental effect of short attack times in a wideband compressor is connected with the comodulation effect described in the Introduction. At the onset of a sound, a fast attack time results in a rapid decrease in gain, so any co-occurring sound has an abrupt decrease in level imposed on it. In other words, an abrupt increase in level of one sound becomes associated with an abrupt decrease in level of the other. The faster the attack time of the compressor, the more likely it will be to ''track'' rapid onsets in level of the input signal. There is evidence that the auditory system contains a mechanism for detecting abrupt changes in level that is not sensitive to the polarity of the change (Macmillan, 1973; Hafter *et al.*, 1996). Also, synchrony of onsets has been shown to be a powerful cue for perceptual grouping (Rasch, 1978; Bregman, 1990; Darwin and Carlyon, 1995). The synchronous abrupt changes in the two sounds produced by gain changes in the compressor may thus promote perceptual fusion of the two sounds even though the changes are in opposite directions. Also, as described in the Introduction for the general case of comodulation, when the signal is split into channels, the abrupt changes may actually be in phase for channels remote from those passing the most intense components.

In this experiment, we varied the attack time of the wideband compressor over a wide range and investigated its effect on intelligibility using a noise-vocoder simulation of a cochlear implant. To isolate the effects described above from effects related to modulation reduction, small adjustments to

the release time were also applied, in such a way that modulation reduction was held constant as the attack time was varied. Because of the number of conditions assessed, the number of channels in the implant simulation during the main test was fixed at a value (8) similar to the number of independent channels typically achieved in cochlear implants (Clark *et al.*, 1990; Friesen *et al.*, 2001; Moore, 2003a).

B. Signal processing for simulation of a cochlear implant

The signal-processing software, written in MATLAB,™ was the same as used by Stone and Moore (2003a). Briefly, the signal to be processed was ''pre-emphasized'' by applying a gain rising at 3.3 dB/octave between 500 and 4000 Hz, giving a total of 10 dB. Below 500 Hz the gain was 0 dB, and above 4000 Hz it was 10 dB. The signal was then filtered into one of a specified number of channels. All filters were implemented using a finite-impulse-response (FIR) method, which introduced a frequency-independent time delay. They were designed such that the low-pass edge of one filter had the complementary response to the high-pass edge of the next higher filter. The outputs of all filters were time aligned. Addition of the channel outputs gave virtually perfect recombination: within the frequency range 100 to 7800 Hz, the ripple was less than ± 0.2 dB.

After filtering, the channel envelopes were extracted by full-wave rectification and low-pass filtering. The channel envelopes were each used to modulate a broadband white noise, after which the modulated noises were bandlimited to the same widths as those of the corresponding analysis filters. Time delays were added to the channel signals to compensate for the time delays introduced by the FIR filters (these delays varied with center frequency), such that the relative timing of envelope signals across frequency was the same for the input and the output. The filtered noise bands were then added back together and a linear-phase 6-pole infinite impulse response high-pass filter with a corner frequency of 100 Hz was applied to remove spurious low-frequency noise.

C. Subjects and equipment

Twenty volunteer subjects (6 males, 14 females, aged 18–41 years), all university undergraduates or graduates, were selected on the basis of their having audiometric thresholds $\leqslant$15 dB HL at octave frequencies between 125 and 8000 Hz and at 3000 and 6000 Hz. None had prior experience of the processing. All were native speakers of British English. Subjects attended one session, which lasted nearly 2 h after initial audiometric screening. Subjects were paid for their attendance. The equipment was the same as used by Stone and Moore (2003a). Signals were generated via a high-quality sound card in a PC, passed through a mixing desk, and presented diotically through Sennheiser HD580 headphones. Each subject was seated in a sound-isolated, double-walled chamber.

The speech material for both target and background was the same as used by Stone and Moore (2003a). The target sentences were from the ASL corpus (MacLeod and Summerfield, 1990), and were spoken by a male speaker of British English. Additionally, four lists of 17 sentences each were prepared from recordings by a female speaker of British English of lists 43 to 57 of the IEEE sentence lists (IEEE, 1968). These four lists were used for training only. All speech (both target and background) was filtered so as to have the same long-term average spectrum as specified in ANSI (1997) for normal conversational levels. The sentence material used for the test proper comprised lists of 15 sentences with 3 keywords in each. The background was continuous running speech produced by a male speaker of British English. Pauses and hesitations were removed by hand editing. The input level of the target speech to the processing was 67 dB SPL (unweighted). The presentation level was 68 dB SPL (unweighted).

D. Procedure

Since the processed signals were novel to the subjects, a training period was necessary before testing proper began. Stone and Moore (2003a) reported that a training period of at least one-half hour was necessary to achieve stable results, but even after this, some learning effects were apparent in the counterbalanced data. Compared to Stone and Moore (2003a), we lengthened the initial training period and made it more interactive.

The initial training comprised two phases. In the first phase, four IEEE sentence lists were presented using progressively decreasing target-to-background ratios (99 to 10 dB) and decreasing numbers of channels (16 to 8). If, after the second presentation of a sentence, the keywords were still not repeated correctly, the experimenter spoke the correct answer and then made a further presentation to allow the subject to verify the answer. The background speaker was not the same as used in the test proper.

In the second phase of initial training, the subject was familiarized with the target and background speakers for the test proper. Two lists were presented, one processed as 8 channels with a $+10$-dB target-to-background, and the second as 8 channels with a $+5$-dB target-to-background. These 30 training sentences were not used again either for training or testing. The processing conditions used in the initial training were varied in a counterbalanced order across subjects. This completed the initial training, but before each condition was tested, there was an additional training phase. Eighteen further sentences were presented up to three times, using 8 channels and a $+2$-dB target-to-background ratio.

After all the training, the intelligibility test was conducted at a signal-to-background ratio of $+2$ dB. A total of 33 target sentences (which had not been used in the training) was used for each condition, with 3 keywords in each. For each presentation of a target sentence, the background was ramped up over 0.25 s, the ramp starting 1 s before the target sentence. After the target sentence was completed, the background continued for about a further 0.5 s before being ramped down over 0.25 s. Subjects were encouraged to respond after each presentation, even if the sentence appeared to be nonsense. Five conditions differing in the compression system used were tested in a counterbalanced order across subjects.

E. Equating compression conditions

The primary purpose of this experiment was to compare four fast-acting compression systems similar to the fast-acting system used by Stone and Moore (2003a), but having attack times of 0.125, 0.5, 2, and 8 ms. A fifth system was also tested which was intended to be more effective than the other systems in reducing envelope modulations over the range of modulation rates important for speech perception (Plomp, 1983; Drullman *et al.*, 1994a; 1994b). This system is described in more detail later. The fast-acting system used by Stone and Moore (2003a) had an attack time of 2 ms, a release time of 240 ms, a compression ratio of 7, and a compression threshold of 55 dB SPL. We used the same compression ratio and compression threshold as before. The compression threshold is lower than usually employed in cochlear implant processors, but is not unrealistically low. The relatively low threshold was chosen to ensure that the compressor was active during presentation of the target speech.

The minimum attack time of 0.125 ms is close to the instantaneous attack time used to prevent overload in some electrical circuits, for example as a limiter in broadcasting. The maximum attack time is similar to the maximum value found in fast-acting compressors in commercial hearing aids. When a relatively long attack time is used with such a high compression ratio, it is desirable to incorporate a small delay to the audio path, typically about half the attack time in duration, so as to prevent large ''overshoots'' at signal onsets. Such a delay has been included in experimental acoustic hearing aids (Robinson and Huntington, 1973; Bustamante and Braida, 1987; Baer *et al.*, 1993; Verschuure *et al.*, 1996; Stone *et al.*, 1999). Here, a delay equal to half the attack time was inserted into the audio path for all the configurations of ''fast'' compression. The maximum delay of 4 ms might be subjectively noticeable in some situations (Agnew and Thornton, 2000), but it would not be subjectively disturbing or objectively degrading (Stone and Moore, 2002; 2003b).

In this experiment, we wished to isolate the effects of fast attack times on spectral spreading and comodulation of onsets from the effect of modulation reduction. To do this, we equated modulation reduction across conditions by making small adjustments to the release time. Stone and Moore (2003a) described a metric, called the fractional reduction in modulation, f_r, which allowed comparison of the effect of different compression systems on envelope modulations. Typically, f_r is a function that varies with envelope frequency, and it depends on both the attack and the release time of a compression system. In order to match f_r between the systems used here, the release time, t_r (ms), was set to

$$t_r = 244 - 2t_a, \tag{1}$$

where t_a is the attack time (ms). For the range of attack times tested here, the release time varied only over a small range (228–243.75 ms). The value of f_r for the four fast compression systems is plotted as a function of envelope modulation frequency in Fig. 1 (curve marked ''Fast''); the curve was almost identical for the four attack times used.

The fifth system tested, mentioned earlier, was similar to that of Robinson and Huntington (1973) and of one of the compressors used in the ''FULL-4'' system of Stone *et al.* (1999). The envelope of the signal was extracted and passed through a 2-pole Bessel-derived low-pass filter, resulting in nearly symmetric attack and release times of 10 and 13 ms, respectively. Note that the attack time is similar to that for the fast system with the longest attack time (8 ms). In order to achieve the near-symmetry of attack and release times for the fifth system, the audio signal was delayed by 6.3 ms before the gain signal was applied. Again, this delay has no significant perceptual effects. The low-pass filter design had negligible overshoot in its step response, and the two-pole design, with a corner frequency of 24 Hz, ensured that the filter output was very low for envelope modulation rates corresponding to voice fundamental frequencies (f_r was nearly zero at an envelope rate of 100 Hz). This envelope compressor achieved the ''target'' f_r for envelope rates up to about 10 Hz. The resulting compressed signal had intermodulation distortion less than about 3% (Moore *et al.*, 1999). The compression ratio, 2.78, was chosen to be equal to the effective compression ratio of the fast-acting systems, for an envelope modulation rate of 2 Hz. This rate corresponds to the lower limit of ''useful'' envelope modulation rates in speech, as found by Drullman *et al.* (1994a; 1994b). The value of f_r for the envelope compression system is plotted as a function of envelope modulation frequency in Fig. 1 (curve marked ''Envelope''). The envelope compressor achieves far more effective compression than the ''fast'' compressors over the range of envelope modulation rates significant for speech perception (Drullman *et al.*, 1994a; 1994b). We therefore expected the envelope compressor to produce lower intelligibility than any of the fast compressors. The compression threshold of the envelope compressor was the same as for the fast compressors, namely 55 dB SPL.

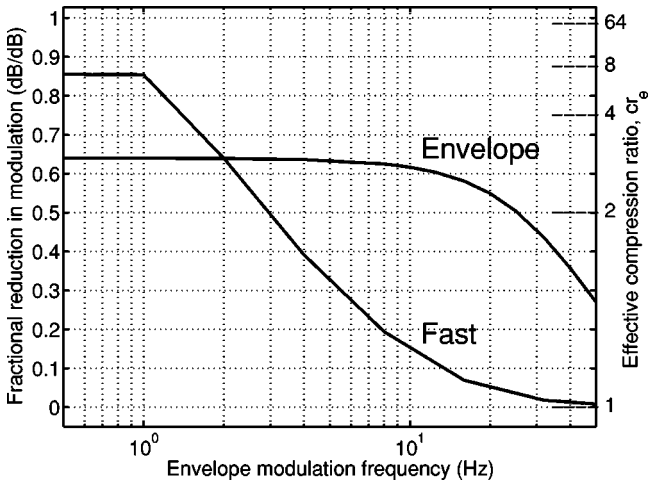


FIG. 1. Effect on envelope modulations of the two compressor types under test, as a function of the modulation rate of the input. The left ordinate shows the fractional reduction in modulation and the right ordinate shows the effective compression ratio. The curve marked ''Fast'' shows results for all four attack times used with the fast compressor. The curve marked ''Envelope'' shows results for the envelope compressor.

Stone and Moore (2003a) used a target-to-background ratio of +5 dB throughout their three experiments, and, for an 8-channel system, the average score was around 70%–75%. Typical slopes of psychometric functions for the intelligibility of speech against an interfering voice have been

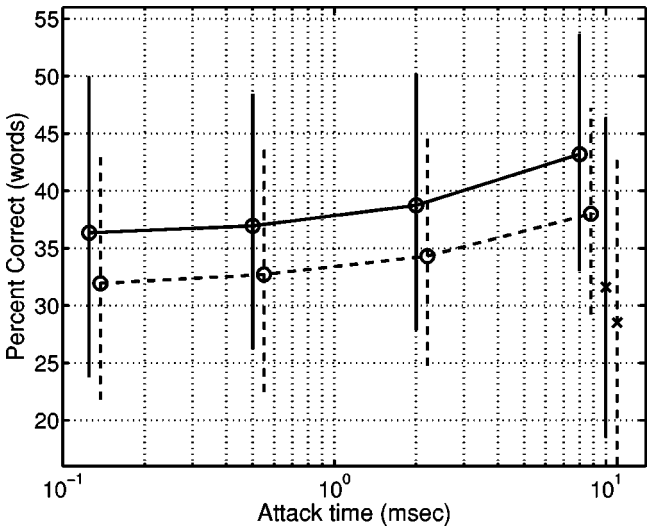


FIG. 2. Results of experiment 1. The solid and dashed lines show data corrected and uncorrected for learning effects, respectively. The connected lines show data for the fast compressors. The separate crosses show results for the envelope compressor. Error bars show ±1 SD. The target-to-background ratio was 2 dB, and eight channels were used in the simulation.

reported as between 7% and 13%/dB around the 50%-point on the psychometric function (Baer and Moore, 1994; Festen and Plomp, 1990). To avoid ceiling effects and to bring the scores nearer to 50%, the target-to-background ratio was reduced to +2 dB for these experiments.

F. Results

The scores were transformed into rationalized arcsine units (RAU; Studebaker, 1985). This was done to make the variance more uniform across conditions. Despite the training given, there was evidence that performance increased across successive conditions, i.e., a further learning effect occurred. To test for this, the scores for each subject were divided by the average score for that subject across the five processing conditions. Then, the score for each condition for a given subject was divided by the mean score for that condition across subjects. The resulting ''normalized'' data were then time ordered and the mean and standard deviation of the scores were calculated across all subjects, for each time slot. Using a *t*-test, we compared the means for the first and last conditions presented. There was a highly significant difference between the two ($t=3.52$, 38 df, $p<0.005$, one-tailed), implying that a learning effect was present.

The time-ordered averaged scores were expressed relative to the average score obtained in the last condition. The resulting relative scoring rates for the first four conditions tested (in time order) were 0.72, 0.91, 0.92, 0.90. The learning effects, if uncorrected, would make the data more ''noisy,'' and might obscure the effects of the different conditions. Hence, to correct for the learning effect, the scores for each of the first four conditions tested for each subject were divided by the relative score for that condition, e.g., the score for the second condition tested was divided by 0.91.

The solid line in Fig. 2 shows the mean learning-corrected word intelligibility scores, transformed back from RAUs into percentages, as a function of the attack time of the fast compressor. The mean corrected score for the envelope compressor is shown as a single cross with solid error bars at the right of the figure. Corresponding uncorrected scores are shown by dashed lines. The error bars show standard deviations (SDs) across subjects. The mean scores tend to improve with increasing attack time for the fast systems. The mean score is somewhat lower for the envelope compressor than for any of the fast systems.

The learning-corrected scores (in RAUs) were subjected to a one-way analysis of variance (ANOVA), with factor compression system. The ANOVA showed a significant effect of compression system; $F(4,76)=5.03$, $p=0.001$. Based on Fisher's least significant differences test, the difference between scores for the fast compressors for attack times of 8 and 2 ms was not significant (one-tailed *t*-test, $t=1.66$, 19 df, $p\cong 0.067$). The difference between the fast compressor with an attack time of 8 ms and all the remaining compression systems was significant at $p<0.025$ ($t>2.33$, 19 df, one-tailed). As expected, the score for the ''envelope'' compressor was significantly worse than for any of the fast compressors, $t\geqslant 1.83$, $p<0.05$ (one-tailed).

Although we were aiming to get overall scores around 50%, the scores were generally lower than this. The target-to-background ratio was lower than used in our earlier experiment, which had given scores around 70%–75% (Stone and Moore, 2003a). Evidently, we reduced the target-to-background ratio too much, possibly because we assumed a slope for the psychometric function that was based on experiments using unprocessed speech in noise. Also, psychometric functions for speech in fluctuating backgrounds, processed using a noise-vocoder simulation, can vary markedly in slope across subjects (Qin and Oxenham, 2003). The large error bars in Fig. 2 reflect a near-bimodal distribution in subject mean scores. Splitting the learning-corrected mean scores of each subject into bins of 5-RAU width, seven subjects scored 20–25, one scored 25–30, seven scored 35–40, and the remaining five scored 40 or more. We have not previously observed such a large spread in individual means in a nominally homogeneous group.

Despite the ''noise'' in the data, the results do show that decreasing the attack time from 8 to 0.5 or 0.125 ms, while holding modulation reduction constant, resulted in a significant reduction in intelligibility. Thus, it would seem desirable to avoid very fast attack times in compression circuits for cochlear implants. On the other hand, the envelope compressor, which had a longer attack time than the other compressors (and a shorter release time), but was more effective in reducing the modulation depth of the signal, produced the lowest intelligibility of all of the compressors. Thus, modulation reduction appears to be an important effect, separate from the effects of spectral spreading and comodulation of onsets resulting from the use of fast attack times.

III. EXPERIMENT 2: ASSESSING THE ROLES OF COMODULATION AND MODULATION REDUCTION

A. Introduction

The aim of this experiment was to assess the relative importance of comodulation and modulation reduction by applying wideband compression to the target and back-

ground, either before or after mixing. The fast compressors used in experiment 1 and by Stone and Moore (2003a) had attack times that were much smaller than the release times. Compressors with asymmetric attack and release times produce large distortion of the shape of the temporal envelope (Stone and Moore, 1992). This distortion on its own may have contributed to the reduction of intelligibility found by Stone and Moore (2003a) for the fast compressor relative to the slow compressor. Since we wished to investigate the role of modulation reduction rather than distortion of envelope shape, we decided to use a form of compression that produced minimal distortion of envelope shape. To do this, we used the envelope compressor of experiment 1, which had near-symmetric attack and release times.

In one condition, the compression was applied before the target and background signals were mixed. This would not be possible in a ''real-world'' environment, but it allowed us to apply modulation reduction to both signals while preserving their independence: comodulation should not then occur. We call this condition ''INDEP.'' In a second condition, the compression was applied after mixing the target and background. This resulted in both modulation reduction and comodulation. We call this condition ''COMOD.'' Each condition was tested using both a 6-channel and an 11-channel simulation; these channel numbers bracket the 8 channels used in experiment 1. On the basis of a pilot trial, target-to-background ratios of +8 and +3 dB were used for the 6- and 11-channel simulations, respectively, in order to achieve similar intelligibility. These ratios led to higher mean scores than for experiment 1.

B. Subjects and equipment

Eight volunteer subjects (4 males, 4 females, aged 19–31 years), all university undergraduates or graduates, were selected on the same basis as for experiment 1, except that one subject was a native speaker of American-, rather than British English. All subjects had previously been exposed to the processing, at times ranging from a few days to 18 months earlier. Subjects attended one session, which lasted just over 1 h. They were paid for their attendance. Since all the subjects had prior exposure to the ASL material, the BKB speech corpus (Bench and Bamford, 1979) was used. The sentences were spoken by a female speaker of British English. The background was based on a male speaker, but a different one from that used for experiment 1. The equipment and signal-processing method were the same as for experiment 1.

The input level of the target signal to the processing was 67 dB SPL (unweighted), and the presentation level of the mixed signal was 68 dB SPL for both conditions, as measured when the target and background were present at the same time. The envelope compressor had a compression ratio of 2.78 and the compression threshold was 55 dB SPL, the same as used in experiment 1.

C. Procedure

The training method for this experiment was almost the same as for experiment 1. The replay of the four IEEE sentence lists was again at increasing levels of difficulty. The second stage of training used the 32 sentences from BKB lists 1 and 2. Consecutive triads of lists drawn from BKB list numbers 3, 5, 6, 8, 9, 10, 11, 13, 14, 18, 19, and 21 were used for assessing each condition. Scores for each condition were the number of keywords correct from 33 sentences. The four conditions (INDEP versus COMOD and two channel numbers) were tested in a counterbalanced design.

D. Equating target-to-background ratios before and after compression

To allow a fair comparison of conditions INDEP and COMOD, it was necessary to ensure that the target-to-background ratios were comparable. In condition INDEP, the target and background were independently compressed, and then had to be mixed at the same target-to-background ratio as used for condition COMOD. This required measurement of the rms value of the uncompressed target and background for condition COMOD, and of the compressed target and background for condition INDEP. However, measurement of the rms value of compressed speech is not straightforward.

If the energy of *uncompressed* speech is estimated for short successive segments (frames), a histogram of the frame energies typically shows two peaks, a main peak defined by the speech energy and a peak at a lower level that represents the background noise in the recording. To measure the ''real'' speech energy, a measurement threshold needs to be selected (note that this threshold has nothing to do with the compression threshold used in the compressor). Only frames with energy above this measurement threshold are used in the calculation of the rms value. Choosing the measurement threshold to lie near the minimum of the valley, Stone and Moore (2003a) found that about 80% of 10-ms frames in continuous running speech were included in their measure of rms value. Compression alters the shape of the histogram of signal levels, which makes it more difficult to estimate an appropriate measurement threshold. Here, we chose the measurement threshold so that the percentage of 10-ms frames included in the measure of rms value was the same as for Stone and Moore, i.e., 80%. The measurement threshold was expressed relative to the level at the output of the compressor for a sinusoidal input with the same rms value as the input speech; this measure is denoted $\mathrm{rms}_{\mathrm{sin}}$. For condition INDEP, the measurement threshold for the target was set to 5 dB below $\mathrm{rms}_{\mathrm{sin}}$. For the background, which was at a lower input level than the target, relatively more of the signal fell below the compression threshold. To allow for this, the measurement threshold was reduced so as to meet the 80% criterion. The measurement threshold for the background was 7 dB below $\mathrm{rms}_{\mathrm{sin}}$ for the +3-dB target-to-background ratio, and about 10 dB for the +8-dB target-to-background ratio.

E. Results

The scores were transformed into RAUs. Each data point was then normalized as described for experiment 1. The data were then time ordered and the mean and standard deviation of the score were calculated for each time slot. Mean scores across time-ordered conditions, expressed rela-

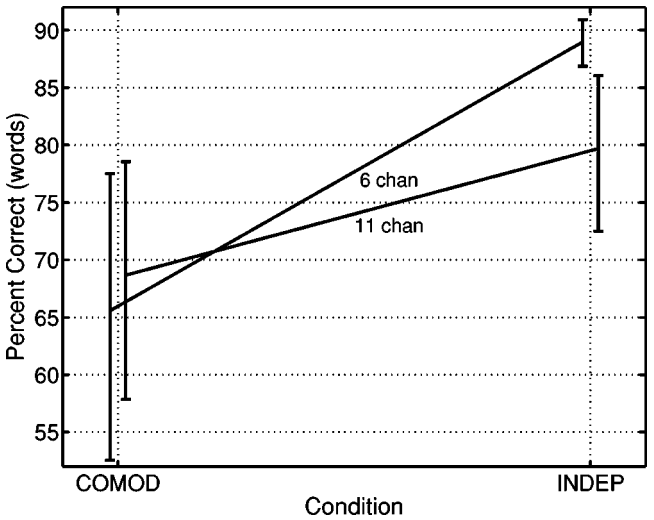


FIG. 3. Results of experiment 2 not corrected for learning effects (which were not significant for this experiment). Error bars show ±1 SD. The target-to-background ratio was 8 dB for the 6-channel system and 3 dB for the 11-channel system.

tive to the mean score for the last condition tested, were 1.03, 0.94, and 0.99. Using a *t*-test, we compared the means for the first and last conditions presented. The difference was not significant ($t=1.71$, 14 df, $p>0.1$, two-tailed). This was not surprising since the subjects were experienced with the processing. Consequently, no compensation was necessary for learning effects.

The mean scores, transformed from RAUs back into percentages, are plotted in Fig. 3. The error bars show SDs across subjects. For both channel numbers, intelligibility was higher for the INDEP condition, where the target and background were mixed after compression.

To assess the significance of the measured effects, an ANOVA was conducted with factors condition (INDEP or COMOD) and number of channels (6 or 11). One potential problem in comparing scores across channel numbers is that the signal-to-background ratio was different for the 6- and 11-channel systems. However, Friesen *et al.* (2001) have shown that psychometric functions for identifying sentences in steady background noise are roughly parallel for systems with different numbers of channels, for channel numbers from 4 to 20. We believe that this makes it reasonable to compare results across channel numbers. There was a significant effect of condition, $F(1,7)=20.67$, $p=0.003$, and a significant interaction between condition and number of channels, $F(1,7)=10.23$, $p=0.015$. The effect of condition was significant for both channel numbers ($t>2.41$, 7 df, $p<0.05$, two-tailed). There was no significant difference between the scores for the 6- and 11-channel systems in condition COMOD ($t=0.89$, 7 df, $p>0.4$), but there was a significant difference between the two channel numbers for the INDEP condition ($t=3.42$, 7 df, $p<0.02$).

The finding that performance was worse in condition COMOD than in condition INDEP supports the idea that at least part of the deleterious effect of fast-acting compression occurs because the compression introduces comodulation between the target and background. The finding that the difference between conditions COMOD and INDEP was larger for 6 channels than for 11 channels can be explained in the following way. In both conditions, the compressors acted on the wideband signal. When a compressed wideband signal is split into a number of frequency channels, the degree of modulation reduction of each channel signal (relative to what would be observed for an uncompressed wideband input signal) is less than that applied to the wideband signal; the channel signals are compressed less than the wideband signal (see experiment 3 and Table I, which is described later). The difference between the wideband modulation reduction and the channel modulation reduction increases with increasing number of channels. Less modulation reduction implies smaller changes in gain over time, which in turn implies less comodulation. This could explain why the difference between conditions INDEP and COMOD was smaller for the 11- than for the 6-channel system.

We conclude that comodulation does contribute significantly to the deleterious effect on intelligibility produced by fast compression, but the influence of comodulation decreases as the number of channels increases.

IV. EXPERIMENT 3: COMPRESSION OF THE WIDEBAND SIGNAL VERSUS COMPRESSION OF THE CHANNEL SIGNALS

A. Introduction

In some cochlear implant systems, the front-end (wideband) compression is followed by a second stage of fast-acting or instantaneous compression applied to individual channels, to map the channel signals into the current range between threshold and the most comfortable level (MCL) at each electrode (Clark *et al.*, 1990; Moore, 2003a; Zeng, 2004). This prevents uncomfortably loud levels (ULL) being reached, even allowing for loudness summation across electrodes. In other systems, all of the compression is performed by a front-end compressor (Eddington, 1983). An alternative configuration of the processing in a cochlear implant would be to have no front-end compressor, and to perform all of the necessary compression using fast-acting compressors within each channel. This approach has two desirable effects. First, comodulation effects would be more restricted in bandwidth. For example, a momentary peak in low-frequency energy of a background sound would lead to a reduction in gain, and a comodulation effect, only for low-frequency channels. Second, the spectral sidebands produced by rapid changes in gain in a given channel would have no influence on other channels. Drawbacks of this approach are that it would reduce modulation depth and intensity contrasts within each channel, *and* would reduce across-channel contrasts, which convey spectral information (Plomp, 1988; Drullman *et al.*, 1996). When a noise-vocoder implant simulator is used, which removes spectral fine structure, subjects may be especially dependent on across-channel contrasts. The same may be true of real cochlear implants.

The purpose of experiment 3 was to compare intelligibility obtained using a single-channel wideband compressor with that obtained using independent compressors within each channel of the implant simulation. For brevity, we refer to these as ''wideband'' compression and ''channel'' com-

TABLE I. Comparison of compression ratios for the wideband and channel compressors of experiment 3. Part A is for the 6-channel system, which used a target-to-background ratio of 5 dB. Part B is for the 11-channel system, which used a target-to-background ratio of 2 dB. The first line in each section shows the channel number. The second line shows the slope of the input–output function for each channel, measured as described in the text, for the system with compression applied to the wideband signal. The third line shows the corresponding effective compression ratio (the reciprocal of the slope). The fourth line shows the nominal compression ratio that was set for each channel of the system with independent channel compression, in order to match the effective compression ratio of the wideband compressor.

Condition	Channel number										
(A) 6-channel system			1	2	3	4	5	6			
Wideband slope			0.49	0.59	0.72	0.61	0.89	1.02			
Wideband effective CR			2.04	1.69	1.39	1.64	1.12	0.98			
Channel nominal CR			4.29	2.34	2.20	2.66	1.57	1.22			
(B) 11-channel system	1	2	3	4	5	6	7	8	9	10	11
Wideband slope	0.55	0.59	0.64	0.74	0.75	0.69	0.66	0.66	0.87	0.90	1.02
Wideband effective CR	1.82	1.69	1.56	1.35	1.33	1.45	1.52	1.52	1.15	1.11	0.98
Channel nominal CR	3.51	2.75	2.20	2.34	2.41	2.57	2.75	2.97	1.83	1.61	1.33

pression, respectively. The simulation used both 6 channels and 11 channels, with target-to-background ratios of +5 and +2 dB, respectively. In order to provide a fair comparison between the two compression arrangements, it was necessary to equate the effective amount of compression for the two. Our procedure for doing this is described next.

B. Equating the effect of a multichannel compression system to that of a single-channel compression system

To make the two conditions comparable, it was necessary to adjust the action of the independent channel compressors so that they produced a mapping of short-term signal levels between input and output similar to that achieved by the single-channel compression system. The steps involved in this are described below.

Two 90-sec-long speech files were generated comprising two male speakers mixed at the same target-to-background ratios as used in the experiment. The speakers were from our corpus of manually edited continuous speech (Stone and Moore, 2003a). First, the implant simulation was performed with no compression active. The level in successive 125-ms rectangular time windows (frames) was measured for each channel signal, prior to the stage of recombination. This provided a statistics file with which the effects of the compression systems were compared. Second, the implant simulation was performed with the single-channel wideband compressor active, and the same analysis was performed. The compressor was the same as one of the fast compressors of experiment 1, having attack and release times of 2 and 240 ms, respectively, a compression ratio of 7, and a compression threshold of 55 dB SPL. For each channel, a scatter plot was produced of the short-term level with compression versus the short-term level without compression (equivalent to a plot of output versus input). As is typical for compression systems (Elberling and Hansen, 1999), the points fell on a straight line (on a dB versus dB scale) only over a small range of levels. A linear regression over this range of levels was performed. Initially, frames were selected which lay in the 30-dB range below the maximum input level for a given channel. The mean and standard deviation of the levels in these frames were calculated. The linear regression was then performed over those frames lying between maximum channel level and the lesser of either 3 dB or 1 standard deviation below the channel mean level. The regression slopes provided a first approximation to the reciprocal of the compression ratio required for the independent channel compressors, so as to provide the same mapping of input to output as for the wideband compressor. These regression slopes are given by the second line of each section in Table I. The corresponding ''effective'' compression ratios are shown in line 3 of each section. The decrease of compression ratio with increasing frequency occurred because the gain of the wideband compressor was largely determined by the more intense low-frequency components in the input speech. This occurred even though, as described earlier, ''pre-emphasis'' was applied before wideband compression, so as to reduce the dominance of low frequencies in setting the gain.

Next, the implant simulation was performed with compression only in the individual channels. The channel compressors had the same attack and release times as the single-channel wideband compressor, but the compression thresholds for each compressor were set 13 dB below the rms value of each channel signal (for the wideband compressor, the compression threshold was 13–14 dB below the rms value of the target plus background at the input to the compressor, the exact value depending on target-to-background ratio). The channel rms value was calculated from the channel envelope signal, which was determined by full-wave rectification of the output of a given channel filter and low-pass filtering using a linear-phase filter with a response that was −3 dB at 50 Hz and −40 dB at 100 Hz (Stone and Moore, 2003a). The channel rms value was calculated in two stages: first, the rms value was calculated for all samples (at the sample rate of 16 kHz) of the channel envelope. Then, the rms value was recalculated using only those samples whose level exceeded the initial rms value−10 dB. Using the same 125-ms frame-selection and measurement criteria as before, a linear regression was performed on the scatter plot of output frame levels versus input frame levels. The channel com-

pression ratios were adjusted iteratively until the regression slopes within each channel were the same for the multichannel compression system and the single-channel compression system. The required compression ratios were larger than the reciprocal of the regression slopes of each channel estimated for the wideband compression system (shown in line 3 of Table I), since the effective compression ratio achieved by the fast compressor was less than the nominal compression ratio (see Fig. 1). The nominal compression ratios needed to achieve the target values are shown in line 4 of each section of Table I.

C. Subjects, equipment, and stimuli

Eight volunteer subjects (3 males, 5 females, aged 20–35 years) were selected on the same basis as for experiment 1, except that three subjects were not native English speakers. These subjects were, however, fluent in English, and the pattern of their results did not differ markedly from that of the other subjects. No subject had participated in experiment 1 or 2, and none was familiar with the speech material or processing method. Subjects were paid for their attendance. The equipment and signal-processing methods were the same as for experiments 1 and 2. The ASL corpus was used. The input level of the target signal to the processing was 67 dB SPL (unweighted), and the presentation level of the mixed signal was 68 dB SPL.

D. Procedure

The four conditions (two arrangements of compression by two channel numbers) were tested in a counterbalanced design. The training procedure was almost identical to that for experiment 1.

E. Reduction of modulation spread in multichannel compression

In the introduction to experiment 1, we discussed how rapid changes in gain produce spectral sidebands around each signal component. In a multichannel acoustic hearing aid, such sidebands could fall outside the passband of the channel containing the signal component that gave rise to the sidebands, resulting in cross-channel effects. In an implant, the output of each channel is fed to a single electrode, so such cross-channel effects do not occur. This was implicitly simulated in our software since, after modulation of the noise carrier by the compressed channel signal, bandpass filtering was used to create the channel signals and to reinstate the long-term average spectral shape of the original speech spectrum plus the pre-emphasis (Stone and Moore, 2003a). The out-of-original channel sidebands were therefore prevented from interfering with adjacent channel signals in our simulation.

F. Results

The scores were transformed into RAUs. We assessed whether there was a learning effect using the same method as for experiments 1 and 2. There was a significant difference between average scores for the first and last conditions ($t=6.16$, 14 df, $p<0.0005$, one-tailed). The normalized scores for the successive conditions were 0.69, 0.88, and 0.95 times the score in the final condition. The change in scores over time is similar to that found for experiment 1. Scores for successive conditions were corrected as described for experiment 1.

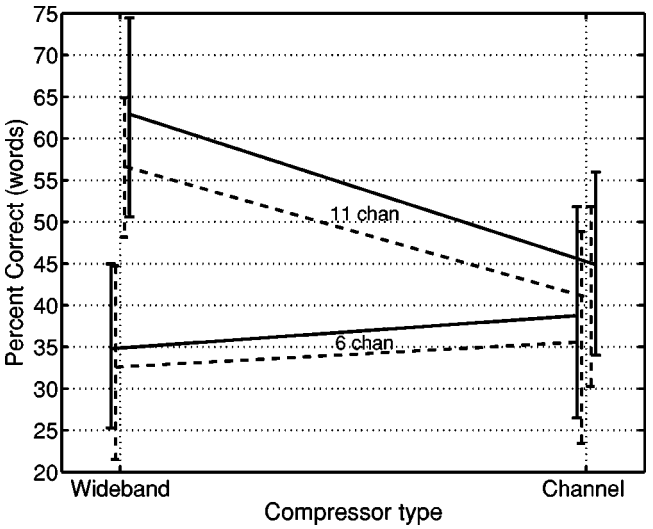


FIG. 4. Results of experiment 3. The solid and dashed lines show data corrected and uncorrected for learning effects, respectively. Error bars show ±1 SD. The target-to-background ratio was 5 dB for the 6-channel system and 2 dB for the 11-channel system.

The corrected scores, transformed from RAUs back into percentages, are plotted for each compressor type in Fig. 4, as solid lines. Corresponding uncorrected scores are shown as dashed lines. The error bars show SDs across subjects. For the 11-channel system, channel compression led to markedly lower intelligibility than wideband compression. For the 6-channel system, intelligibility was similar for channel and wideband compression.

The corrected scores were subjected to an ANOVA with factors number of channels (6 or 11) and type of compression (wideband or channel). There was a significant effect of number of channels, $F(1,7)=96.0$, $p<0.001$, and a significant effect of compressor type, $F(1,7)=10.18$, $p=0.015$. There was also a significant interaction between number of channels and compressor type, $F(1,7)=38.72$, $p<0.001$. This interaction arose because, for the 6-channel system, there was no significant difference in intelligibility between compressor types ($t=1.44$, 7 df, $p>0.1$, two-tailed), while, for the 11-channel system, performance for wideband compression was significantly better than for channel compression ($t=6.40$, 7 df, $p<0.001$, two-tailed).

The lower intelligibility with channel than with wideband compression for the 11-channel system presumably occurred because of the reduction in across-channel spectral contrasts produced by the channel compression. It seems that this effect was more than enough to offset the potentially beneficial effect of channel compression in restricting the bandwidth of comodulation effects. However, for the 6-channel system, intelligibility was similar for channel and wideband compression. Perhaps this reflects a more even balance between the two opposing factors. With the 6-channel system, spectral information was represented in a

highly quantized manner, and subjects may have relied more on within-channel patterns of amplitude modulation than on across-channel comparisons.

The score for the 6-channel wideband compressor is lower (35%) than the 55% found previously under similar test conditions (Stone and Moore, 2003a). This may be related to the use of a different background talker in the two studies. The background talker in the earlier study spoke with a high degree of amplitude modulation, possibly allowing more ''listening in the dips.''

V. DISCUSSION

A. Recommended attack times

The data presented here suggest that several factors can interact in a complex way to influence the efficacy of compression systems, when speech is processed so as to present primarily envelope cues in different frequency bands. We have established previously that fast-acting compression leads to lower intelligibility than slow-acting compression (Stone and Moore, 2003a). We have shown here that changes in the attack time of the fast compressor, while holding the fractional reduction in modulation constant, can also affect intelligibility. Attack times less than about 2 ms, typically associated with the need to avoid overload in electroacoustic systems with limited dynamic range, are more deleterious than longer attack times. The deleterious effect of very short attack times can probably be attributed to the two factors discussed earlier, namely the spectral spreading and the onset comodulation produced by rapid changes in gain.

The results of experiment 1 did not reveal a ''threshold'' attack time below which intelligibility rapidly deteriorated. Rather, there was a small progressive deterioration as the attack time was decreased below 8 ms, with the biggest change occurring between 8 and 2 ms. We tentatively conclude that attack times should exceed ''several'' ms in order to avoid the worst effects of fast compression. Since the best performance was achieved for the longest attack time of 8 ms, the ''optimal'' attack time may well be longer than 8 ms.

B. Importance of preserving envelope shape

In experiment 1, we also assessed an envelope compressor which had an attack time of 10 ms, longer than that of the fast system with the longest attack time. However, the envelope compressor had a markedly shorter release time than any of the fast compressors. The near-symmetric attack and release times of this compressor, plus the fact that gain changes were applied to the delayed audio signal, meant that the compressor produced much less distortion of the envelope *shape* than the fast compressors. This might be expected to improve intelligibility. However, the envelope compressor was much more effective than the fast compressors at reducing modulation depth over the range of modulation rates important for speech perception. This might be expected to impair intelligibility. The results showed that the envelope compressor produced significantly lower intelligibility than any of the fast compressors, which suggests that, for this particular compressor, the deleterious effect of reduced envelope modulation outweighed any potential advantage from better preservation of envelope shape. However, it would be desirable to conduct further experiments manipulating the asymmetry of attack and release times while holding the fractional reduction in modulation as constant as possible. This would help in deciding whether preservation of envelope shape is important.

It should be noted that modulation reduction, as measured by the fractional reduction in modulation, depends on both the attack and release time (as well as compression ratio and threshold). It seems to us likely that modulation reduction is a more relevant and important factor than attack or release times alone.

C. Relative importance of modulation reduction and comodulation

Experiment 2 compared the effect of compressing the target and background before mixing (condition INDEP) and after mixing (condition COMOD). The two conditions introduced similar amounts of modulation reduction, but only condition COMOD introduced comodulation of the target and background. The results clearly showed poorer performance for condition COMOD, indicating that intelligibility can be reduced by comodulation of the target and background. The difference in intelligibility between conditions INDEP and COMOD was smaller for the 11-channel system than for the 6-channel system. This smaller change, despite the more adverse target-to-background ratio used, which would have led to greater comodulation, may have happened because increasing the number of channels in condition COMOD led to a decrease in the overall degree of comodulation of the target and background.

Experiment 3 compared the effect of wideband compression and channel compression. For the 11-channel system, performance was significantly worse for the latter, suggesting that reduction of across-channel contrasts, which convey spectral information, can produce a reduction in intelligibility, as argued by Plomp (1988). However, the effect was not apparent when 6 channels were used.

It seems clear that both modulation reduction and comodulation can lead to reduced intelligibility in a competing speech task. Comodulation appears to play a smaller role as the number of channels increases, while modulation reduction plays a greater role.

D. Hybrid fast and slow compression

A possible way of avoiding the deleterious effects of fast-acting compression described above is to use a two-stage compression process, similar to the DUAL-4 system described by Stone *et al.* (1999) for acoustic hearing aids. In that system, a slow-acting compressor, with a low-to-moderate compression ratio, was followed by fast-acting channel compression, again with low-to-moderate compression ratios. The slow-acting compressor reduced gross variations in signal level between environments so as to present a narrower range of levels to the fast-acting compressors. These compressors then mapped the remaining dynamic range of the signal into the reduced dynamic range of the impaired cochlea. For an implant, where the dynamic range

that can be applied to a single electrode is very small, more compression is necessary. This ''extra'' compression should probably be incorporated in the slow-acting compressor, especially in implants with a large number of channels (greater than 6); otherwise, the deleterious effects associated with reduced within- and across-channel intensity contrasts, as revealed in our experiments, will play a significant role. There is at least one commercially available cochlear implant system that incorporates a slow-acting wideband compressor followed by fast-acting channel compressors.

E. Applicability to acoustic hearing aids for profound hearing loss

We have focused here on the effects of fast-acting compression on performance with a cochlear-implant simulator. However, the results may also have implications for the design of acoustic hearing aids for people with sensorineural hearing loss, especially severe to profound loss. Such people have reduced frequency selectivity (Pick *et al.*, 1977; Glasberg and Moore, 1986; Moore, 1998), and some are relatively insensitive to temporal fine structure cues (Rosen and Fourcin, 1986; Moore, 1998; Moore and Moore, 2003), but have a good ability to use temporal envelope cues (Bacon and Gleitman, 1992; Moore *et al.*, 1992; Turner *et al.*, 1995). Any form of signal processing that adversely affects the use of temporal envelope cues might therefore be expected to have a deleterious effect on performance.

F. Limitations of the simulation

The noise-vocoder stimulation of a cochlear implant used here has also been employed by many others. However, it should be noted that it does not adequately simulate all aspects of a cochlear implant system. With a real cochlear implant, the available dynamic range on each electrode is very small, while the simulation is used with normally hearing listeners who have a wide dynamic range. In an implant, an additional stage of fast-acting or instantaneous compression is nearly always applied to the signal for each electrode. Such a stage was not used in our simulation (or by others using similar simulations), because this would result in only a small portion of the dynamic range of the (normally hearing) listeners being used. Also, the use of instantaneous compression would lead to severe spectral spreading, which would mean that the ''channel'' signals were no longer confined to specific spectral regions.

A second limitation of the simulation is that, because the cutoff frequency of the envelope filters used in the simulation was relatively low (50 Hz), information about the fundamental frequency of the target and background talkers was largely removed. In a real cochlear implant, at least some information about fundamental frequency may be conveyed. Thus, it would be desirable to evaluate different forms of amplitude compression with real cochlear implant users, and not just with a simulation.

VI. CONCLUSIONS

Using a simulation of a cochlear implant assessed with normal-hearing listeners, we have explored the factors underlying the reduction in speech intelligibility produced by fast-acting compression, as demonstrated by Stone and Moore (2003a). The following are our main findings and conclusions.

(1) When the fractional reduction in modulation depth is held constant over a wide range of modulation rates, intelligibility decreases as the attack time of a fast compression system is decreased from 8 to 0.125 ms, keeping release times in the range 228–244 ms. This can probably be accounted for by the rapid changes in gain associated with fast attack times, which lead to two main effects: (a) correlated abrupt changes in amplitude of the target and background, which may inhibit their perceptual segregation; and (b) spectral artifacts associated with the modulation of components in the original signal.
(2) A fast-acting envelope compressor with an attack time of 10 ms and a release time of 13 ms led to lower intelligibility than any of the compressors described in (1) above, even though the envelope compressor produced less distortion of envelope shape and avoided most of effects (a) and (b) above. The poorer intelligibility produced by the envelope compressor can be attributed to the fact that it was much more effective than the compressors described in (1) in reducing the modulation depth of the input signal.
(3) Applying compression independently to the target and background, prior to mixing, led to higher intelligibility than when compression was applied after mixing. This indicates that comodulation of the target and background contributes to the reduced intelligibility produced by fast-acting compression. The effect was greater for a 6-channel simulation than for an 11-channel simulation, suggesting that the importance of comodulation decreases with increasing number of channels.
(4) For an 11-channel simulation, compression applied independently to the channel signals resulted in lower intelligibility than compression applied to the wideband signal, even though comodulation effects would have been smaller for the independent channel compression. For this number of channels, reduction of across-channel contrasts, which convey spectral information, was probably the dominant factor leading to poor performance for the channel compression. For the 6-channel simulation, performance was similar for wideband and channel compression, suggesting that the deleterious effects of reduced across-channel contrasts for the latter were offset by the reduction in comodulation effects.

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (UK). We thank Tom Baer for comments on an earlier version of the manuscript. We also thank Andrew Faulkner and two anonymous reviewers for helpful comments.

Agnew, J., and Thornton, J. M. (**2000**). ''Just noticeable and objectionable group delays in digital hearing aids,'' J. Am. Acad. Audiol. **11**, 330–336.

ANSI (**1996**). ANSI S3.22-1996. ''Specification of hearing aid characteristics'' (American National Standards Institute, New York).
ANSI (**1997**). ANSI S3.5-1997. ''Methods for the calculation of the speech intelligibility index'' (American National Standards Institute, New York).
Bacon, S. P., and Gleitman, R. M. (**1992**). ''Modulation detection in subjects with relatively flat hearing losses,'' J. Speech Hear. Res. **35**, 642–653.
Baer, T., and Moore, B. C. J. (**1994**). ''Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech,'' J. Acoust. Soc. Am. **95**, 2277–2280.
Baer, T., Moore, B. C. J., and Gatehouse, S. (**1993**). ''Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times,'' J. Rehabil. Res. Dev. **30**, 49–72.
Bench, J., and Bamford, J. (**1979**). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London).
Braida, L. D., Durlach, N. I., De Gennaro, S. V., Peterson, P. M., and Bustamante, D. K. (**1982**). ''Review of recent research on multiband amplitude compression for the hearing impaired,'' in *The Vanderbilt Hearing-Aid Report*, edited by G. A. Studebaker and F. H. Bess (Monographs in Contemporary Audiology, Upper Darby, PA).
Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bradford, MIT Press, Cambridge, MA.).
Bregman, A. S., Abramson, J., Doehring, P., and Darwin, C. J. (**1985**). ''Spectral integration based on common amplitude modulation,'' Percept. Psychophys. **37**, 483–493.
Bustamante, D. K., and Braida, L. D. (**1987**). ''Multiband compression limiting for hearing-impaired listeners,'' J. Rehabil. Res. Dev. **24**, 149–160.
Carrell, T. D., and Opie, J. M. (**1992**). ''The effect of amplitude comodulation on auditory object formation in sentence perception,'' Percept. Psychophys. **52**, 437–445.
Clark, G. M., Tong, Y. C., and Patrick, J. F. (**1990**). *Cochlear Prostheses* (Churchill Livingston, Edinburgh).
Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory grouping,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).
Dillon, H. (**1996**). ''Compression? Yes, but for low or high frequencies, for low or high intensities, and with what response times?'' Ear Hear. **17**, 287–307.
Drullman, R., Festen, J. M., and Houtgast, T. (**1996**). ''Effect of temporal modulation reduction on spectral contrasts in speech,'' J. Acoust. Soc. Am. **99**, 2358–2364.
Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). ''Effect of reducing slow temporal modulations on speech reception,'' J. Acoust. Soc. Am. **95**, 2670–2680.
Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). ''Effect of temporal envelope smearing on speech reception,'' J. Acoust. Soc. Am. **95**, 1053–1064.
Eddington, D. K. (**1983**). ''Speech recognition in deaf subjects with multichannel intracochlear electrodes,'' Ann. N.Y. Acad. Sci. **405**, 241–258.
Elberling, C., and Hansen, K. V. (**1999**). ''Hearing instruments—interaction with user preference,'' in *Auditory Models and Non-linear Hearing Instruments*, edited by A. N. Rasmussen, P. A. Osterhammel, T. Andersen, and T. Poulsen (Holmens Trykkeri, Copenhagen, Denmark).
Festen, J. M., and Plomp, R. (**1990**). ''Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.
Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). ''Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,'' J. Acoust. Soc. Am. **110**, 1150–1163.
Glasberg, B. R., and Moore, B. C. J. (**1986**). ''Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,'' J. Acoust. Soc. Am. **79**, 1020–1033.
Hafter, E. R., Bonnel, A. M., and Gallun, E. J. (**1996**). ''Detection of change without regard to its valence,'' J. Acoust. Soc. Am. **99**, 2541.
Hall, J. W., and Grose, J. H. (**1990**). ''Comodulation masking release and auditory grouping,'' J. Acoust. Soc. Am. **88**, 119–125.
Hall, J. W., and Grose, J. H. (**1991**). ''Some effects of auditory grouping factors on modulation detection interference (MDI),'' J. Acoust. Soc. Am. **90**, 3028–3035.
Hall, J. W., Grose, J. H., and Mendoza, L. (**1995**). ''Across-channel processes in masking,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).
IEEE (**1968**). *Recommended Practice for Speech Quality Measurements* (Institute of Electrical and Electronics Engineers, New York).
MacLeod, A., and Summerfield, Q. (**1990**). ''A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use,'' Br. J. Audiol. **24**, 29–43.
Macmillan, N. A. (**1973**). ''Detection and recognition of intensity changes in tone and noise: The detection-recognition disparity,'' Percept. Psychophys. **13**, 65–75.
Moore, B. C. J. (**1990**). ''How much do we gain by gain control in hearing aids?'' Acta Oto-Laryngol., Suppl. **469**, 250–256.
Moore, B. C. J. (**1992**). ''Across-channel processes in auditory masking,'' J. Acoust. Soc. Jpn. (E) **13**, 25–37.
Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).
Moore, B. C. J. (**2003a**). ''Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants,'' Otol. Neurotol. **24**, 243–254.
Moore, B. C. J. (**2003b**). ''Speech processing for the hearing-impaired: Successes, failures, and implications for speech mechanisms,'' Speech Commun. **41**, 81–91.
Moore, B. C. J., and Moore, G. A. (**2003**). ''Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects,'' Hear. Res. **182**, 153–163.
Moore, B. C. J., Peters, R. W., and Stone, M. A. (**1999**). ''Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips,'' J. Acoust. Soc. Am. **105**, 400–411.
Moore, B. C. J., and Shailer, M. J. (**1992**). ''Modulation discrimination interference and auditory grouping,'' Philos. Trans. R. Soc. London, Ser. B **336**, 339–346.
Moore, B. C. J., Shailer, M. J., and Black, M. J. (**1993**). ''Dichotic interference effects in gap detection,'' J. Acoust. Soc. Am. **93**, 2130–2133.
Moore, B. C. J., Shailer, M. J., and Schooneveldt, G. P. (**1992**). ''Temporal modulation transfer functions for bandlimited noise in subjects with cochlear hearing loss,'' Br. J. Audiol. **26**, 229–237.
Pick, G., Evans, E. F., and Wilson, J. P. (**1977**). ''Frequency resolution in patients with hearing loss of cochlear origin,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London).
Plomp, R. (**1983**). ''The role of modulation in hearing,'' in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer, Berlin).
Plomp, R. (**1988**). ''The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function,'' J. Acoust. Soc. Am. **83**, 2322–2327.
Plomp, R. (**1994**). ''Noise, amplification, and compression: Considerations of three main issues in hearing aid design,'' Ear Hear. **15**, 2–12.
Qin, M. K., and Oxenham, A. J. (**2003**). ''Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers,'' J. Acoust. Soc. Am. **114**, 446–454.
Rasch, R. A. (**1978**). ''The perception of simultaneous notes such as in polyphonic music,'' Acustica **40**, 21–33.
Robinson, C. E., and Huntington, D. A. (**1973**). ''The intelligibility of speech processed by delayed long-term averaged compression amplification,'' J. Acoust. Soc. Am. **54**, 314.
Rosen, S., and Fourcin, A. (**1986**). ''Frequency selectivity and the perception of speech,'' in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Stone, M. A., and Moore, B. C. J. (**1992**). ''Syllabic compression: Effective compression ratios for signals modulated at different rates,'' Br. J. Audiol. **26**, 351–361.
Stone, M. A., and Moore, B. C. J. (**2002**). ''Tolerable hearing-aid delays. II. Estimation of limits imposed during speech production,'' Ear Hear. **23**, 325–338.
Stone, M. A., and Moore, B. C. J. (**2003a**). ''Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task,'' J. Acoust. Soc. Am. **114**, 1023–1034.
Stone, M. A., and Moore, B. C. J. (**2003b**). ''Tolerable hearing-aid delays. III. Effects on speech production and perception of across-frequency variation in delay,'' Ear Hear. **24**, 175–183.
Stone, M. A., Moore, B. C. J., Alcántara, J. I., and Glasberg, B. R. (**1999**).

''Comparison of different forms of compression using wearable digital hearing aids,'' J. Acoust. Soc. Am. **106**, 3603–3619.
Studebaker, G. (**1985**). ''A rationalized arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.
Summerfield, Q., and Culling, J. F. (**1992**). ''Auditory segregation of competing voices: Absence of effects of FM or AM coherence,'' Philos. Trans. R. Soc. London, Ser. B **336**, 357–366.
Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). ''Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **97**, 2568–2576.
Verschuure, J., Maas, A. J. J., Stikvoort, E., de Jong, R. M., Goedegebure, A., and Dreschler, W. A. (**1996**). ''Compression and its effect on the speech signal,'' Ear Hear. **17**, 162–175.
Villchur, E. (**1973**). ''Signal processing to improve speech intelligibility in perceptive deafness,'' J. Acoust. Soc. Am. **53**, 1646–1657.
Zeng, F.-G. (**2004**). ''Compression and cochlear implants,'' in *Compression: From Cochlea to Cochlear Implants*, edited by S. P. Bacon, A. N. Popper, and R. R. Fay (Springer, New York).

Secant lines in a preliminary study of phonetic reduction of /ɪ/ and /ɛ/

Richard S. McGowan[a)]
CReSS LLC, 1 Seaborn Place, Lexington, Massachusetts 02420-2002

(Received 8 May 2003; revised 13 May 2004; accepted 13 July 2004)

Two subjects from the X-Ray Microbeam Speech Production Database were examined in their production of the vowels /ɪ/ and /ɛ/ in alveolar and dental consonant contexts. Secant lines, or first-order splines, between the three most anterior pellets were examined at vowel critical times. These critical times were zero crossings in the tangential acceleration of the midpoints of the secant lines. We expected and found, in general, that vowel reduction occurred as a function of vowel duration in measures of the secant line midpoint-to-palate distance and secant line orientation at vowel critical times. The shorter the vowel, the smaller the distance of the secant line midpoints to the palate and the less downward the orientation of the secant lines at the vowel critical times. Phonetic reduction was also apparent in the formant frequencies. There were differences between the speakers in terms of the range of vowel duration and degree of reduction. The subjects differed in the functional parts of the tongue spanned by the secant lines and the shape of their palates. These differences were factors in the observed relations between formant frequencies and the articulatory, secant line measures for each subject.
[DOI: 10.1121/1.1789491]

PACS numbers: 43.70.Aj [AL] Pages: 2324–2337

I. INTRODUCTION

A. Previous work

Tongue kinematics in consonant–vowel and vowel–consonant transitions needs to be understood in order to perform articulatory synthesis of English and other languages. One particular phenomenon associated with these transitions is phonetic vowel reduction. Phonetic vowel reduction in syllables is a complex, speaker-dependent phenomenon that is often characterized by greater coarticulation of the vowel with the surrounding consonants as the duration of the vowel decreases. In this paper, we seek to understand some articulatory and acoustic aspects of phonetic vowel reduction in English front vowel production in terms of tongue positions and shapes and their consequent acoustics.

Lindblom (1963) quantified reduction of Swedish lax vowels in the acoustic domain by considering formant frequencies of the vowels as a function of vowel duration for one speaker. He showed that the shorter duration vowels tended to possess formant frequencies at their centers that were more greatly affected by the surrounding segmental context than did those of longer duration. Thus, he concluded that vowel reduction in languages with ''heavy stress,'' like Swedish and English, is a phenomenon of assimilation and not of neutralization. Lindblom (1964) verified what he observed in his 1963 study on another Swedish speaker with both tense and lax vowels. Stevens and House (1963) also showed that the first two formants in American English vowels spoken in words with symmetric consonant context were affected by the places of articulation of those consonants. Again, the vowels showed assimilation, or coarticulation, with the surrounding consonants.

There have been a number of acoustic studies of coarticulation in vowel production since the original Lindblom and Stevens and House papers appeared. Stevens, House, and Paul (1966) noted asymmetries in second formant frequency, or F2, trajectories for vowels in symmetric consonant context of nonsense syllables. Further, for lax vowels, there was a tendency for F2 to approach a neutral value at closure of the final consonant. Ohde and Scharf (1975) observed that the onset consonant had more influence on the succeeding vowel than did the vowel's succeeding consonant. The Stevens *et al.* (1966) and Ohde and Scharf (1975) observations could indicate both a coarticulatory effect of surrounding consonants on the vowel, and, perhaps, a certain degree of lenition of the final consonant, particularly when it is preceded by a lax vowel. Broad and Fertig (1970) were able to show that a linear superposition of the consonant effects on the vowel /ɪ/ in nonsymmetrical CVC syllables provided a good statistical model of formant trajectories. Broad and Clermont (1987) refined this model to include more vowels and to obtain particular formant trajectory shapes.

Acoustic investigations have progressed substantially in considerations of factors other than vowel duration in vowel reduction. Gay (1978) found no evidence of vowel reduction with changes in speaking rate and stress. However, this could have been the result of using vowels in voiceless, bilabial consonant context, and the result of instructions given to the speakers. Nord (1986) employed acoustic analyses to show that factors other than duration, such as phrase position and stress, have an effect on the degree of vowel reduction. He studied vowels in syllables with alveolar/dental consonant onsets and codas, and these syllables were parts of two syl-

a)Electronic mail: rsmcgowan@earthlink.net

lable words. Di Benedetto (1989) studied the properties of F1 trajectories in three English speakers and found further evidence of assimilation with the surrounding consonants. She also showed that F1 and F2 sampled at maximum F1 alone could not be used to distinguish between /ɪ/ and /ɛ/. However, by considering other F1 trajectory properties, such as the proportion of total vowel duration when the maximum is attained, these vowels could be distinguished. Van Son and Pols (1990, 1992), in studying the read speech of a professional newscaster, found minimal effects on formant tracks when vowel durations were changed as a result of speech rate changes. The only consistent change that they found was that of a higher first formant frequency for all vowels in the fast speaking rate condition. This change in F1 could have been the result of a consistently more open mouth, or other postural change in the vocal tract, during faster rate speech. They concluded that the speaker changed speaking style with speaking rate in order to maintain midvowel formant frequencies.

Further work has confirmed that speaking style and suprasegmental conditions are major factors in vowel reduction phenomena. van Bergem (1993) performed a study on the effects of word stress, word class, and sentence stress on the vowels in Dutch read sentences. He found that word stress and word class were the major determiners of F1 and F2. Further, he confirmed that vowel reduction was a consonant assimilation phenomenon and not the result of neutralization. van Bergem concluded that vowel reduction is as much a matter of speaking style as it is a matter of vowel duration. Moon and Lindblom (1994) also studied vowel reduction using formant frequency data. They asked speakers to produce speech in clear and citation-form styles at a constant speaking rate. They found the expected assimilation of vowel formants with surrounding segments, but that clear speech involved an ''undershoot-compensating'' mechanism. By adding the extra effect of speaking style, they extended the results of Lindblom (1963).

The intricacies of vowel reduction are also apparent in articulatory studies of tongue movement. Lindblom (1964) confirmed that, for a single speaker producing a single consonant-vowel combination, articulators moved a smaller distance from consonant release into the vowel when the duration of the vowel was diminished. Kuehn and Moll (1976) noted that simple instructions to talkers to increase the rate of speech of nonsense words and syllables sometimes resulted in either increased articulatory velocity, decreased articulatory displacement, or a combination of these effects. Flege (1988), who studied rate-induced vowel reduction in schwa and bilabial consonant environments, noted some vowel reduction, as well as production strategies that mitigated vowel reduction. These strategies included increased peak tongue velocities, as well as early movement toward the stressed vowel and preservation of the vowel in movements away from the vowel. Engstrand (1988) experimented with Swedish tense vowels, also in schwa and bilabial consonant contexts. He noted that the stressed vowels were more extreme in the formant frequencies and they were produced with tighter vocal tract constrictions. He found little effect of rate on vowel reduction, but noted that his results could have been different in other consonantal contexts. More recently Adams, Weismer, and Kent (1993) have confirmed and extended the observations of Kuehn and Moll (1976) that speakers have a variety of articulatory strategies when changing speaking rate. Finally, Perkell, Zandipour, Matthies, and Lane (2002) have shown that speakers do use different articulatory means to change from normal style to clear style speech.

B. The present study

The present study seeks to supplement and extend the previous findings on changes in vowel articulation and acoustics with vowel duration. In particular, front vowels /ɪ/ and /ɛ/ in alveolar and dental consonant environments for two English speakers were examined using *secant lines*, which are described below. The data were from the Wisconsin X-Ray Microbeam Speech Production Database, the XRMB-SPD (Westbury, 1994), which contains digitized tongue flesh point trajectories for adults producing various word lists, sentences, and paragraphs.

There are many factors that determine the duration of the vowel, such as rate, degree of emphasis or stress, phrase position, and the idiolect of individual speakers. We chose to examine the effect of differences in vowel duration within individual speakers, without regard to the linguistic or nonlinguistic causes of these differences. Part of this was driven by a desire to see whether the degree of vowel reduction or consonant assimilation effects could be detected as a function of vowel duration alone, absent reference to the causes of the vowels' absolute durations. The other reason is that the data could not be readily separated according to rate, stress, and phrase position without reducing the number of tokens within each category to the point that statistically significant trends could be detected.

This study of vowel reduction is a part of a larger project to base articulatory synthesis on kinematic data. More fundamentally, the goal is to understand and verify models of the articulatory-acoustic relations. With the most readily available kinematic data in the form of flesh point data as collected by the x-ray microbeam system (Fujimura, Kiritani, and Ishida, 1973; Westbury, 1994) and the ElectroMagnetic Midsagittal Articulometry, EMMA (Perkell, Cohen, Svirsky, Matthies, Garabieta, and Jackson, 1992), it is important to be able to analyze these data employing articulatory representations that support these goals. Primarily, the articulatory representations of flesh point data needs to provide tongue shape information, because the overall shape of the tongue surface determines the vocal-tract area function, and, thus, the resonance properties of the vocal tract. The line segments between neighboring flesh points are called secant lines here. Secant lines provide a first step in characterizing the geometric relations between the tongue surface and the fixed structures, such as the hard palate, for flesh point data. Although secant line kinematics can provide only a first approximation to tongue shape kinematics, it does so without the much more complicated continuum description. Further, it is a local description in that changes in one secant line do not change a secant line that does not share a pellet. This is unlike PCA or RARAFAC analyses (e.g., Hoole, 1999),

which characterize statistical regularities in the global shape of the tongue by a small number of basic global shapes. The immediate motivation here was to find how vowel reduction phenomena could be characterized within a secant line representation. In this work we were concerned with following the movement of two of the most anterior secant lines between the three most anterior pellets on the tongues of two subjects.

The procedure of tracking secant line kinematics has been foreshadowed by Stone (1990), who studied CV and VC transitions in read speech with the x-ray microbeam and ultrasound, where the consonants were /l/ and /s/ and the vowels were /i/, /ɑ/, and /o/. She concluded: "*The tongue can be segmented sagittally into functional segments that operate as semi-independent units within the tongue. Movements of these segments result in local displacement and anterior to posterior tongue rotation*" (Stone, 1990; p. 2216). The main evidence for this statement came from two kinds of measurement. One was that the change in the distance between adjacent pellets among various phonemes depended on which part of the tongue the pellets were attached and on the phoneme type. For instance, the distance between the third and fourth of five pellets always seemed to be smaller than for the others in Stone's test utterances, despite the fact they were set on a protruded tongue at equal distances. The other evidence came from the relative timing of articulatory events. The anterior-most or the posterior-most pellets reached their extreme positions for neighboring phonemes in the same temporal order. The differing inter-pellet lengths and relative timings indicated that the sections of the tongue were controlled somewhat independently. Stone hypothesized five independent sagittal segments in all from the tongue tip to the tongue root. Further, Westbury, Hashi, and Lindstrom (1998) have used secant lines to characterize the tongue shapes of various tokens of [ɹ] in the XRMB-SPD.

C. Secant line kinematics

Typical tongue flesh point data contain time-sampled positions of two to five pellets placed on the midline of the tongue from just behind the tongue tip to, perhaps, 6 cm or so behind the tip. The XRMB-SPD usually has records of four pellets on the midline of the tongue surface. The motion of four midsagittal flesh points on the tongue can be characterized by the motion of each of two secant lines: one between the two most anterior flesh points and another between the two most posterior flesh points. In this paper we were most concerned with the kinematics of the anterior of the tongue during front vowel production. Thus, we considered the kinematics of the two most anterior secant lines, which involved the three most anterior pellets. Numbering the pellets from anterior to posterior, we considered the secant lines between the first and second pellets and the secant line between the second and the third pellets. The former secant line is referred to as the 1–2 *secant line*, and the latter secant line is referred to as the 2–3 *secant line* (see Fig. 1). Each secant line possesses four degrees of freedom in the midsagittal plane: two translation degrees of freedom of the midpoint of the secant line, the rotation about the midpoint, and the dilation, or change in length, of the secant line. We studied the changes in translation and rotational degrees-of-freedom as functions of vowel duration in this work.

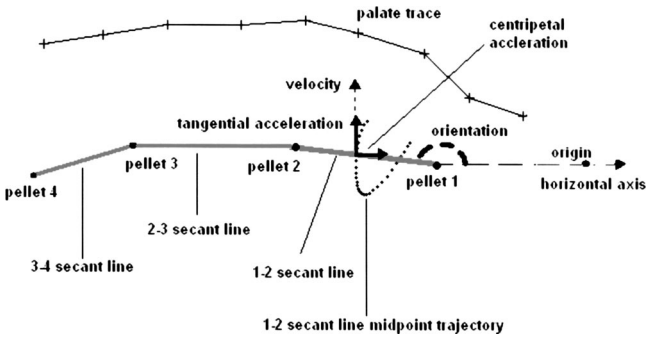


FIG. 1. The 1–2, 2–3, and 3–4 tongue secant lines for JW29. The dashed line with the arrow at the end of pellet 1 is the *x* axis (horizontal direction). The palate trace is also shown. The details of the 1–2 secant line kinematics are shown. The locus of dots, or points, represents the time-sampled path of the midpoint of the 1–2 secant line through a vowel: one for each time sample at 160 Hz. At any particular time, the velocity and tangential acceleration vectors are parallel, and the centripetal acceleration is orthogonal to these. The orientation of the 1–2 secant line is the angle between the horizontal direction and the 1–2 secant line.

The kinematic quantities defined below are illustrated in Fig. 1 for the 1–2 secant line. The midpoint of each secant line travels along a path, or locus of points, in the midsagittal plane. The time derivative of the midpoint position vector is the velocity of the midpoint. This is a two-dimensional vector with a direction that specifies the *tangent direction* and a magnitude, or *speed*, at each point on the path. The time derivative of the velocity can also be calculated to obtain acceleration. Because the velocity has two spatial degrees of freedom, the acceleration has two spatial degrees of freedom. One of these specifies the time rate of change of the speed in the tangent direction, and the other specifies the time rate of change of the tangent direction itself. The first component is parallel to the tangent direction and is known as *tangential acceleration*. While translation speeds are positive by definition, tangential acceleration can either be negative or positive, because the translation speed can either be increasing or decreasing. The second component of acceleration is directed orthogonal to the tangent and is known as *centripetal acceleration*. Another part of the specification for the secant line is in terms of *orientation angle* with respect to the horizontal direction. The orientation angle increases when the secant line rotates about its midpoint in the counterclockwise direction and the orientation angle decreases when the rotation is clockwise. In the former case the anterior end of the secant line points more upward, and in the latter case it points more downward. The origin of the coordinate system is that defined in the XRMB-SPD handbook (Westbury, 1994, p. 37), and is near the tips of the upper incisors in the midsagittal plane. The *x*-axis is formed by the intersection of the midsagittal plane and a plane that intersected the tips of the central incisors and at least two other maxillary teeth on opposite sides of the mouth. The *y* axis is the line in the midsagittal plane, passing through the origin that is orthogonal to the *x* axis.

TABLE I. Speech tokens for subjects JW11 and JW29.

Word	Context	Number of tokens for JW11	Range of vowel durations for JW11	Number of tokens for JW29	Range of vowel durations for JW29
this	Word list	5	163–225 ms	5	225–256 ms
	"...damage **this** ship's hull."	1	100 ms	1	106 ms
	"Put **this** one right here."	1	94 ms	1	88 ms
	"But **this** March,..."	0		1	100 ms
	"**This** March, Tom..."	0		1	88 ms
sense	Word list	5	106–150 ms	5	163–181 ms
	"...provide a **sense** of order."	4	106–131 ms	5	100–125 ms
	"...make **sense** of the problem."	5	81–113 ms	5	100–118 ms
said	Word list	1	194 ms	1	250 ms
	"They all know what I **said**."	5	163–194 ms	5	206–238 ms

II. METHOD

Two subjects, JW11 (male) and JW29 (female), were chosen from the XRMB-SPD for examination. These subjects were chosen based on gender and because they had different palate shapes, and part of the inquiry was to address differences caused by palate shape. There turned out to be other differences that appeared more obvious as the data were analyzed: the two speakers had different articulatory behaviors as a function of vowel duration and corresponding pellets did not mark the same functional parts of the tongue. The XRMB-SPD contains articulatory data from several channels and acoustic data that have been synchronized using mathematical interpolation. All articulatory data channels have the common frame rate of 160 samples/s, which means that the duration between frames is 6.25 ms (Westbury, 1994, p. 57). The output speech was sampled at 21379 samples/s with a microphone placed 10 cm anterior and 5 cm lateral to the mouth opening (Westbury, 1994, p. 29).

We noted those words that contained either the front vowel /ɪ/ or /ɛ/ for subjects JW11 and JW29, and that were immediately preceded and followed by alveolar or dental consonants or consonant clusters. The words that met these criteria, and for which there were more than one or two tokens, were "this," "sense," and "said." Thus, the initial consonants considered were /s/ and /ð/, and the final consonants or consonant clusters were /d/, /ns/, and /s/. These words appeared in various contexts: word lists, lists of sentences, and stories. Table I gives the numbers of tokens, contexts, and range of vowel durations for JW11 and JW29. The times of release of the initial consonant and closure of the final consonant or first consonant of a final consonant cluster were marked based on the acoustic record. The times of release of the fricatives were taken at the cessation of frication. Closures were marked by abrupt changes in acoustic amplitude in frequencies above 1 kHz for /d/ and /ns/ and by the beginning of frication for /s/. These were marked to the nearest articulatory sample time, rounding down to the preceding articulatory sample time for releases and up to the following articulatory sample time for closures. The time interval between each release and the closure was the vowel duration.

We chose to analyze the 1–2 and 2–3 secant lines. Without constraints on the movement of the tongue, movements of the 1–2 or 2–3 secant lines cannot be used to infer what happens in posterior regions of the tongue. For example, the translational movement of either or both the 1–2 or 2–3 secant lines could be caused by a general downward translation of the tongue body or a rotation of a portion of the tongue in the oral cavity with an axis of rotation posterior to the midpoints of the secant lines in question, as could be caused by jaw rotation. In the data examined here, we argue that differences in the acoustics caused by the parts of the tongue posterior to the 1–2 and 2–3 secant lines is minimal within a vowel, because of bracing of the sides of the tongue on the palate and teeth for mid to high front vowels (Perkell, 1991). This physical constraint should be further enhanced because these vowels appear in coronal consonant contexts. Also, not only would jaw motion be limited, but also its effect on the tongue should be limited by the bracing and by the fact that the tongue is not rigidly connected to the jaw.

For subjects in the XRMB-SPD, four pellets were placed along the midline (longitudinal sulcus) of the tongue, so that the most anterior pellet was about 1 cm posterior to the apex (tongue tip) and the most posterior about 6 cm from the apex, as long as a gag reflex was not encountered. The two other pellets were approximately evenly spaced between these extremes, and there were no gender-related differences in pellet spacing and placement (Westbury, 1994). Pellet placement can differ between subjects in the XRMB-SPD. This issue can be addressed by briefly considering the distances between the pellets for our two subjects in sustained vowels /ɑ/ and /ʌ/, because these vowels were produced with a fairly flat tongue blade. For JW11 the average distance between pellets one and two was 16.2 mm, and between pellets two and three the average distance was 14.4 mm. For JW29 the average distance between pellets one and two was 17.3 mm, and between pellets two and three the average distance was 17.7 mm. Thus, the first three pellets had a greater separation for JW29 than for JW11. On the other

hand, the vocal tract appeared to be larger for JW11, an adult male, than for JW29, an adult female. The midsagittal palate outlines of the two subjects indicated a horizontal span of about 60 mm for JW11 and 50 mm for JW29. The line that was meant to be a rough indication of the position of the posterior pharyngeal wall was about 88 mm behind the anterior of the palate trace for JW11 and about 75 mm behind the anterior of the palate trace for JW29. These data indicate a smaller vocal tract, and probably, a smaller tongue for JW29 than for JW11. Further, the palate of JW11 was more domed than that of JW29 [see Figs. 4(a) and 4(b) later for palate shapes], which is yet another factor that makes the effective acoustic path length still longer for JW11 than for JW29. These factors, together with the fact that the pellets were more widely spaced for JW29 than for JW11, indicates that a functionally more posterior part of the tongue was considered for JW29 than for JW11 when examining the positions and orientations of the 1–2 and 2–3 secant lines.

An analysis program was written to calculate kinematic quantities for the 1–2 secant lines and the 2–3 secant lines. Midpoint velocities were calculated numerically using centered differences for each of the horizontal and vertical components of the secant line midpoints. The centered-difference for a digitized signal, $s(t)$, at sample time t_i, where i is a sample index, is $(s(t_{i+1})-s(t_{i-1}))/(t_{i+1}-t_{i-1})$. Speed is the magnitude of the velocity vector. Tangential accelerations were computed numerically using centered differences of the speeds.

In the analysis, we were concerned with characterizing tongue position and shape at consonant release, closure, and in the middle of the vowel. (The choice of time for ''middle of vowel,'' denoted here as *vowel critical time*, is discussed later in Sec. III.) The articulatory measures used were the distance of the midpoint of each secant line to the palate and the orientation of each secant line with respect to the horizontal direction. The former measure is called *secant line midpoint-to-palate distance*, and the latter quantity is called *secant line orientation*. Four or five line segments approximated the midsagittal palate outlines for the purpose of calculating the distances between the secant lines and the palate. The midpoint-to-palate distance was calculated by calculating the lengths of the perpendicular from each palate line to the secant line midpoint. If the resulting perpendicular intersected its palate line outside of the range for which it represented the palate, the length of the line from the closest end point on the palate to the secant line midpoint was taken to be the distance to that portion of the palate. The minimum of the distances between each palate segment and the secant line midpoint was taken to be the midpoint-to-palate distance. We avoided discontinuities in this quantity by using such a calculation. An alternative strategy for calculating distances to the palate has been proposed by Lindstrom and Westbury. They calculate the minimum distance to the palate using an exponentially weighted average distance from each of the sampled palate points to avoid abrupt changes in this quantity with small changes in tongue position (Simpson, 2002).

Based on Lindblom's 1963 analyses, it was expected that the differences between the measures at release or closure and at vowel centers could show a clearer trend with vowel duration than the absolute measures at centers alone, because vowel reduction is a coarticulation, or assimilation, phenomenon. Reference can be made to the 1–2 secant line and its time-sampled midpoint path for /ɛ/ in the word ''sense'' spoken by JW29 illustrated in Fig. 2. Three snapshots of the 1–2 secant line are shown: (1) at release, (2) at the maximum midpoint-to-palate distance, and (3) at closure. We can expect that the positions of the secant lines at release and closure will be mutually dependent on its position at vowel center. [Interestingly, consonant *jaw* position is affected more than vowel jaw position in English (Keating, Lindblom, Lubker, and Kreiman, 1994). We do not know of a comparable study for tongues.] Thus, not only were distances and orientations recorded in the vowel center, but the differences with the corresponding values at release and closure were also calculated.

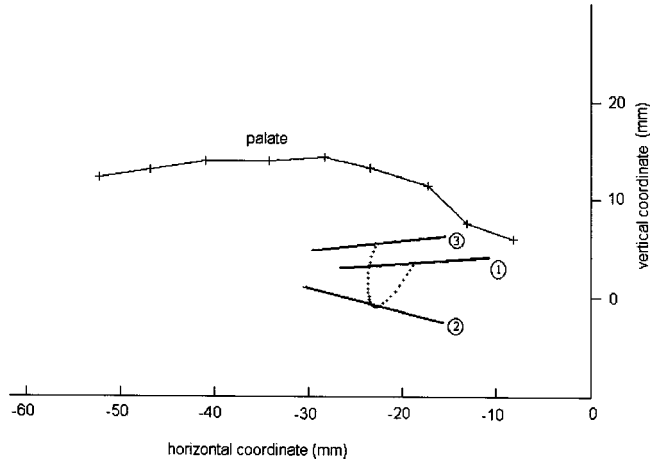


FIG. 2. Examples of a time sample's 1–2 secant line midpoint trajectory as a dotted path for /ɛ/ in the word ''sense'' spoken by JW29. The dark line segments are the 1–2 secant line at (1) the release of /s/, (2) at a time when the movement goes from opening to closing, and (3) at the closure of /n/.

The errors in articulatory measurement are discussed in the manual that accompanies the CDs containing the XRMB-SPD, Westbury (1994). From the manual, we expect the error in measuring the position of a moving pellet to be less than 1 mm. Thus, the error in measuring the position of the midpoint of a secant line will be less than 1 mm. The error in measuring the orientation of a secant line will be approximately, at most, 6°. The microbeam data were low-passed filter using spline fitting that is equivalent to fourth-order, zero delay Butterworth filtering. Corner frequencies for which the attenuation was 6 dB down from full pass were 12 Hz for the most anterior tongue pellet and 10 Hz for the other tongue pellets (Milenkovic, 2003). There were unequal numbers of tokens for each word extracted for each subject most often because of tracking problems during the recording of particular tokens.

These data allowed an exploration of some of the physics of the articulation-acoustic relationship. Formant frequencies were measured using 256-point windows to produce DFT cross sections at vowel centers. Formant frequency measurement errors made by the author were estimated using the measurements taken by another speech scientist who was

uninformed as to the purpose of this study. The average and standard deviation in the absolute value of the percent difference in each of the measured values by subject and formant frequency number were calculated. For JW11's F1s the average was 4% with a standard deviation of 4% and the average for his F2s was 2% with 1% standard deviation. For JW29's F1 the average was 7% with a standard deviation of 3% and the average for her F2s was 3% with 3% standard deviation.

III. RESULTS

A. Qualitative patterns and vowel critical times

Comparisons of vowel durations between JW11 and JW29 indicate that JW11 did not distinguish contexts in which the tokens were spoken as much as JW29 did. For the vowel /ɪ/ the ratio of the maximum to minimum vowel duration was 2.4 for JW11 and 3.7 for JW29. The range of durations for the vowel /ɛ/ produced a ratio of maximum to minimum duration of 2.4 for JW11 and 2.7 for JW29 (see Table I). The smaller ratios for JW11 indicated that JW11 and JW29 could be using different articulatory strategies in producing their vowels. In order to test the hypothesis that JW11 used articulatory compensation to mitigate vowel reduction, regression analyses of maximum 1–2 secant line midpoint speeds in the opening phase for the two subjects and two vowels were performed. A significant negative slope ($p<0.1$) occurred only for the case of JW11 producing /ɪ/. Thus, JW11 sped his articulations up when the /ɪ/ in ''this'' became shorter. Other investigators have noted changes in articulatory behavior that diminish vowel reduction effects (e.g., Kuehn and Moll, 1976; Harris, 1978; Ostry and Munhall, 1985). Speakers will often increase articulatory velocities or the force with which a consonant is released to mitigate the assimilation of the vowel to the surrounding consonants (Moon and Lindblom, 1994).

Because the vowels of interest were front vowels and the surrounding consonants were coronal, there were consistencies in secant line kinematics among the word tokens. This was particularly true for vowels of longer or moderate duration. Consideration of these consistencies allowed for a reasonable choice of criteria for vowel critical times (i.e., vowel centers) for vowels of at least moderate duration. Some of these criteria turned out to be useful for, and could be generalized to, vowels of the shortest duration that did not exhibit these consistencies. The consistencies that are being referred to are illustrated in Fig. 2. The 1–2 secant line midpoint generally had a downward movement away from the palate after the release of the initial consonant, with subsequent upward movement before the closure of the final consonant. Before the end of the downward movement, the midpoint slowed, and at the beginning of the upward movement the speed increased. (The closer the points on the path illustrated in Fig. 2, the slower the midpoint movement.) Thus, the maximum midpoint-to-palate distance often corresponded closely to a speed minimum and a zero crossing in tangential acceleration. Further, the general pattern of rotation for the 1–2 secant line was clockwise or decreasing orientation angle, after consonant release, as illustrated by the snapshots of the secant lines at times (1) and (2) in Fig. 2. Also, the rotation tended to be in the opposite (counterclockwise) direction from sometime in the mid-portion of the vowel to the following consonant closure, as illustrated by the snapshots of the secant lines at times (2) and (3) in Fig. 2. This meant that the tip of the tongue tended to point downward more in the vowel than at the release or closure of the surrounding consonants in many, but not all, cases.

For the pattern just described, the center of the vowel was considered to be sometime during the change in the midpoint direction from downward to upward. Various criteria were considered for choosing times within this time interval to record quantities of interest: the maximum midpoint-to-palate distance, local midpoint speed minima, or zero crossings in the tangential acceleration. In some of the shorter utterances, there were no local speed minima, or they were very close to release or closure, the maximum midpoint-to-palate distance occurred near the release or closure, and/or the only zero crossing in tangential acceleration corresponded to a speed maximum. Because of these different patterns it is a nontrivial problem on how to define the middle of the vowel articulation kinematically. Further, the same kinds of problems occur when defining the middle of the vowel acoustically, because there is not necessarily an apparent local maximum or minimum in the formant trajectories sometime during vowel production. For the 1–2 secant line, the two kinematic quantities that were the least often close to a release or closure times were times of the maximum midpoint-to-palate distance and times of tangential acceleration zero crossing. (In cases where there were multiple zero crossings in the tangential acceleration, the one closest to the acoustic temporal center of the vowel was chosen.) Figures 3(a) and 3(b) show the proportion of total vowel time where these critical times for the 1–2 secant line occur as a function of vowel duration for both /ɪ/ and /ɛ/ for each subject. The times of zero crossings of tangential accelerations were more consistently near the acoustic temporal centers than were the times of the maximum midpoint-to-palate distance for the 1–2 secant lines. Further, the 2–3 midpoint-to-palate distance maxima were not always well defined because this distance often varied little over the course of a vowel and because there were sometimes multiple local maxima. Thus, the zero crossings of the tangential acceleration were used for the kinematically defined midvowel, or vowel critical time. The phrase ''vowel critical time'' will sometimes be shortened to ''vct'' here.

B. Articulation of /ɪ/ at vowel critical times and vowel duration

Figure 4(a) shows the positions of the 1–2 secant lines and the positions of the 2–3 secant lines at their respective vowel critical times from tokens of the word ''this'' for subject JW11. The midpoints of the secant lines are indicated as well. For JW11 there was nearly a continuum of 1–2 secant line midpoints at the 1–2 vowel critical times with continuous variation in their distances to the palate, except for one outlier. Also, the 2–3 secant line midpoints clustered together, except for one outlier. The outliers for both the 1–2 and 2–3 secant lines were from the same token in sentence

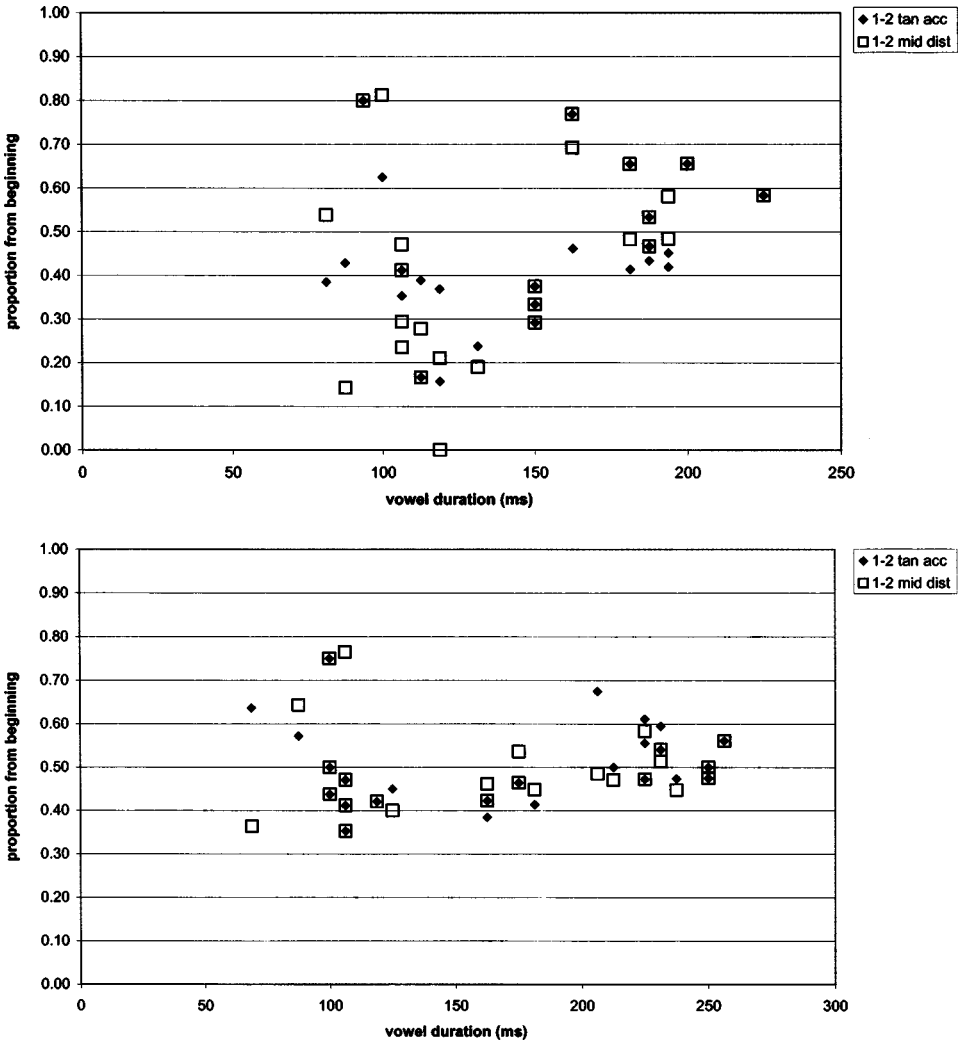


FIG. 3. Proportion of the total vowel duration where the vowel critical time occurred versus the vowel duration for all tokens of /ɪ/ and /ɛ/: (a) JW11; (b) JW29.

context: ''...damage this ship's hull.'' It appears that this /ɪ/ was highly coarticulated with the /dʒ/ of the preceding word. The other /ɪ/ in sentence context had a 1–2 secant line midpoint that was the closest to the palate in the continuum seen in Fig. 4(a).

For JW29 there were clearly two groups of 1–2 secant lines at their vowel critical times [Fig. 4(b)]. These two groups were, perhaps, less clearly defined for the 2–3 secant midpoints. The tokens of the group of four 1–2 secant lines closest to the palate were from the word ''this'' spoken in sentence context, with vowel durations from 88 to 106 ms (Table I). The five 1–2 secant lines farthest from the palate were from the word ''this'' spoken in word lists, with vowel durations from 225 to 256 ms (Table I).

The first column of Table II contains the linear regression statistics for the secant line midpoint-to-palate distance at the vowel critical times versus vowel duration and for secant line orientation at the vowel critical times versus vowel duration for both subjects. The same statistics for the differences of these quantities with their values at release and closure are shown in the second and third columns of this table. Both JW11 and JW29 exhibited, at least, very significant ($p<0.05$) positive slopes in the linear regression of the 1–2 secant line midpoint-to-palate distance at the 1–2 vcts versus the vowel duration (Table II, rows 1 and 2, column 1), i.e., the longer the duration the greater the distance between the palate and the midpoint of the secant lines. A highly significant positive slope occurred for JW29's 2–3 secant line's midpoint-to-palate distance versus vowel duration (Table II, row 4, column 1), but there was no such trend for JW11 (Table II, row 3, column 1). The differences in secant line midpoint-to-palate distances at vct and at either release or closure are shown in rows 1–4, columns 2 and 3 of Table II. Comparing across rows 1–4, and with the exception of JW11's 2–3 secant line in row 3, the regression slopes in column 2 were less significant than those of column 1, and the regression slopes of column 3 were more significant than those in column 1. Thus, generally, differences in the midpoint-to-palate at vct and at closure followed the expected trend with vowel duration with more statistical significance than the midpoint-to-palate distance at either vct or its difference at vct and at release.

As noted in Sec. III A, the negative rotation of the tongue blade into a more downward pointing direction was a commonly observed pattern. It might be expected that the longer vowels would involve more of this negative orientation change than the shorter vowels. For JW11's 1–2 secant line, it was only when the differences in orientation at vowel critical times and at release or closure that significant negative linear trends with vowel duration appeared (Table II,

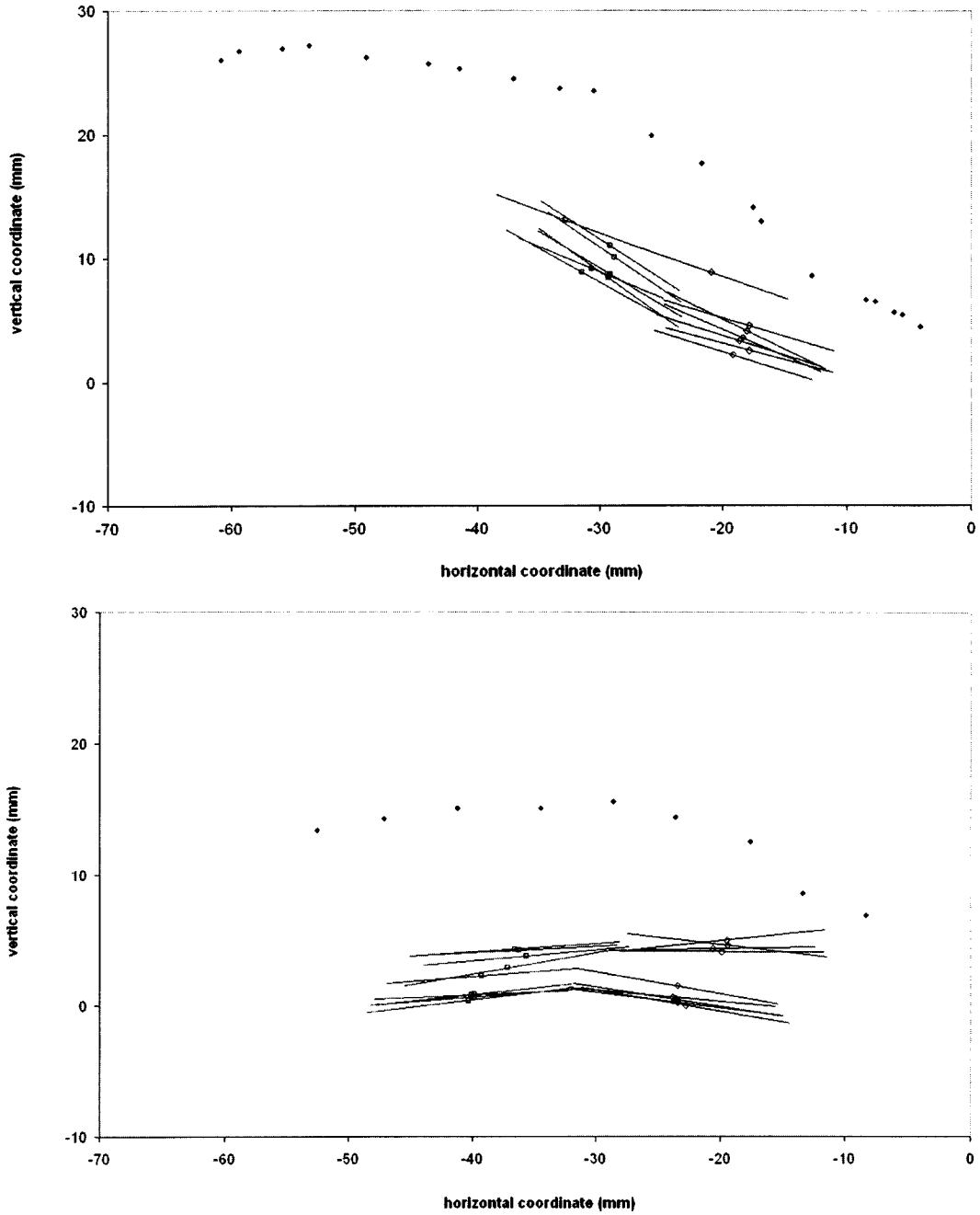


FIG. 4. 1–2 secant lines at their vowel critical times and 2–3 secant lines at their vowel critical times for tokens of "this." The midpoints of the 1–2 secant lines are drawn as diamonds and the midpoints of the 2–3 secant lines are drawn as squares. (a) JW11, (b) JW29.

row 5). It was JW11's 2–3 secant line orientation at the 2–3 vowel critical times or its difference with its value at release or closure that showed very significant negative trends with vowel duration (Table II, row 7). It can be noted that removing the outlier for JW11 did not significantly increase the degree of significance of any of these regression slopes. For JW29 the 1–2 secant line orientations at vcts, as well as differences of 1–2 orientations between vowel critical times and release or closure times, exhibited either very significant or highly significant negative slope with vowel duration (Table II, row 6). JW29's 2–3 secant line orientation versus vowel duration showed significance only when differences between values at vct and closure were considered (Table II, row 8). Also, this trend was a *positive* slope. Other than for JW29's 2–3 secant line, the trend in orientation was for a more downward orientation with increased vowel duration.

C. Formant frequencies and their relation to articulation of /ɪ/

Acoustic vowel reduction of /ɪ/ was not as evident for JW11, as it was for JW29. According to linear regression statistics, JW11 exhibited significance ($p<0.1$) only in a linear increase of F1 at 1–2 vowel critical times with vowel duration (slope=4.17, $R^2=0.46$). Linear regression statistics for JW29 showed highly significant linear increases in F1 at vcts with increased vowel duration (slope=9.73, $R^2=0.89$ at

TABLE II. Statistics for linear regression of the midpoint-to-palate distance and orientation at vowel critical time and the differences of these quantities at vowel critical time and at release and at closure versus the duration of the vowel /ɪ/ in ''this'' for JW11 and JW29 (*=0.1 confidence, **=0.05 confidence, and ***=0.01 confidence). Distances are in mm, orientations in degrees, and durations in ms. ''vct'' stands for the vowel critical time.

Subject	Regression	@ vct	@ vct–@ rel	@ vct–@ clo
JW11	1–2 midpoint-to-palate distance *versus* vowel duration	slope=0.019 $R^2=0.71$**	slope=0.010 $R^2=0.38$	slope=0.021 $R^2=0.80$***
JW29	1–2 midpoint-to-palate distance *versus* vowel duration	slope=0.035 $R^2=0.96$***	slope=0.037 $R^2=0.88$***	slope=0.033 $R^2=0.83$***
JW11	2–3 midpoint-to-palate distance *versus* vowel duration	slope=0.001 $R^2=0.00$	slope=−0.023 $R^2=0.63$**	slope=−0.009 $R^2=0.15$
JW29	2–3 midpoint-to-palate distance *versus* vowel duration	slope=0.016 $R^2=0.77$***	slope=−0.008 $R^2=0.29$	slope=0.020 $R^2=0.73$***
JW11	1–2 orientation *versus* vowel duration	slope=−0.016 $R^2=0.04$	slope=−0.125 $R^2=0.56$*	slope=−0.060 $R^2=0.64$**
JW29	1–2 orientation *versus* vowel duration	slope=−0.053 $R^2=0.64$***	slope=−0.089 $R^2=0.78$***	slope=−0.044 $R^2=0.50$**
JW11	2–3 orientation *versus* vowel duration	slope=−0.103 $R^2=0.82$***	slope=−0.073 $R^2=0.69$**	slope=−0.100 $R^2=0.91$***
JW29	2–3 orientation *versus* vowel duration	slope=−0.004 $R^2=0.02$	slope=−0.001 $R^2=0.00$	slope=0.028 $R^2=0.46$**

1–2 vct and slope=10.36, $R^2=0.86$ at 2–3 vct), while she did not show any significant linear trend for F2 versus vowel duration.

One of the main reasons for using secant lines is that they can provide some insight into the articulatory acoustic relations. Table III provides statistics on linear regressions between F1 or F2 and the midpoint-to-palate distance or secant line orientation at vowel critical times for both subjects. For JW29, F1 significantly increased with both 1–2 secant line midpoint-to-palate distances at 1–2 vct and 2–3 secant line midpoint-to-palate distances at 2–3 vct (Table III, column 3, rows 1 and 2). There were no other significant linear trends, except between F2 and the orientation of the 2–3 secant line for JW11.

D. Articulation of /ɛ/ at vowel critical times and vowel duration

In considering the vowel /ɛ/, the tokens of ''sense'' and ''said'' were combined for each speaker. 1–2 secant line midpoint-to-palate distances at vowel critical times, at release, and at closure are plotted as a function of the vowel duration of JW11's and JW29's /ɛ/ productions in Figs. 5(a) and 5(b). The figures indicate increased 1–2 secant line

TABLE III. Statistics for the linear regression of the secant line midpoint-to-palate distance and orientation versus formant frequencies at the vowel critical times (vct) for JW11 and JW29 producing /ɪ/ in ''this'' (*=0.1 confidence, **=0.05 confidence, and ***=0.01 confidence). Distances are in millimeters, orientations in degrees, and frequencies in Hz.

Versus	JW11 F1	JW11 F2	JW29 F1	JW29 F2
1–2 midpoint-to-palate distance	slope= 0.0011 $R^2=0.27$	slope= 0.0010 $R^2=0.28$	slope= 0.0020 $R^2=0.86$***	slope= −0.0006 $R^2=0.03$
2–3 midpoint-to-palate distance	slope= 0.0015 $R^2=0.34$	slope= −0.0009 $R^2=0.35$	slope= 0.0009 $R^2=0.72$***	slope= −0.0009 $R^2=0.16$
1–2 orientation	slope=−0.04 $R^2=0.09$	slope=−0.03 $R^2=0.17$	slope=−0.03 $R^2=0.61$	slope=−0.01 $R^2=0.05$
2–3 orientation	slope=−0.04 $R^2=0.09$	slope=−0.07 $R^2=0.69$**	slope=−0.02 $R^2=0.02$	slope=−0.01 $R^2=0.07$

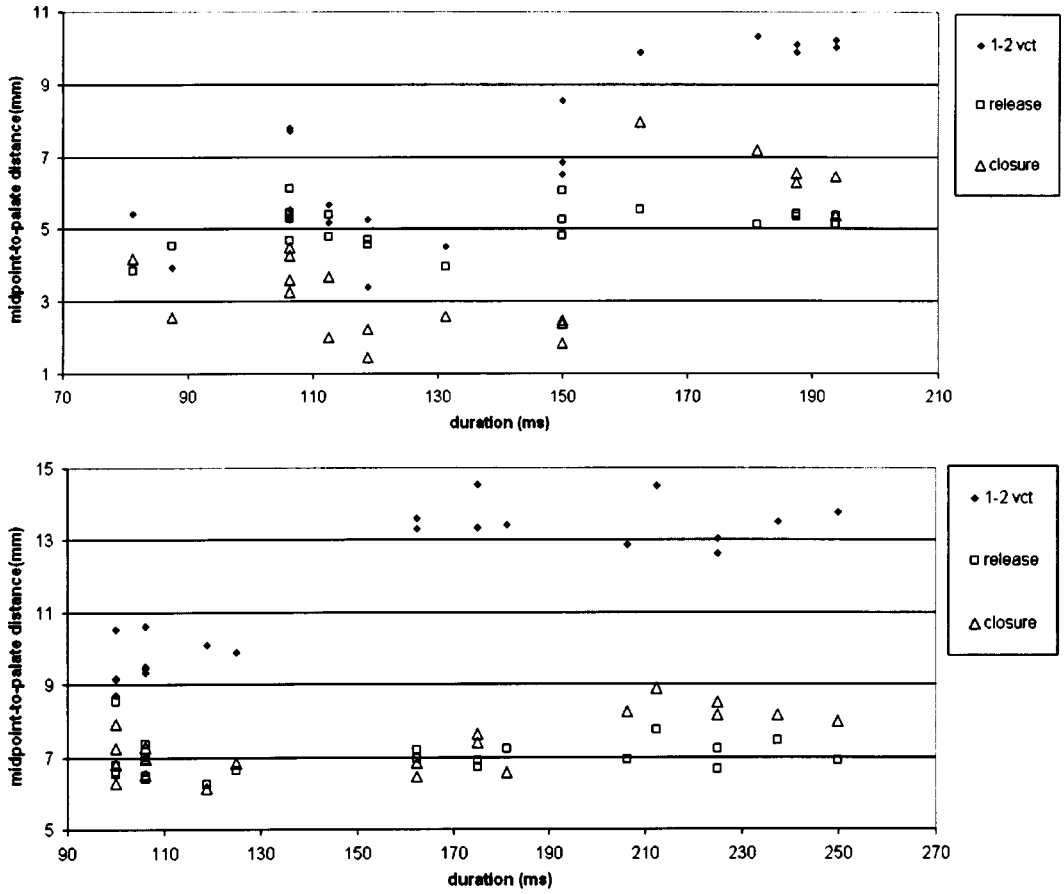


FIG. 5. The 1–2 midpoint-to-palate distance (mm) at release, at closure, and at vowel critical time versus vowel duration (ms) for tokens of /ɛ/ in the words ''sense'' and ''said''. (a) JW11, (b) JW29.

midpoint-to-palate distance at vct with increased vowel duration for both subjects. However, variability in these distances for JW11 at release and at closure makes it difficult to find a trend visually in the differences at vct and at release and closure for this subject [Fig. 5(a)]. For both JW11 and JW29 the slopes of the regression lines for the 1–2 midpoint-to-palate distances at vcts versus vowel duration were all highly significant in the expected positive direction (Table IV, rows 1 and 2, column 1). Similarly, the linear trends between differences in the midpoint-to-palate distance measures taken at vowel critical times and at release or closure and vowel duration were, at least, very significant (Table IV, rows 1 and 2, columns 2 and 3). Also, the midpoint-to-palate distance at vct increased with vowel duration for JW11's 2–3 secant line, as did the differences of this distance between vct and release or closure (Table IV, row 3). JW29's 2–3 secant line behaved differently, with an increase in the midpoint-to-palate distance between vct and release with vowel duration, and *a decrease* when differences of these distances at vct and at closure were considered (Table IV, row 4, columns 1 and 3).

Secant line orientations at vowel critical times, at release, and at closure for the 1–2 secant lines of JW11's and JW29's /ɛ/ productions are shown in Figs. 6(a) and 6(b). A visual inspection of the figure does not reveal any particular trend for JW11, but there does seem to be a trend toward decreasing orientation with vowel duration at for JW29 [Fig. 6(b)]. For JW11 [Fig. 6(a)] there was some statistical indication of decreasing orientation at vct and more negative differences in orientation at vct and at release and closure (Table IV, row 5). These trends were statistically more significant for JW29 than for JW11 (Table IV, rows 5 and 6). As in the case of JW11's /ɪ/ utterances, there were very significant linear trends 2–3 secant line orientation at vct and differences with this quantity at release versus vowel duration (Table IV, row 7, columns 1 and 2). The orientation of JW29's 2–3 secant line orientations at vowel critical times *increased* as a function of vowel duration, as did differences with orientations at release (Table IV, row 8, columns 1 and 2).

A closer look at Figs. 5(b) and 6(b) revealed that for JW29's all vowels with durations greater than about 150 ms behaved similarly in the 1–2 secant line midpoint-to-palate distance and orientation measures. This was a saturation effect that means that a linear fit to the data did not capture all the regularity in the 1–2 secant line distance to palate and orientation versus vowel duration relations. A more appropriate function fit might be provided by an exponential function of vowel duration; similar to the one that Lindblom (1963) used to model behavior of the formant frequencies in midvowel as a function of vowel duration.

E. Formant frequencies and their relation to articulation of /ɛ/

There were individual differences in the way formant frequencies at vowel critical times behaved as a function of

TABLE IV. Statistics for linear regression of the midpoint-to-palate distance and orientation at vowel critical time and the differences of these quantities at vowel critical time and at release and at closure versus the duration of the vowel /ɛ/ in ''sense and said'' for JW11 and JW29 (*=0.1 confidence, **=0.05 confidence, and ***=0.01 confidence). Distances are in mm, orientations in degrees, and durations in ms. ''vct'' stands for vowel critical time.

Subject	Regression	@ vct	@ vct—@ rel	@ vct—@ clo
JW11	1–2 midpoint-to-palate distance *versus* vowel duration	slope=0.053 $R^2=0.69$***	slope=0.047 $R^2=0.72$***	slope=0.020 $R^2=0.32$**
JW29	1–2 midpoint-to-palate distance *versus* vowel duration	slope=0.033 $R^2=0.74$***	slope=0.032 $R^2=0.39$***	slope=0.024 $R^2=0.23$**
JW11	2–3 midpoint-to-palate distance *versus* vowel duration	slope=0.042 $R^2=0.67$***	slope=0.024 $R^2=0.31$**	slope=0.025 $R^2=0.75$***
JW29	2–3 midpoint-to-palate distance *versus* vowel duration	slope=−0.005 $R^2=0.06$	slope=0.087 $R^2=0.21$**	slope=−0.013 $R^2=0.38$***
JW11	1–2 orientation *versus* vowel duration	slope=−0.086 $R^2=0.16$*	slope=−0.082 $R^2=0.20$**	slope=−0.121 $R^2=0.26$**
JW29	1–2 orientation *versus* vowel duration	slope=−0.144 $R^2=0.72$***	slope=−0.159 $R^2=0.81$***	slope=−0.119 $R^2=0.65$***
JW11	2–3 orientation *versus* vowel duration	slope=−0.126 $R^2=0.40$***	slope=−0.206 $R^2=0.56$***	slope=0.028 $R^2=0.04$
JW29	2–3 orientation *versus* vowel duration	slope=0.071 $R^2=0.77$***	slope=0.027 $R^2=0.15$*	slope=−0.001 $R^2=0.03$

vowel duration. For JW29, F1 showed highly significant linear trends with vowel duration at both the 1–2 and 2–3 secant line vowel critical times, just as for her /ɪ/ tokens (slope=11.52, $R^2=0.82$, $p<0.01$ at 1–2 vct, and slope =10.39, $R^2=0.72$, $p<0.01$ at 2–3 vct). For JW11, the linear relation between F1 and vowel duration was significant only at the 2–3 secant line vcts (slope=5.24, $R^2=0.26$, $p<0.05$ at 2–3 vct). However, only JW11 showed a highly significant linear trend in F2 with vowel duration in /ɛ/ (slope=6.78, $R^2=0.34$, $p<0.01$ at 1–2 vct, and slope=11.29, $R^2=0.32$, $p<0.01$ at 2–3 vct).

Further insight is gained when the dependencies of the formant frequencies on the articulatory measures are examined. F1 of JW11 showed a significant dependence on the 1–2 secant line and a very significant dependence on the 2–3 secant line midpoint-to-palate distances (Table V, rows 1 and 2, column 1). The linear trend of JW11's F1 versus secant line orientation was negative and highly significant for both the 1–2 and 2–3 secant lines (Table V, rows 3 and 4, column 1). The F2 versus midpoint-to-palate distance and F2 versus orientation relations were, at least, very significant for each of JW11's secant lines (Table V, column 2). For JW29's F1 there was a highly significant dependence on both the 1–2 secant line midpoint-to-palate distance and orientation (Table V, rows 1 and 3, column 3). There was little dependence of JW29's F2 on any articulatory quantity, except for a significant negative slope for F2 versus 1–2 orientation. A significant *negative* trend in F1 versus the 2–3 secant line midpoint-to-palate distance was recorded (Table V, row 2, column 3).

There is further evidence of individual differences in articulatory-acoustic relations in /ɛ/ production when horizontal and vertical dimensions of the midpoints are considered separately. The secant line that shows the most significant trend in the midpoint-to-palate distance versus F1 is the 2–3 secant line for JW11 and the 1–2 secant line for JW29. For JW11, a highly significant linear relation exists between the horizontal coordinate of the 2–3 secant line midpoint and F1 (slope=−0.016, $R^2=0.48$, and $p<0.01$), but not for the vertical coordinate at the 2–3 secant line vct (slope =−0.010, $R^2=0.13$). For JW29, the linear fits between F1 and the horizontal coordinate and between F1 and the vertical coordinate were both highly significant (slope=−0.013, $R^2=0.65$, and $p<=0.01$ and slope=−0.010, $R^2=0.36$, and $p<0.01$, respectively).

IV. DISCUSSION

The results indicate that phonetic vowel reduction occurred for both subjects, to varying degrees, as measured in terms of midpoint-to-palate distances, orientations, and formant frequencies. After reviewing the important, particular facts about the dataset examined here, we will summarize the findings regarding articulatory changes with vowel duration. The second part of the discussion will focus on the physics of the observed relations between articulatory and acoustic trends. This will entail some discussion on individual differences.

There are three facts that have relevance to the following discussion. First, the limited range of some of the observations could have had an effect on the linear regressions. In some cases a lack of sufficient range in the midpoint-to-palate or secant line orientation could have made it difficult to find significant linear trends, because the variability due to noise might have outweighed variability that resulted in regular linear trends. The lack of range could have been the result of a compensatory articulatory strategy, as discussed for JW11's /ɪ/ in Sec. III A. Second is the fact that the same secant line probably spanned different functional regions of the tongue for the two subjects, as discussed in Sec. II. The final fact is that the subjects had substantially different palate shapes.

The following general trends in articulation as a function of vowel duration have been noted using Tables II and IV. The 1–2 secant line midpoint-to-palate distances at vct increase with vowel duration for /ɪ/ and /ɛ/ for both subjects. For JW11's 2–3 secant line there was no such trend for /ɪ/, but there was for /ɛ/. The lack of a trend for JW11's /ɪ/ could have been due to his limited range of vowel durations for this vowel. For JW29's 2–3 secant line, there was no consistent trend in the midpoint-to-palate distance for /ɛ/, but there was

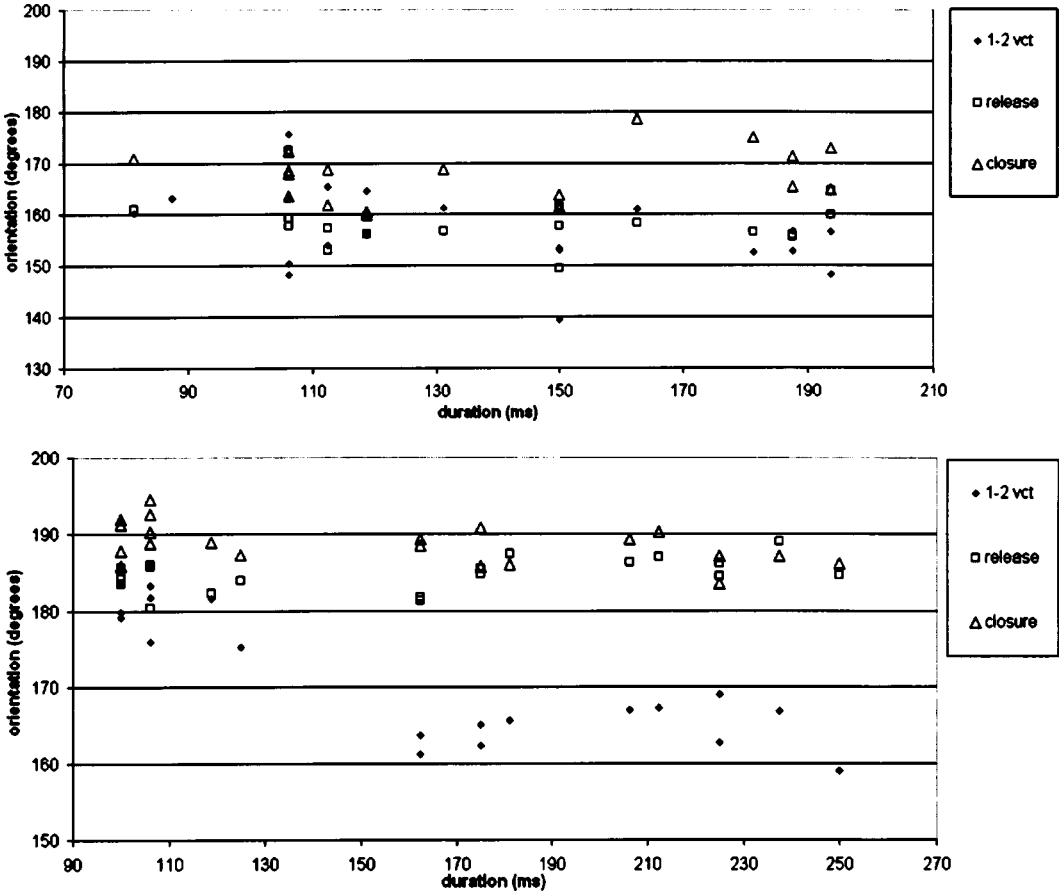


FIG. 6. 1–2 secant line orientation (degrees) at release, at closure, and at 1–2 vowel critical time versus vowel duration (ms) for tokens of /ɛ/ in the words ''sense'' and ''said''. (a) JW11; (b) JW29.

for /ɪ/. The differences between the subjects in the behavior of the 2–3 secant line was highlighted when orientations were considered.

JW29's 1–2 secant line showed increased downward (i.e., more negative) orientation with increased vowel duration to higher degrees of significance than did JW11's 1–2 secant line for /ɪ/ and /ɛ/. The opposite is true for the 2–3 secant line, with JW11's showing a more significant increase in downward orientation with vowel duration than JW29's, for both vowels. In fact, JW29's 2–3 secant line had some significant increases in orientation with vowel duration. The differences in the behavior of JW29's and JW11's 2–3 secant line could be attributed to a difference in the functional parts of the tongue spanned by this secant line for the two subjects. Palate shape differences could also have played a role.

We now turn to the relation between articulation and acoustics reported in Tables III and V. There were no significant trends between F1 and the midpoint-to-palate distance for JW11 in the case of /ɪ/, while there were highly significant trends for JW29 (Table III). This could have been the result of his limited range of midpoint-to-palate distances for this subject. The ranges of midpoint-to-palate distances for JW11's /ɪ/ productions were less than 3.5 mm for both the 1–2 and 2–3 secant lines, while the range for JW29's 1–2

TABLE V. Statistics for linear regression of the secant line midpoint-to-palate distance and orientation versus formant frequencies at the vowel critical times (vct) for JW11 and JW29 producing /ɛ/ in ''sense and said'' (*=0.1 confidence,**=0.05 confidence, and***=0.01 confidence). Distances are in millimeters, orientations in degrees, and frequencies in Hz.

Versus	JW11 F1	JW11 F2	JW29 F1	JW29 F2
1–2 midpoint-to-palate distance	slope= 0.0021 $R^2=0.16$*	slope= 0.0024 $R^2=0.50$***	slope= 0.0015 $R^2=0.60$***	slope= 0.0013 $R^2=0.09$
2–3 midpoint-to-palate distance	slope= 0.0017 $R^2=0.28$**	slope= 0.0009 $R^2=0.32$***	slope= 0.0004 $R^2=0.15$*	slope= 0.0000 $R^2=0.00$
1–2 orientation	slope=−0.11 $R^2=0.41$***	slope=−0.06 $R^2=0.24$**	slope=−0.06 $R^2=0.54$***	slope=−0.07 $R^2=0.15$*
2–3 orientation	slope=−0.08 $R^2=0.45$***	slope=−0.05 $R^2=0.70$***	slope=−0.02 $R^2=0.22$**	slope=0.01 $R^2=0.02$

secant line was 6.9 and 3.6 mm for her 2–3 secant line in /ɪ/ (see Figs. 4 and 5). In the case of JW29's /ɪ/ productions, these trends in F1 versus midpoint-to-palate distances provide examples of the well-known relation between tongue height and F1. An increase in the area of the front part of the vocal tract can be conceived as a decrease in the magnitude of the mass element of the Helmholtz resonator for high articulations. For less close articulations, increases in the midpoint-to-palate distance represented increases in the portion of the acoustic tube that contains the volume velocity maximum for the standing wave associated with F1, in an application of perturbation theory (e.g., Schroeder, 1967; Stevens, 1998, pp. 148–152).

For /ɛ/ the positive F1 versus midpoint-to-palate distance trend was most significant for the 2–3 secant line of JW11 and the 1–2 secant line of JW29. The difference between subjects appears partly to be due to the differences in functional regions spanned by their 2–3 secant lines. However, it can be noted that changes in F1 were not strictly related to tongue height, because the shape of the palate plays a role. The fact that the relation between the horizontal coordinate of JW11's 2–3 secant line midpoint and F1 was highly significant, but there was no significant relation between the vertical component and F1, could be due to the fact that the 2–3 secant line opposed the relatively steep diagonal of JW11's palate. Thus, pulling the tongue back for JW11's /ɛ/ production had the same effect as JW29's pulling the tongue down and back from her flatter palate: an increase in F1.

The palate shape differences could also explain why there were significant increases of F2 with midpoint-to-palate distances for JW11 in the vowel /ɛ/ and no significant trends between midpoint-to-palate distances and F2 for JW29 in either vowel. The highly significant increase of JW11's F2 in /ɛ/ with 1–2 and 2–3 midpoint-to-palate distances could have been caused in two ways. The 1–2 and 2–3 segments of JW11 could have been far enough forward that their translations affected only the maximum velocity portion of the F2 standing wave, or the increased midpoint-to-palate distance of these segments was coincident with a decreasing orientation angle, which actually caused the F2 increase. The reason to expect an increased F2 with a decreased orientation angle is explained below.

Based on perturbation theory, it can be expected that both F1 and F2 of /ɛ/ increase with decreasing orientation angle for secant lines in the front of the tongue. Decreasing orientation angles in the front secant lines imply that the regions for velocity maxima of the F1 and F2 have an increasing cross-sectional area, while the regions posterior to this, which could be associated with either regions of pressure maxima or regions that are neither velocity or pressure maxima, possess decreasing areas. There were the expected, and very significant, negative trends between F1 and 1–2 and 2–3 secant line orientation in the vowel /ɛ/ for JW11 and JW29, and in the case of JW11, between F2 and 1–2 and 2–3 secant line orientation (Table V). However, it must be kept in mind that these effects could simply be due to the coincidence of increasing the midpoint-to-palate distance and decreasing orientation, with the physical cause for changing F1 and F2 being the increasing midpoint-to-palate distance. The significances of trends for /ɪ/ (Table III) were probably hindered by a lack of range of orientations. For JW11, the ranges of orientations for both secant lines in /ɪ/ were less than 15°, compared to ranges greater than 25° for /ɛ/. The JW29 ranges of orientations for both secant lines in /ɪ/ were less than 16°, compared to ranges greater than 17° for /ɛ/. The fact that the 2–3 secant line appeared to have consistent effects on F1 and F2 for JW11, but not for JW29, may have been the result of the fact that the functional extent of the tongue blade accounted for between the first and third pellets is greater for JW29 than for JW11.

In summary, while we have claimed that pellet placement did have an effect on how tongue placement in vowel production was measured, we believe that there were speaker differences that cannot be attributed to any differences to pellet placement differences. In effect we have added a small amount of evidence for the conclusion of Johnson, Ladefoged, and Lindau, "*We are left with the conclusion that speakers of the same language and dialect may use different articulatory plans; that the universal articulatory phonetics hypothesis is wrong*" (Johnson *et al.*, 1993, p. 713). It can be seen here that some of this difference could be the result of palate shape differences. This also has implications for the normalization procedure of Hashi, Westbury, and Lindstrom (1998), which essentially maps the midsagittal palate outlines onto a straight line for all speakers. They found that the variation in the pellet-to-palate distance between talkers was reduced with this mapping. We have an analogous normalization with the secant line midpoint-to-palate distance measure. Our results indicate that if such a normalization procedure is to be employed, then secant line orientations need to be measured in relation to the palate, and the oral space near the palate needs to be warped to account for directions of translational tongue movement in relation to the local palate surface.

V. CONCLUSION

In this specialized sample of lax, front vowels in alveolar or dental consonant context, both midpoint-to-palate distance differences and orientation differences contributed to acoustic assimilation to the surrounding consonants with decreasing vowel duration. Further, speaker-dependent differences in articulation and the formant frequency-articulation relations appeared to be, in part, attributable to palate shape differences. Secant line coordinates is a representation that is closely related to flesh point data that successfully emphasizes the relation between local tongue shape and acoustics.

There are at least two directions for current research to continue. The first is to examine the same phenomena in other, so-called, front vowels, such as /i/ and /æ/, and to treat linguistic factors, such as stress, separately from duration. Other extensions involve studying the properties of the secant line trajectories themselves. Data-driven articulatory synthesis should be achievable through the control of secant line kinematics.

ACKNOWLEDGMENTS

This work was supported by Grant No. NIDCD-01247 to CReSS LLC. The comments of Anders Löfqvist, Tine Mooshammer, and an anonymous reviewer helped to greatly improve this paper.

Adams, S. G., Weismer, G., and Kent, R. D. (**1993**). ''Speaking rate and speech velocity movement profiles,'' J. Speech Hear. Res. **36**, 41–54.

van Bergem, D. R. (**1993**). ''Acoustic vowel reduction as a function of sentence accent, word stress, and word class,'' Speech Commun. **12**, 1–23.

Broad, D. J., and Fertig, R. H. (**1970**). ''Formant-frequency trajectories in selected CVC-syllable nuclei,'' J. Acoust. Soc. Am. **47**, 1572–1582.

Broad, D. J., and Clermont, F. (**1987**). ''A methodology for modeling vowel formant contours in CVC context,'' J. Acoust. Soc. Am. **81**, 155–165.

Di Benedetto, M.-G. (**1989**). ''Vowel representation: Some observations on temporal and spectral properties of the first formant frequency,'' J. Acoust. Soc. Am. **86**, 55–66.

Engstrand, O. (**1988**). ''Articulatory correlates of stress and speaking rate in Swedish VCV utterances,'' J. Acoust. Soc. Am. **83**, 1863–1875.

Flege, J. E. (**1988**). ''Effect of speaking rate on tongue position and velocity of movement,'' J. Acoust. Soc. Am. **84**, 901–916.

Fujimura, O., Kiritani, S., and Ishida, H. (**1973**). ''Computer controlled radiography for observation of movements of articulatory and other human organs,'' Comput. Biol. Med. **3**, 371–384.

Gay, T. (**1978**). ''The effect of speaking rate on vowel formant movements,'' J. Acoust. Soc. Am. **63**, 223–230.

Harris, K. S. (**1978**). ''Vowel duration change and its underlying physiological mechanisms,'' Lang Speech **21**, 354–361.

Hashi, M., Westbury, J. R., and Honda, K. (**1998**). ''Vowel posture normalization,'' J. Acoust. Soc. Am. **104**, 2426–2437.

Hoole, P. (**1999**). ''On the lingual organization of the German vowel system,'' J. Acoust. Soc. Am. **106**, 1020–1032.

Johnson, K., Ladefoged, P., and Lindau, M. (**1993**). ''Individual differences in vowel production,'' J. Acoust. Soc. Am. **94**, 701–714.

Keating, P. A., Lindblom, B., Lubker, J., Kreiman, J. (**1994**). ''Variability in jaw height for segments in English and Swedish VCVs,'' J. Phonetics **22**, 407–422.

Kuehn, D. P., and Moll, K. L. (**1976**). ''A cineradiographic study of VC and CV articulatory velocities,'' J. Phonetics **4**, 303–320.

Lindblom, B. (**1963**). ''Spectrographic study of vowel reduction,'' J. Acoust. Soc. Am. **35**, 1773–1781.

Lindblom, B. (**1964**). ''Articulatory activity in vowels,'' KTH Speech Technology Laboratory-QPSR, Vol. 2/1964, pp. 1–5.

Milenkovic, P. (**2003**). Personal communication.

Moon, S.-J., and Lindblom, B. (**1994**). ''Interaction between duration, context, and speaking style in English stressed vowels,'' J. Acoust. Soc. Am. **96**, 40–55.

Nord, L. (**1986**). ''Acoustic studies of vowel reduction in Swedish,'' KTH Speech Transmission Laboratory-QPSR, Vol. 4/1989, pp. 19–36.

Ohde, R. N., and Scharf, D. J. (**1975**). ''Coarticulatory effects of voiced stops on the reduction of acoustic targets,'' J. Acoust. Soc. Am. **58**, 923–927.

Ostry, D. J., and Munhall, K. G. (**1985**). ''Control of rate and duration of speech movements,'' J. Acoust. Soc. Am. **77**, 640–648.

Perkell, J. S. (**1991**). ''On the nature of distinctive features: Implications of a preliminary vowel production study,'' in *Frontier of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, London), pp. 365–380.

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., and Jackson, M. T. T. (**1992**). ''Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements,'' J. Acoust. Soc. Am. **92**, 3078–3096.

Perkell, J. S., Zandipour, M., Matthies, M. L., and Lane, H. (**2002**). ''Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues,'' J. Acoust. Soc. Am. **112**, 1627–1641.

Schroeder, M. R. (**1967**). ''Determination of the geometry of the human vocal tract by acoustic measurements,'' J. Acoust. Soc. Am. **41**, 1002–1010.

Simpson, A. P. (**2002**). ''Gender-specific articulatory-acoustic relations in vowel sequences,'' J. Phonetics **30**, 417–435.

van Son, R. J. J. H., and Pols, L. C. W. (**1990**). ''Formant frequencies of Dutch vowels in a text read at normal and fast rate,'' J. Acoust. Soc. Am. **88**, 1683–1693.

van Son, R. J. J. H., and Pols, L. C. W. (**1992**). ''Formant movements of Dutch vowels in a text read at normal and fast rate,'' J. Acoust. Soc. Am. **92**, 121–127.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Stevens, K. N., and House, A. S. (**1963**). ''Perturbation of vowel articulation by consonantal context: An acoustical study,'' J. Speech Hear. Res. **6**, 11–128.

Stevens, K. N., House, A. S., and Paul, A. P. (**1966**). ''Acoustic description of syllable nuclei: An interpretation in terms of a dynamic model of articulation,'' J. Acoust. Soc. Am. **40**, 123–132.

Stone, M. (**1990**). ''Three-dimensional tongue movement,'' J. Acoust. Soc. Am. **87**, 2207–2217.

Westbury, J. R. (**1994**). *X-Ray Microbeam Speech Production Database User's Handbook* (unpublished manuscript, University of Wisconsin).

Westbury, J. R., Hashi, M., and Lindstrom, M. J. (**1998**). ''Differences among speakers in lingual articulation for American English /ɹ/,'' Speech Commun. **26**, 203–226.

The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts

Joseph S. Perkell,[a] Frank H. Guenther,[b] Harlan Lane,[c] Melanie L. Matthies,[d] Ellen Stockmann,[e] Mark Tiede,[e] and Majid Zandipour[f]

Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

(Received 14 March 2003; revised 30 April 2004; accepted 8 July 2004)

This study addresses the hypothesis that the more accurately a speaker discriminates a vowel contrast, the more distinctly the speaker produces that contrast. Measures of speech production and perception were collected from 19 young adult speakers of American English. In the production experiment, speakers repeated the words *cod, cud, who'd,* and *hood* in a carrier phrase at normal, clear, and fast rates. Articulatory movements and the associated acoustic signal were recorded, yielding measures of contrast distance between /ɑ/ and /ʌ/ and between /u/ and /ʊ/. In the discrimination experiment, sets of seven natural-sounding stimuli ranging from *cod* to *cud* and *who'd* to *hood* were synthesized, based on productions by one male and one female speaker. The continua were then presented to each of the 19 speakers in labeling and discrimination tasks. Consistent with the hypothesis, speakers with discrimination scores above the median produced greater acoustic contrasts than speakers with discrimination scores at or below the median. Such a relation between speech production and perception is compatible with a model of speech production in which articulatory movements for vowels are planned primarily in auditory space. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1787524]

PACS numbers: 43.70.Aj, 43.70.Bk, 43.71.Es [AL] Pages: 2338–2344

I. INTRODUCTION

Recent brain imaging studies have provided evidence supporting the hypothesis of an intimate relationship between speech production and speech perception. Investigators have shown that motor areas of the brain are active during speech perception (cf. Mathiak *et al.*, 2002; Rizzolatti and Arbib, 1998) and auditory areas are active during speech production (cf. Hickok and Poeppel, 2000). A study using transcranial magnetic stimulation showed an increase of motor-evoked potentials in listeners' tongue muscles when they heard words whose production strongly involves tongue movements (Fadiga *et al.*, 2002).

The current study addresses this hypothesis in another way, by seeking correlations between measures of production and perception across speakers. Specifically, we hypothesize that speakers who discriminate well between vowel stimuli with subtle acoustic differences will produce relatively more clear-cut vowel contrasts while speakers who are less able to discriminate between the same vowel stimuli will produce less clear-cut vowel contrasts.

This hypothesis is based on a model of speech production in which goals for vowel movements are regions in multidimensional auditory-temporal space (Guenther, 1995; Guenther *et al.*, 1998). In this model, DIVA,[1] speech motor planning is influenced by two competing constraints: the listener's need for clarity and the speaker's motivation to achieve an economy of effort (Guenther; 1995; Lindblom and Engstrand 1989). The degree of clarity is related to the amount of separation among the auditory goal regions for different sounds, which is determined by their location and size in auditory space. The model forms goal regions initially by monitoring sounds from the speaker's native language and learning, for each phoneme, the region of auditory space that encompasses examples of that phoneme (see Guenther *et al.*, 1998 for details). According to the way the model functions, speakers who can perceive fine acoustic-phonetic details will learn goal regions spaced further apart. This may be because they are more likely than people with less acute auditory perception to reject poorly produced tokens of a phoneme when learning the goal regions.

II. BACKGROUND

Studies of the relation between production and perception have employed a variety of experimental paradigms and measures. In the most common approach, experimenters examine a group of speakers who vary in measures of production and perception and look for relationships between the two types of measures. For example, Fox (1982) had 11 subjects judge similarities of synthetic vowel stimuli in paired comparisons and identified three underlying dimensions, two of which were interpreted as representing the first two formants. The dimensional weightings that best predicted a given speaker's judgments of vowel similarities correlated with the formant values of that speaker's produced corner

[a]Corresponding author; Research Laboratory of Electronics, M.I.T. Dept. of Brain and Cognitive Sciences, M.I.T.; Department of Cognitive and Neural Systems, Boston University, Boston, MA; email address: perkell@speech.mit.edu

[b]Also of Department of Cognitive and Neural Systems, Boston University, Boston, MA.

[c]Also of Department of Psychology, Northeastern University, Boston, MA.

[d]Also of Department of Communication Disorders, Boston University, Boston, MA.

[e]Also of Haskins Laboratories, New Haven, CT.

[f]Also of Department of Cognitive and Neural Systems, Boston University, Boston, MA.

vowels, providing evidence of a relation between production and perception. Recently, Newman (2003) found modest but significant cross-subject correlations in measures of 20 subjects' speech perception and properties of their speech productions. Speakers having perceptual prototypes with longer voicing onset times (VOTs) for the /p/ in /pɑ/ also tended to produce /pɑ/ with longer VOTs. Similarly, there was a relation between the frequency of the spectral peak of their prototypes and productions of the /ʃ/ in /ʃɑ/.

Other results have supported the idea that perception and production are linked, but somewhat less directly. Based on electromyographic (EMG) measurements, Bell-Berti *et al.* (1979) inferred that subjects formed two groups in their implementation of tense-lax differences among the vowels /i, ɪ, e, ɛ/: one group apparently used tongue height, the other, tongue tension. In a labeling test of an /i/ to /ɪ/ continuum with an anchoring condition, the magnitude of the induced boundary shift was greater for the tongue-height group. Frieda *et al.* (2000) had subjects produce the vowel /i/ in citation and hyperarticulated conditions and also had them select their ideal exemplar of /i/ from among 330 synthetic stimuli. A synthesized stimulus was classified as a prototype if chosen as an exemplar more than 13% of the time it was presented. According to this method, 24 of the 35 subjects had perceptual prototypes for /i/. The formant values of those perceptual prototypes were approximated more closely by subjects' hyperarticulated productions of /i/ than they were by their citation productions, but only for the subject group with relatively clear-cut prototypes.

There have also been studies of parallel changes in production and perception. For example, Bradlow *et al.* (1997) studied perceptual learning of English /r/ and /l/ by 11 Japanese speakers. The subjects recorded /r/ and /l/ productions before and after a number of sessions in which they were trained to identify /r/ and /l/. For measures of those productions, a group of English-speaking listeners identified and rated the phonemes produced before and after training. Improvements in these production measures covaried across speakers with measures of their perceptual learning. Rvachew (1994) obtained a similar finding when children were given perceptual training to correct phonological errors in their speech production.

In a study of eight postlingually deafened cochlear implant users, Vick *et al.* (2001) examined covariation of production and perception of vowel contrasts. Measures were made prior to the subjects' receiving their implants and 24–52 weeks postimplant. For the most part, subjects who produced vowel pairs with reduced contrast pre-implant and who showed improved perception of the contrast post-implant, also had enhanced production contrasts postimplant.

A link between perception and production has also been shown within individual subjects, using a ''sensorimotor adaptation'' paradigm. Houde and Jordan (1998, 2002) observed compensatory changes in the productions of vowels whispered by subjects whose auditory feedback had been altered. The feedback alteration shifted the first two formants along the /i, ɪ, ɛ, æ, ɑ/ axis, which had the effect of changing the identity of the vowel that was fed back to the subject. For example, when subjects initially pronounced *bed* and then were fed back an increasingly *bead*-like acoustic result, they compensated for the perceived vowel raising by vowel lowering and ended up pronouncing *bad*. The effect generalized to various consonant environments and vowels. For example, training with *get* yielded compensations in test words *geck, guess, debt*, and *pet* and also in test words *geet, git, gat*, and *got*. This demonstration of a tight link between production and perception within individual speakers is compatible with the functionality of mappings between articulatory and auditory frames of reference in DIVA (Guenther, 1995; Guenther and Ghosh, 2003).

Although there have been negative findings (cf. Paliwal *et al.*, 1983; Ainsworth and Paliwal, 1984—discussed further below), on balance, these rather diverse studies provide support for some kind of link between production and perception. Many of them are consistent with the idea that speakers who have relatively sensitive perceptual capabilities produce more distinct sound contrasts. The current study investigates this idea in more detail, using articulatory, acoustic, and perceptual measures, with an explicit, model-based hypothesis: Subjects who have relatively higher vowel discrimination scores will produce more distinct vowel contrasts than those who do not.

III. METHODS

Subjects: Production and perception experiments were performed on a group of 19 young adult speakers of American English, 9 females, and 10 males. The subjects were paid volunteers who had no history of speech or hearing disorders.

A. Production experiment

Each subject participated in a speech production experiment, in which his or her articulatory movements and speech signal were recorded.

1. Speech materials

The speech materials, which took an hour to read aloud, consisted of the words *cod, cud, who'd*, and *hood* embedded in the phrase, Say ___ hid it. There were 27 repetitions of each *cod* and *who'd* utterance and 9 repetitions of each *cud* and *hood* utterance (which were initially included as foils for a slightly different design). The two contrasts were chosen because the members of each contrasting pair are acoustically and articulatorily close to one another.[2] These utterances were part of a larger set that included other materials. In order to investigate the effects of speaking condition, the entire corpus was read in three different conditions, ''fast,'' ''normal,'' and ''clear.'' For the fast condition, the subject was asked to speak as rapidly as possible without eliminating any sounds. If necessary, the subject was reminded of this criterion as the experiment progressed. For the clear speech condition, the subjects were asked to pronounce the words carefully without increasing their loudness.

2. Recordings

An electromagnetic midsagittal articulometer system (EMMA—Perkell *et al.*, 1992) was used to record the posi-

tion versus time of points on the subject's tongue, lips and jaws in the midsagittal plane. The subject was seated in an adjustable chair in a sound-attenuated room. A directional microphone was positioned about 14 in. from the subject's lips, and the subject read three repetitions of the utterance set to make an acoustic recording without the movement transducer system in place.[3] Then the transmitter assembly of the EMMA system was fit snugly to the subject's head. Small EMMA transducer coils (2 mm×5 mm) were attached at the midline with biocompatible adhesive to the vermillion border of upper lip (UL) and lower lip (LL), the gingival papilla between the lower central incisors (LI), and three places on the tongue, 1 cm from the tongue tip (TT), the tongue blade (TB—about 3 cm from the tongue tip), and the tongue dorsum (TD—about 5 cm back from the tongue tip). Transducer coils were also attached to the bridge of the nose and the gingival papilla between the upper central incisors for a maxillary frame of reference.

A custom-written program was used to control the experiment, record the movement and acoustic signals to disk and display the utterance materials one at a time on a computer screen located about three feet in front of the subject. To allow monitoring of the progress of the experiment, the program also generated a real-time display of the acoustic signal and movement trajectories of the transducers.

The audio signal was low-pass filtered at 7.8 kHz and sampled at 16 kHz. Each signal channel from the EMMA system (three per transducer coil) was hardware low-pass filtered at 100 Hz and sampled at 500 Hz.

3. Data extraction

The data of interest were the *x* and *y* positions of a transducer coil on the tongue at the time the tongue reached its target position, and measures of the corresponding acoustic spectrum. As the first step in data extraction, one of the experimenters labeled the beginning and end of the target vowel in the sound pressure waveform for each token, using an interactive MATLAB procedure with displays of the sound pressure waveform and a spectrogram.

Approximately 4100 tokens were processed for this study (27 repetitions each of *cod* and *who'd*, 9 each of *cud* and *hood*, three conditions, 19 subjects). To make the analysis time tractable, algorithms were developed to automate the remaining data extraction steps.

The first step in the extraction of articulatory data for each token was the conversion of the raw EMMA voltage signals from each transducer to midsagittal-plane values of *x* (horizontal) and *y* (vertical) versus time. The voltage signals are first lowpass filtered using a Butterworth filter with a 100 Hz corner frequency and a roll-off rate of 18 dB/oct. Next, the transducer position versus time signals are translated and rotated into the maxillary frame of reference. This is followed by lowpass filtering using a ninth-order Butterworth with a 12.5 Hz corner frequency. For reasons explained below, the TB transducer was chosen as being most representative of the tongue body position for the vowel. Next, the time of the minimum in the velocity magnitude (speed) of the TB transducer during the vowel interval was located as the target point at which *x* and *y* data were extracted. For this purpose, TB velocity magnitude versus time was computed as the square root of the summed squares of central-differenced *x* and *y* values at each time step.

The first three formant frequencies were extracted with a method designed to minimize the occurrence of missing or spurious values. The audio signal was first pre-emphasized by first-order differencing (mu=0.98). Formants were obtained by peak-picking from a spectral envelope that was derived by averaging linear predictive coding (LPC) spectra taken over the interval from 30 to 40% of the delimited vowel interval using a sliding 30 ms window overlapped at 1 ms steps. In order to ameliorate problems with missing or incorrectly identified formants, each analysis was repeated using LPC orders of 16 through 22 inclusive, plus an ''optimal'' order chosen by a heuristic similar to that suggested by Vallabha and Tuller (2002). The intermediate result of this LPC analysis was eight largely overlapping sets of possible formant frequencies for each token (one for each LPC order). These frequency values from all tokens of a given vowel type were then binned within a histogram of 100 Hz resolution. Starting from expected values (mean results from Peterson and Barney, 1952) the closest peaks of the smoothed histogram were then used to determine subject-specific formant ''targets'' to use in choosing among the formants computed for each token, with the three closest values to the F1, F2, and F3 targets retained as the values for that token. Outliers exceeding two standard deviations were marked as missing data.

Two measures of vowel contrast were derived for each vowel pair, /ɑ-ʌ/ and /u-ʊ/, in each of the three speaking conditions, fast, normal, and clear. To derive *articulatory contrast distance* in each speaking condition, the mean values of TB*x* and TB*y* were calculated for each vowel and the Euclidean distance between the two sets of coordinates was found.

Correspondingly, to derive *acoustic contrast distance*, mean values of F1 and F2 were calculated for each vowel in each condition, and formant separation was calculated as the Euclidean distance between the means in the F1, F2 plane.

B. Perception experiments

The same 19 subjects also participated in perception experiments, consisting of vowel labeling and discrimination tasks.

1. Stimuli

For each of the two word pairs, *cod-cud* and *who'd-hood*, two sets of seven stimuli ranging from *cod* to *cud* and *who'd* to *hood* were synthesized, based on natural productions of the words in isolation by a male speaker and a female speaker (who were not subjects in the experiment). Values of the first three formants at 10 ms intervals during the vowel were calculated by interpolation between the naturally produced end-point values. The fourth and fifth formants were kept constant. The Klatt synthesizer (Klatt, 1980) as implemented in the SPEECHSTATION 2 SPEECH ANALYSIS WORKSTATION software (Sensimetrics, Inc., Somerville, MA) was used to generate the seven stimuli in each continuum. Within each continuum, the vowel portions of all seven

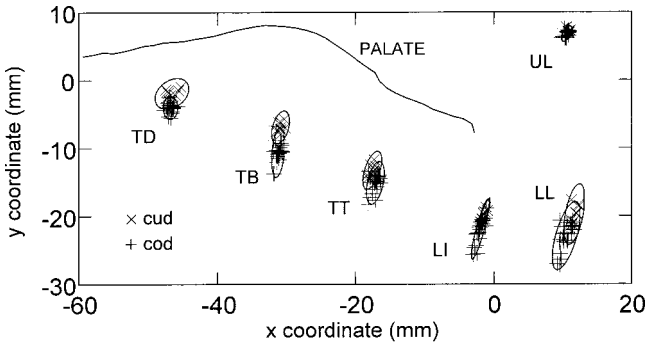


FIG. 1. EMMA transducer coil locations during all the productions of tokens containing the /ɑ/ in *cod* (+) and the /ʌ/ in *cud* (×) by one female subject in the normal condition. The transducer coils are located at points on the tongue dorsum (TD), tongue blade (TB), tongue tip (TT), lower incisor, (LI), lower lip (LL), and upper lip (UL). The midsagittal palatal contour (up to about the dento-alveolar junction on the right) is shown for reference.

stimuli were constructed to have the same naturally produced F0 contour and duration; the chosen duration was a compromise between the durations of the naturally produced endpoint utterances. For each stimulus, the synthesized vowel was inserted carefully between the same naturally produced initial-consonant release and final-consonant voicing signal so as to avoid discontinuities in the waveform. The resulting stimuli sounded quite natural in informal listening tests.

2. Tasks

Stimuli from all four of the continua (2 contrasts×2 genders) were presented to each of the subjects in labeling and discrimination tasks. Stimulus presentations for the tasks were created and administered using the Ecos/Win interface (AVAAZ Innovations, 1997 London, Ontario) and played through Bose noise-reducing headphones™. The labeling test was administered first, then the discrimination test. Stimuli were blocked by word pair, and all tasks were subject-paced.

The labeling task served as a control to verify the expected behavior of the listeners. In this task, each stimulus was presented individually and the subject was asked to identify the word using a computer mouse to select from the two choices on the monitor (*cod* or *cud*, *who'd* or *hood*). Each of the seven stimuli was presented 18 times.

The discrimination task was a classic ABX design (Liberman *et al.*, 1951). Stimuli were grouped into 60 sets of three stimuli where the first and second were one, two, or three steps apart on the synthesis continuum and the third was the same as either the first or second. After each set was played, the subject decided whether the third stimulus was the same as the first or second and indicated the decision by selecting button 1 or 2 on the computer screen using the mouse. Each set was played only once in a block and the task consisted of two blocks, for a total of 120 trials.

IV. RESULTS

Figure 1 is a plot of the EMMA transducer coil locations during all the productions of tokens containing the /ɑ/ in *cod* (+) and the /ʌ/ in *cud* (×) by one female subject in the normal condition. The transducer coils are located at points on the tongue dorsum (TD), tongue blade (TB), tongue tip (TT), lower incisor, (LI), lower lip (LL), and upper lip (UL). It was determined that the middle of the three tongue transducers TB represented the /ɑ/ target location with the least amount of coarticulatory influence of the preceding /k/ and the following /d/. Therefore, the analysis of the articulatory target for the /ɑ/ in *cod* and the /ʌ/ in *cud* focused on the location of the TB transducer. For uniformity across the four test words, the same transducer coil was used to represent the tongue body location during the /u/ in *who'd* and the /ʊ/ in hood. As exemplified in this figure, the TB category separation is the distance in millimeters between the centroids of the *cod* and *cud* TB distributions.

To clarify how the perceptual results were parametrized, Fig. 2 displays results for six of the subjects for the cod-cud continuum. These six illustrate the ranges of variation in labeling slopes and discrimination accuracy that were observed. Each panel contains a labeling function (filled circles connected by solid lines) whose values are the percentage of instances in which the presented stimulus (numbered 1–7 on the horizontal axis) was labeled *cod*. The remaining three functions show the percent correct for the one-step (number 1 connected by dotted lines), the two-step, and the three-step ABX stimuli. Subjects 3, 9, 14, and 7 are males and subjects 19 and 6 are females. The top three plots show responses to the male synthesized stimuli and the bottom three plots, responses to the female synthesized stimuli. There is considerable variation among the subjects in the steepness of the slope of their labeling functions and in the shapes and peak amplitudes of their discrimination functions. Such cross-speaker variation is consistent with previous findings (cf. Pisoni, 1971).

The one-, two-, and three-step stimulus comparisons were not equally successful in differentiating among the subjects' perceptual performances. The peak three-step discriminations produced a ceiling effect for most subjects, and the peak one-step stimuli led to highly variable scores, many at chance level. Moreover, many one-step intervals did not cross the labeling category boundary and the peak one-step scores did not correlate with any production measures. We retained, therefore, the peak discrimination scores obtained with stimuli separated by 2 steps on the synthetic continua.[4]

Figure 3 presents frequency distributions of the 19 subjects with respect to those two-step peak discrimination scores for each of the two vowel contrasts. For the following reasons, we classified subjects into two groups based on those scores. Even the two-step discrimination triads yielded substantial ceiling effects, with 37% of the subjects (*who'd-hood*) or 26% of the subjects (*cod-cud*) scoring 100 percent. Those subjects presumably had higher discriminative acuity than this discrimination task was able to measure. All subjects' scores fell at one of five values (*who'd-hood*) or one of four values (*cod-cud*) between 75 and 100% correct. (Scores varied in increments of 6.25% because there were 16 discrimination trials per two-step stimulus comparison.) Finally, the distribution for *who'd-hood* was right-skewed. These considerations and the observation that production contrast appeared to grow nonlinearly with phoneme discrimination led us to classify all subjects who scored 100% as ''high''

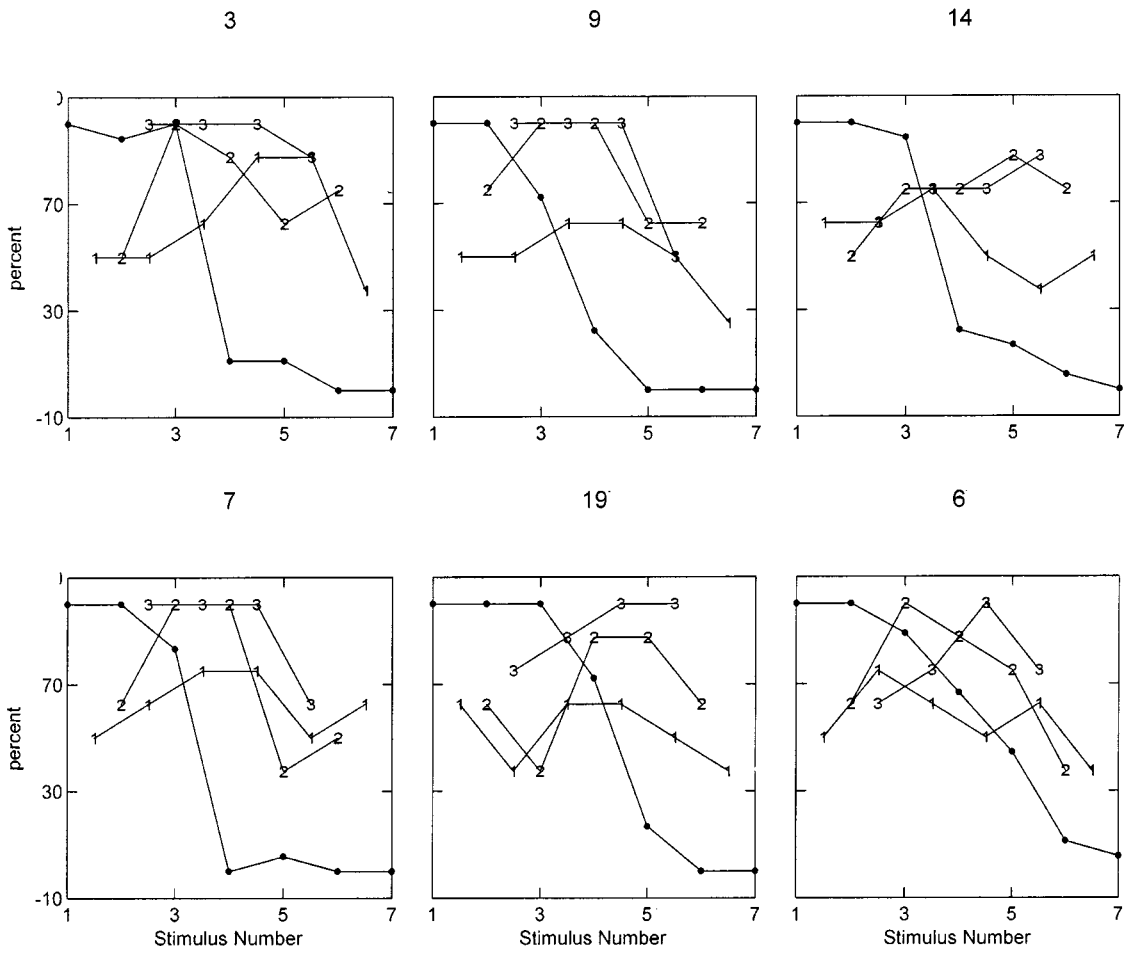


FIG. 2. Examples of perception test results for the cod-cud continua from six of the subjects. Each panel contains a labeling function (filled circles connected by solid lines) whose values are the percentage of instances in which the presented stimulus (numbered 1–7 on the horizontal axis) was labeled cod. The remaining three functions in each panel show the percent correct for the one-step (number 1 connected by dotted lines), the two-step, and the three-step ABX discrimination stimuli.

discriminators. All others subjects were classified as "low" discriminators. (As it turned out, high discriminators scored above the median peak two-step score and low discriminators at or below the median.)

Figure 4 plots articulatory contrast distance (for tongue body position) and acoustic contrast distance (for separation in the formant plane) in the upper and lower panels, respectively, as a function of the three speaking conditions. The left-hand panel gives results for *who'd-hood*, the right for *cod-cud*. For both articulatory and acoustic parameters and for both vowel contrasts, high discriminators (values labeled "*H*") produced greater contrast distances than low discriminators ("*L*") on the average. This effect was reliable for the articulatory parameter for *who'd-hood* ($F=28.6$, df=1,131, $p<.01$) and *cod-cud* ($F=17.9$, df=1,116, $p<0.01$). The effect was also reliable for the acoustic parameter for *cod-cud* ($F=26.4$, df=1,70, $p<0.01$) but not for *who'd-hood* where unexplained between-speaker variability was great ($F=1.6$, df=1,73, $p>0.05$; error bars are one standard error of the mean). As inspection of Fig. 4 reveals, speaking condition had an effect on both articulatory and acoustic contrast distance for *who'd-hood* (respectively, $F=49.1$, df=2,262, $p<.01$; $F=5.5$, df=2,146, $p<0.01$) but not for *cod-cud* ($F=0.5$, df=2,232, $p>.05$; $F=1.8$, df=2,140, $p>0.05$).

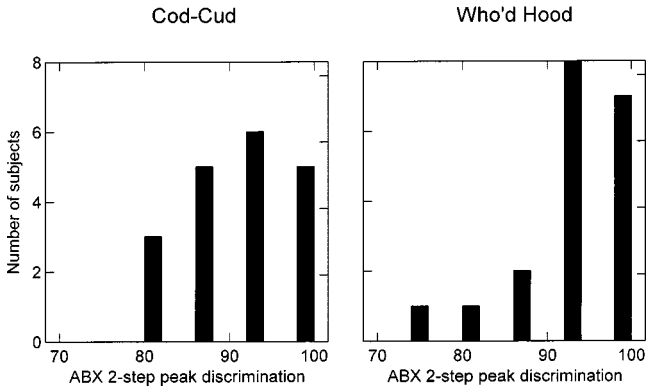


FIG. 3. Frequency distributions of the 19 subjects with respect to two-step peak discrimination scores on an ABX task for each of the two vowel contrasts.

V. DISCUSSION

We have found, for two vowel contrasts, that the more accurately a speaker discriminates a contrast, the more distinctly the speaker produces that contrast. This finding is consistent with and extends the results of other investigations

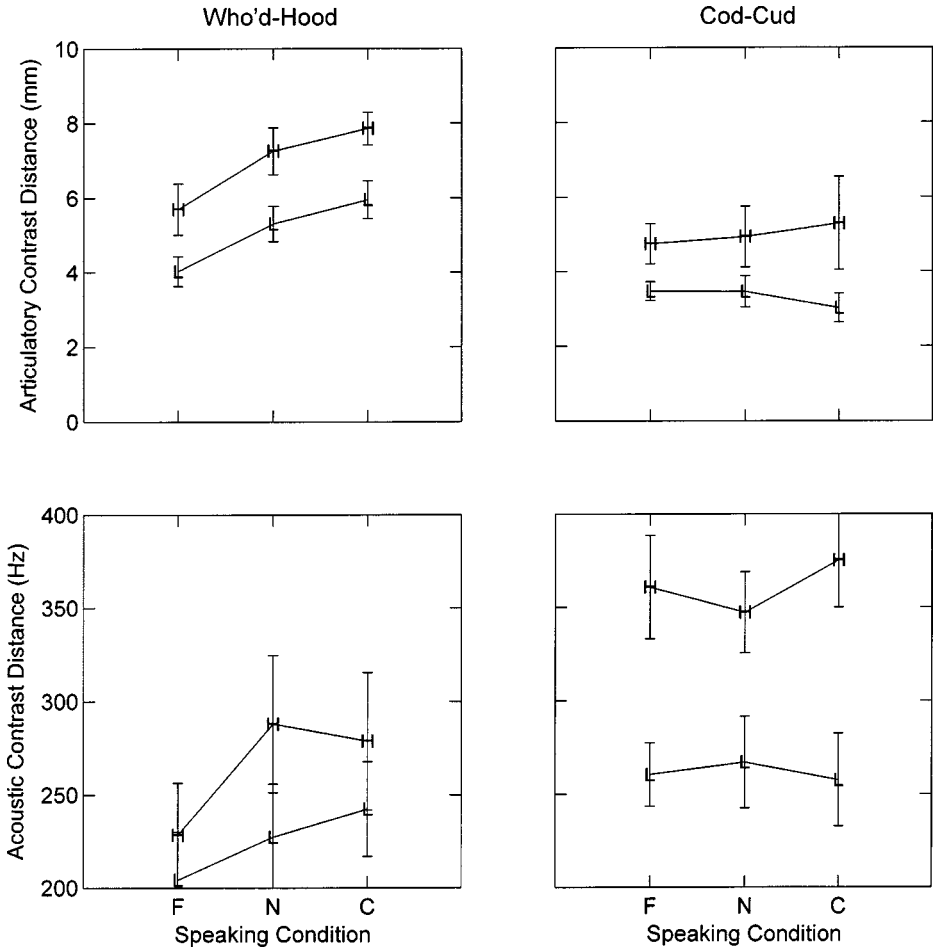


FIG. 4. Articulatory contrast distance (for tongue body position—upper panel) and acoustic contrast distance (for separation in the formant plane—lower panel) as a function of the three speaking conditions. The left-hand panel gives results for *who'd-hood*, the right for *cod-cud*. Findings for high discriminators (labeled "*H*") and low discriminators ("*L*") are plotted separately. Error bars are one standard error about the mean.

(see Background; also Ladefoged *et al.*, 1972). It is also compatible with the results of Perkell *et al.* (in press), who studied individual differences in producing the sibilant contrast in American English and their relation to two speaker characteristics: speakers' use of a quantal biomechanical effect in producing the sibilants and their performance on a test of sibilant discrimination. Parallel to the current results, Perkell *et al.* (in press) found that speakers with more acute sibilant discrimination produced greater sibilant acoustic contrast distances.

In Sec. II, we cited two studies with objectives and methods similar to the present experiment, which failed to find support for the hypothesis that speakers who discriminate a contrast more acutely produce that contrast more distinctly. Paliwal *et al.* (1983) performed a vowel production and perception experiment on ten subjects to evaluate the hypothesis that a listener refers to his own articulation for perceiving speech. Although this hypothesis, that production regulates perception, is the converse of ours, their general objective was the same: to find relations between production and perception. The subjects produced 11 vowels of English in /hVd/ context and they identified synthesized vowels (in the same context) in which values of F1 and F2 covered the entire F1, F2 plane. The results of correlations of the formant frequencies of produced and perceived vowels, within and between subjects, led to rejection of the hypothesis. Ainsworth and Paliwal (1984) used a similar approach in studying English glides /w, r, l, j/, with the same results and conclusion. There are several differences between these two studies and the current one that could account for the different outcomes, most notably the fact that we used discrimination scores not identification results as the perceptual measure.

As mentioned in the Introduction, our general hypothesis, that perception influences production, is compatible with the DIVA model (Guenther, 1995; Guenther *et al.*, 1998) in which the basic phonemic units for vowels are multidimensional regions in auditory-temporal space. These regions are utilized in speech perception and they are also goals for the planning of articulatory movements.

We hypothesize that language learners find it advantageous in their communicative interactions to be as intelligible as possible across a range of acoustic transmission conditions; therefore, according to the functionality of DIVA, speakers who have more acute perception of fine acoustic differences between vowels will learn auditory goal regions for vowels that are smaller and spaced further apart than speakers with less acute vowel perception. Such differences in goal regions among speakers would account for the current experimental results.

ACKNOWLEDGMENTS

This research was supported by Grant No. DC01925 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

[1]"DIVA" stands for "directions into velocities of articulators."

[2]Other contrasts could have been considered; however, a more systematic exploration of different kinds of distinctions (i.e., tense/lax, low/high, back/front) is beyond the scope of the current study.

[3]These utterances were used to verify informally that vowel productions with the EMMA system in place were not distorted by the presence of the EMMA transducer coils after a short period of accommodation.

[4]For both sets of continua (*who'd-hood* and *cod-cud*), correlations across subjects between peak one-step percent correct and peak two-step percent correct were not significant. A cross-subject correlation between values of peak two-step discrimination percent correct for *who'd-hood* and those for *cod-cud* was close to but missed significance ($r = 0.43$, $p = 0.068$).

Ainsworth, W. A., and Paliwal, K. K. (**1984**). "Correlation between the production and perception of the English glides /w, r, l, j/," J. Phonetics **12**, 237–243.

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., and Sawusch, J. R. (**1979**). "Some relationships between speech production and perception," Phonetica **36**, 373–383.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Fadiga, L., Craighero, L., Buccino, G., and Rizzolati, G. (**2002**). "Speech listening specifically modulates the excitability of tongue muscles: a TMS study," Eur. J. Neurosci. **15**, 399–402.

Fox, R. A. (**1982**). "Individual variation in the perception of vowels: Implications for a perception-production link," Phonetica **39**, 1–22.

Frieda, E. M., Walley, A. C., Flege, J. E., and Sloane, M. E. (**2000**). "Adults' perception and production of the English vowel /i/," J. Speech Lang. Hear. Res. **43**, 129–143.

Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation and rate effects in a neural network model of speech production," Psych. Review **102**, 594–621.

Guenther, F. H., and Ghosh, S. S. (**2003**). "A model of cortical and cerebellar function in speech," Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Aug. 3–9 (Universitat Autònoma de Barcelona, Barcelona, Spain), 169–174.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). "A theoretical investigation of reference frames for the planning of speech movements," Psych. Review **105**, 611–633.

Hickok, G., and Poeppel, D. (**2000**). "Towards a functional neuroanatomy of speech perception," Trends Cogn. Sci. **4**, 131–138.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295–310.

Klatt, D. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Ladefoged, P., DeClerk, J., Lindau, M., and Papcun, G. (**1972**). "An auditory-motor theory of speech production," UCLA Working Papers in Phonetics **22**, 48–94.

Liberman, A. M., Harris, K. S., Kinney, J. A., and Lane, H. (**1951**). "The discrimination of relative onset time of the components of certain speech and nonspeech patterns," J. Exp. Psychol. **61**, 379–388.

Lindblom, B., and Engstrand, O. (**1989**). "In what sense is speech quantal?" J. Phonetics **17**, 107–121.

Mathiak, K., Hertrich, I., Grodd, W., and Ackermann, H. (**2002**). "Cerebellum and speech perception: a functional magnetic resonance imaging study," J. Cogn Neurosci. **14**, 902–912.

Newman, R. S. (**2003**). "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report," J. Acoust. Soc. Am. **113**, 2850–2860.

Paliwal, K. K., Lindsay, D., and Ainsworth, W. A. (**1983**). "Correlation between production and perception of English vowels," J. Phonetics **11**, 77–83.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M., (**1992**). "Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3078–3096.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., and Stockmann, E. (in press). "The distinctness of speakers' /s-ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect," J. Speech Lang. Hear. Res.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Pisoni, D. B. (**1971**). "On the Nature of Categorical Perception of Speech Sounds," Supplement to Status Report on Speech Research, Haskins Laboratories, New Haven, CT.

Rizzolatti, G., and Arbib, M. A. (**1998**). "Language within our grasp," Trends Neurosci. **21**, 188–194.

Rvachew, S. (**1994**). "Speech perception training can facilitate sound production learning," J. Speech Lang. Hear. Res. **37**, 347–357.

Vallabha, G., and Tuller, B., (**2002**). "Systematic errors in the formant analysis of steady-state vowels," Speech Commun. **38**, 141–160.

Vick, J., Lane, H., Perkell, J. S., Matthies, M. L., Gould, J., and Zandipour, M. (**2001**). "Covariation of cochlear implant users' perception and production of vowel contrasts and their identification by listeners with normal hearing," J. Speech Lang. Hear. Res. **44**, 1257–67.

The control of aerodynamics, acoustics, and perceptual characteristics during speech production

Jessica E. Huber[a)]
Purdue University, Department of Audiology and Speech Sciences, West Lafayette, Indiana 47907

Elaine T. Stathopoulos and Joan E. Sussman
University at Buffalo, Department of Communicative Disorders and Sciences, Buffalo, New York 14214

(Received 7 January 2004; revised 21 May 2004; accepted 30 June 2004)

One of the most important areas of study in speech motor control is the identification of control variables, the variables controlled by the nervous system during motor tasks. The current study examined two hypotheses regarding control variables in speech production: (1) pressure and resistance in the vocal tract are controlled, and (2) perceptual and acoustic accuracy are controlled. Aerodynamic and acoustic data were collected on 20 subjects in three conditions, normally (NT), with an open air pressure bleed tube in place (TWB), and with a closed bleed tube in place (TNB). The voice recordings collected from the speakers in the production study were used in the perceptual study. Results showed that oral pressure (P_o) was significantly lower in the TWB condition than in the NT and TNB conditions. The P_o in the TWB condition seemed to be related to maintenance of subglottal pressure (P_s). Examination of the perceptual and acoustic data indicated that perceptual accuracy for [α] was achieved by maintaining P_s to preserve a steady sound pressure level, fundamental frequency, and voicing. Overall, it appeared speakers controlled pressure in compensating, but for the ultimate goal of maintaining acoustic and perceptual accuracy. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1785571]

PACS numbers: 43.70.Bk, 43.70.Aj [AL] Pages: 2345–2353

I. INTRODUCTION

One of the most important areas of study in motor control in general, and in speech motor control specifically, is the identification of control variables, the variables controlled by the nervous system during motor tasks. This was demonstrated by Stein's (1982) review of the available literature in an attempt to answer the question ''What muscle variable(s) does the nervous system control in limb movements?'' This review was followed by numerous responses from scientists regarding control variables in limb movement and speech. The focus on the control variables in movement is related to the belief that knowing the goals of movement leads directly to information about motor programming. Understanding motor control is important both to the study of normal speech behavior and to the study of treatment for speech disorders associated with neuromotor control disorders.

The current paper presents evidence for control variables in speech production as they relate to two specific hypotheses about what is controlled in speech production. One of these hypotheses was presented by Warren and colleagues (cf., Warren, 1986; Warren *et al.*, 1990, 1989, 1981), who suggest air pressure and resistance are the variables controlled during speech. These authors have presented data from numerous studies of normal speakers and individuals with velopharyngeal incompetence indicating that the speech subsystems actively compensate to maintain a constant air pressure and resistance in the vocal tract. Another hypothesis, that of Perkell and colleagues (cf., Guenther, 1995; Guenther *et al.*, 1998; Perkell *et al.*, 1995, 1997), suggests that acoustic and perceptual accuracy are the variables controlled during speech. These authors, too, have a substantial amount of data to support their hypothesis, using both normal speakers and individuals with hearing impairment. Motor equivalence, the idea that more than one articulatory configuration can result in a single percept, is often used to support Perkell and colleagues' theory (Guenther *et al.*, 1999).

One of the difficulties in studying control variables in speech is that often speech production data can be interpreted to support more than one theory. For example, according to Warren, in the attempt to control pressure and resistance in the vocal tract, perceptual adequacy may be compromised (Warren, 1986; Warren *et al.*, 1990, 1989). One example which Warren uses to demonstrate this idea is that individuals with cleft palate or velopharyngeal incompetence often use glottal stops in the place of oral stops. Warren hypothesizes that this substitution allows these speakers to maintain pressure and resistance in the vocal tract, but at the expense of perceptual adequacy. It could be argued that an alternate explanation for the use of a glottal stop is an unsuccessful attempt to control perceptual accuracy. In other words, since these individuals can not produce a perceptually accurate oral stop, they use a sound which is perceptually accurate in manner, but not place. The difference between these two theories is the hypothetical reason for the glottal stop. Warren's theory suggests that the maintenance of manner is unimportant as long as pressure and resistance in the vocal tract are controlled. However, in Perkell and colleagues' theory, the acoustic-perceptual goal is primary and the maintenance

a)Electronic mail: jhuber@purdue.edu

of the pressure characteristics of the stop is one of the ways acoustic-perceptual accuracy could be maintained.

In his review of limb control variables, Stein (1982) concluded that ''there does not seem to be a single control variable in all movements produced by muscles'' (p. 540). He states that *multiple* control variables are regulated by the neuromuscular mechanism. It would be logical to think that this is true in speech as well. Moon and Folkins (1991) demonstrated the multiplicity of control variables relative to perceptual characteristics and aerodynamics. These authors found that when they attenuated the auditory feedback for the frication for [f], speakers increased intraoral pressure, indicating that perceptual accuracy is controlled to some extent in speech production. However, the speakers did not increase intraoral pressure as much as would be required to compensate completely for the reduction in frication intensity. This may be explained in two ways: (1) the speakers were concerned with the control of articulatory aerodynamics and may have been keeping intraoral pressure below a specific aerodynamic criterion, or (2) speakers could not increase intraoral pressure as much as would be required to compensate completely for the reduction in intensity. The Moon and Folkins (1991) data suggest that both aerodynamics and acoustic-perceptual accuracy are possible control variables in speech, suggesting that the two theories presented above could be related in some way.

The current study was a perturbation study, one in which normal speech production is disturbed. Studies that perturb the oral cavity, both statically and dynamically, have been used extensively to examine compensation (cf., Folkins and Abbs, 1975; Kelso *et al.*, 1984; Lindblom *et al.*, 1979; McFarland and Baum, 1995; McFarland *et al.*, 1996; Munhall *et al.*, 1994; Putnam *et al.*, 1986). Compensation, in speech production, reflects the mechanisms used to overcome changes or disruptions to speech. One example of an everyday compensation is when speakers adjust their articulation while holding something between their teeth, like a pencil. In the current study, allowing air pressure to leak out of the oral cavity during the production of a [p] changes the aerodynamic environment and the acoustic and perceptual accuracy of the production. Since perturbation leads to active responses to maintain control variables, at the expense of other variables, the study of compensation provides information regarding what variables are controlled (McFarland *et al.*, 1996).

Huber and Stathopoulos (2003) demonstrated responses from the respiratory and laryngeal subsystems to an oral air pressure bleed. When air pressure was allowed to leak out of the oral cavity during production of the syllable [pɑ], lung and rib cage volume excursions significantly increased and rib cage volume terminations significantly decreased, showing the respiratory subsystem compensated for the air pressure bleed. Additionally, there was a decrease in average airflow when the bleed was in place, strongly suggesting an increase in laryngeal airway resistance due to an active response from the laryngeal subsystem. However, these data did not explain the goal of the compensations from the respiratory and laryngeal subsystems. The present study was completed in concert with the production study described in Huber and Stathopoulos (2003).

The purpose of the present study was to examine possible acoustic-perceptual and aerodynamic control variables in speech production by exploring speakers' acoustic and aerodynamic compensations to perturbation (an air pressure bleed—''tube with bleed'' condition). Further, listener's perceptions of the productions under the normal and perturbing conditions were examined. The first hypothesis of this study, related to Warren and colleagues theory, is that if pressure or resistance is controlled in speech production, air pressure and/or resistance will be maintained, at least to some extent, in the ''tube with bleed'' condition. Huber and Stathopoulos (2003) demonstrated an increase in respiratory drive and possibly laryngeal airway resistance in response to the air pressure bleed. These mechanisms may have resulted in the maintenance of air pressure and/or resistance at a level comparable to the NT condition. Since [p] is a stop consonant, measurements of oral pressure and vocal tract resistance were used to examine this hypothesis.

The second hypothesis of this study, related to the Perkell and colleagues theory, is that if acoustics and perceptual accuracy are controlled in speech production, measures of acoustic and perceptual accuracy will be maintained, at least to some extent, in the ''tube with bleed'' condition. The silent closure interval and burst, consistently present in normal voiceless stops, relate to the articulatory dynamics of stop consonant production, and contribute to our perception (cf., Bailey and Summerfield, 1980; Halle *et al.*, 1957; Liberman *et al.*, 1952; Repp, 1984). Therefore, measurements of oral acoustics (the presence of a burst during the production of [p] and addition of supraglottal noise during the normally silent [p] closure interval) were used, along with more global acoustic measurements (sound pressure level and fundamental frequency) and a perceptual study, to examine this hypothesis.

II. METHODS

A. Production study

The methods and some of the data reported in this study in the production study sections have been published in Huber and Stathopoulos (2003). However, additional production data, specifically relevant to the current study, are also presented.

1. Speaker participants

Twenty normal speakers, ten women and ten men, were participants. The average age of the women was 23 years, 8 months (range=20 years, 6 months to 27 years, 3 months; standard deviation (SD)=1 year, 10 months) and the average age of the men was 24 years, 8 months (range=20 years, 3 months to 29 years, 6 months; SD=2 years, 10 months). Speakers were grouped by sex because resistance of the vocal tract may be different in women and men, and, therefore, production of absolute pressures and the viscoelastic properties of the laryngeal and respiratory components may be different (Stathopoulos and Sapienza, 1997). For example, women demonstrate less lung elasticity than men (cf. Bode *et al.*, 1976; Mansell *et al.*, 1977). It has been suggested that

this difference may account for some of the differences in oral airflow between women and men during plosive production (Stathopoulos and Weismer, 1985).

Participants reported no history of voice or respiratory problems (including asthma), neurological disease, head or neck surgery, or formal speaking or singing training; no recent colds or infections; and no smoking in the last 5 years. They had normal height and weight as measured on the day of testing (Expert Panel on the Identification, 1998), normal speech, language, and voice as judged by a certified speech-language pathologist (the first author), and General North American dialect of English as judged by the first author. Participants had normal hearing as indicated by a hearing screening at 20 dB HL for octave frequencies between 250 and 8000 Hz, bilaterally (ANSI, 1996).

2. Procedures and speech tasks

A perturbation paradigm was used that included an intraoral air pressure bleed during the production of the voiceless plosive [p]. Speakers were instructed to produce a train of seven syllables of [pɑ] on one breath at comfortable loudness and pitch, using a slow rate (approximately 1 to $1\frac{1}{2}$ syllables per second). The syllable train was produced ten times under the each of the following conditions:

(1) ''tube no bleed'' (TNB)—A 5-mm internal diameter (8-mm external diameter, 19.63-mm^2 cross-sectional area) bleed tube was placed between the lips on the right side of the mouth, with one end extending into the oral cavity. The end of the tube inside the mouth was sealed so that no bleed was present in the oral cavity.
(2) ''tube with bleed'' (TWB)—A bleed tube, of the same type as the one used for the TNB condition, was placed between the lips on the right side of the mouth, with one end extending into the oral cavity; however, both ends of the tube were open, allowing air to leak out during the production of [p]. The tube remained open throughout the ten syllable train productions.
(3) ''no tube'' (NT)—normal production without a bleed tube.

The NT condition was produced first to ensure that the speakers understood how to complete the task. The order of the TNB and TWB conditions was counterbalanced across subjects, and speakers were not informed as to whether the bleed tube would be open or not. Speakers also produced syllable trains composed of [bɑ], [fɑ], and [sɑ] without a bleed tube as foils for the perceptual study. They produced three trains of seven syllables for each consonant–vowel combination at comfortable intensity.

3. Equipment

The acoustic signal was sensed with a dynamic, omnidirectional, lavalier microphone. The microphone was placed within a circumferentially vented pneumotachograph mask. The mouth-to-microphone distance was held constant at 3.5 cm. Resonances from within the mask were not a concern as all conditions included the mask. Therefore, any added resonances associated with the mask were the same across all conditions. The signal was stored on digital audio tape, using a two-channel DAT Recorder (Tascam model DA-P2).

Oral airflow was sensed using the circumferentially vented pneumotachograph mask and a pressure transducer (Glottal Enterprises, PTW-1). Oral pressure was sensed via a 1-mm internal diameter (2-mm external diameter) pitot tube placed between the lips just inside the participant's mouth on the left side. The distal end of the pitot tube was connected to a pressure transducer. Oral airflow and air pressure were only collected during production of [pɑ] syllable trains, and not during the production of foils.

The oral airflow and air pressure signals were digitized on-line through an IBM compatible computer and an A-D/D-A board (Data Translations DT2821) at 10 kHz. The oral airflow signal was low-pass filtered at 3400 Hz and the oral pressure signal was low-pass filtered at 36.5 Hz as the signals passed through the Glottal Enterprises system. The microphone signal was digitized separately from the digital audiotape to the computer after data collection was complete. It was low-pass filtered at 8 kHz for anti-aliasing (Frequency Devices 900C) and digitized to the IBM compatible computer through the same A-D/D-A board at a sampling rate of 20 kHz.

4. Measurements

Acoustic measurements were made to describe the signal characteristics of the [pɑ] syllables produced with and without the bleed in place.

(1) Sound pressure level (SPL) was measured from the microphone within the mask, calibrated for a 3.5-cm mouth-to-microphone distance. Sound pressure level was measured across the syllable train.
(2) Fundamental frequency (F0) was measured from the middle 100 ms of the center five [ɑ] productions in each syllable train, using the inverse-filtered airflow signal. Fundamental frequency was measured as a gross indicator of changes in voicing or pitch.
(3) Presence of a burst during production of [p]: Spectrograms of the acoustic waveforms, computed with a bandwidth of 257 Hz, were examined to identify a burst during the production of [p] for each of the conditions. A burst was indicated by a discrete darkened band of noise, of any amplitude, during the production of [p] which was confirmed by a spike in energy in the acoustic waveform (see Fig. 1). Data are presented as the percent of time a burst was present out of the total number of productions for each condition.
(4) Addition of noise during [p] closure interval: The same spectrograms as above were examined to identify noise during the production of [p] for each of the conditions, indicated by a band of noise across the usually quiet closure interval for the production of [p] (see Fig. 1). A production was counted as containing noise during the closure interval when a substantial amount of noise was visible on the spectrogram at any time during the closure interval; the noise did not have to fill the closure interval (see Fig. 1(d)). However, a small amount of noise at the start and/or end of the closure interval (see Fig. 1(b))

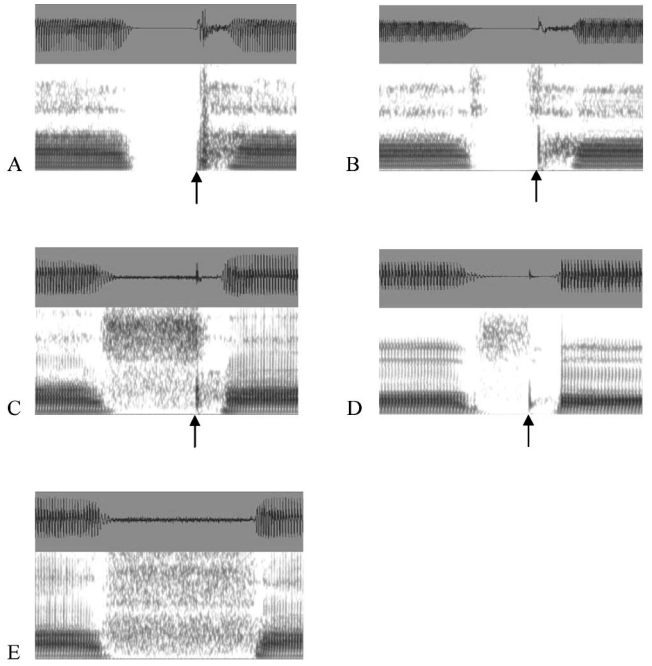


FIG. 1. Examples of the presence of a burst and addition of noise during the closure interval in the production of [p]. Arrows indicate burst location. (a) ''No tube'' condition—burst is present, noise is not. (b) ''Tube no bleed'' condition—burst is present, noise is not. (c and d) ''Tube with bleed condition''—burst and noise are present. (e) ''Tube with bleed'' condition—noise is present, burst is not.

was not enough to be counted as a noisy closure interval. The purpose of this measurement was to determine if noise was introduced during the normally silent closure interval due to the introduction of the bleed or the bleed tube. Data are presented as the percent of time noise was present out of the total number of productions for each condition.

Oral aerodynamic measurements were made to evaluate whether pressure and vocal tract resistance were maintained during the bleed condition and whether speech was perturbed by the TWB condition.

(1) Peak intraoral air pressure (P_o) was measured at the peak pressure for the [p] in the five middle syllables of the syllable train.
(2) Amount of airflow leak (V_{leak}) was measured from the average airflow signal at the point where the intraoral air pressure was highest for the [p] consonant. This measurement indicated two things. If the participant performed the task appropriately in the NT and TNB conditions, this value was close to 0, indicating no air was leaking out of the oral cavity. During the TWB condition, it suggested the effectiveness of the bleed condition in perturbing the oral cavity.
(3) In the NT and TNB conditions, vocal tract resistance was estimated using laryngeal airway resistance (R_{law}). Since P_o can be used to estimate subglottal pressure (P_s) and average oral airflow (V_o) can be used to estimate translaryngeal airflow (V_{tl}), R_{law} was computed by dividing the estimated V_{tl} during the [a] by the estimated P_s during the [p] (Smitheran and Hixon, 1981). Average oral airflow was measured from the airflow signal of the five middle vowels in the syllable train. The oral airflow signal was low-pass filtered at 30 Hz and V_o was measured from the center of the vowel. During the TWB condition, R_{law} could not be estimated because P_o could not be used as an estimate of P_s since the oral cavity was not closed. Therefore, vocal tract resistance was estimated by taking oral port resistance, a common technique for estimating resistance during a fricative (Warren, 1986). Oral port resistance was estimated by dividing the peak P_o during the [p] production by the airflow leak at the lips during the production of [p] (V_{leak}). The comparison of oral port resistance to laryngeal airway resistance was based on work by Warren (1986).

5. Reliability

Intrameasurer reliability was completed for all measurements on 10% of the data, one male and one female who were randomly selected. For all measurements except ''presence of a burst'' and ''addition of noise,'' the difference between the measurements was computed by subtracting the mean of the second set of measurements for each dependent measure from the mean of the first set. Also, Pearson product moment correlations were performed between the first and second set of measurements. Correlations above 0.9 were expected between the two sets of measurements for all dependent measures. For the ''presence of a burst'' and ''addition of noise'' ratings, intrameasurer reliability was determined by having the measurer rerate the syllables for these characteristics, without knowledge of the original rating. The number of times the first rating of each production agreed with the second rating was determined.

Results of the intrameasurer reliability indicated that there were small differences between the two sets of measurements (less than 1 unit for each set). All correlations between the first and second set of measurements were above 0.95, indicating excellent reliability of measurement. For the ''presence of a burst'' ratings, the first set of ratings agreed with the second set 96% of the time. For the ''addition of noise'' ratings, the first set of ratings agreed with the second set 100% of the time.

6. Statistical analysis

Participant means were calculated for each trial. Since the V_{leak} values represented the adequacy of the productions in the NT and TNB conditions, if the V_{leak} value was higher than 0.15 L/s during these conditions, it was determined that the subject did not complete the task appropriately for that syllable production. If V_{leak} values were higher than 0.15 L/s for a syllable in the NT or TNB conditions, all measurements made on that syllable were removed from the trial mean. This occurred in less than 2% of the trials.

The trial values were averaged together for each individual to obtain means and standard deviations that reflected the individual's performance in each condition. Comparisons among conditions were used to obtain information about compensation to each condition. Multivariate analyses of

TABLE I. Summary table for statistical analyses for condition, sex, and sex by condition effects. Significant effects ($p<0.05$) are indicated with an asterisk. $F=F$ ratio. Degrees of freedom are in parentheses.

	Condition (2)		Sex (1)		Sex X condition (2)	
Measure	*F*	*p*	*F*	*p*	*F*	*p*
Sound pressure level	1.568	0.237	1.822	0.194	0.280	0.759
Fundamental frequency	1.710	0.211	215.862	0.000*	2.691	0.097
Presence of burst	29.509	0.000*	1.949	0.180	2.702	0.096
Addition of noise	546.916	0.000*	0.003	0.957	2.195	0.142
Peak intraoral air pressure	37.635	0.000*	0.273	0.608	3.262	0.063
Airflow leak	3.816	0.043*	0.788	0.386	0.656	0.531
Airway resistance	32.777	0.000*	0.076	0.786	3.224	0.065
Percent correct—Identification task	95.991	0.000*	N/A	N/A	N/A	N/A

variance (MANOVA) were used to examine differences between the conditions on the dependent variables. The conditions were used as multiple dependent variables. Sex effects were assessed in repeated measures analyses of variance (ANOVA), with sex as the between subject factor. Tukey A tests were used *posthoc* for any significant condition effects in the MANOVAs. An alpha level of $p<0.05$ was used for all tests to determine significance. A summary of the statistical results from the MANOVAs and ANOVAs is given in Table I.

B. Perceptual study

1. Listener participants

A total of ten normal adult listeners participated in the perceptual study. Average age of the listeners was 30 years, 9 months (range=23 years, 9 months to 52 years, 11 months; SD=9 years, 4 months). Listeners had normal hearing, indicated on the day of testing by a hearing screening at 20 dB for octave frequencies between 250 and 8000 Hz, bilaterally (ANSI, 1996). Only individuals who reported no formal education regarding perception of speech sounds, including an introductory phonetics course, and no formal singing or speech training were included in the study. Näive rather than trained listeners were chosen to get a more typical opinion regarding the identity of the consonant across conditions. Listeners had normal speech and language as judged by a certified speech-language pathologist (the first author). Their first language was American English, and they were right-handed. None of the listeners knew about or participated in the production study.

2. Equipment

The second, single syllable was isolated from each syllable train. The waveforms were played through an IBM compatible computer and an A-D/D-A Board (Data Translations DT 2821) at 10 kHz. The signal was low-pass filtered at 4200 Hz for anti-aliasing (White Instruments 4681 Filter).

a. Procedures. The speech signal was presented at about 77 dB SPL (±2 dB), depending on the original SPL of the production, measured with B&K SPL meter coupled to the headphones. Participants were tested individually in a sound-treated room and heard the speech samples via high-quality headphones in the right ear. The speech waveforms were placed into a randomly determined order, not blocked by talker, and presented to each participant. Each participant received the same order of presentation. The identification task was designed to provide information about whether the addition of the tube or the bleed significantly changed the perceptual identity of the [p] sound.

3. Procedures and stimuli

a. Practice. All listeners participated in practice trials to complete a criterion test (to ensure they could complete the listening task) and to become familiar with the procedures. Stimuli included [pɑ] syllables from all three conditions and foil syllables from the [bɑ], [fɑ], and [sɑ] trains. The stimuli for the practice trials were not included in the experimental test session. First, participants were presented with five [pɑ] stimuli from the NT condition, and two [bɑ], two [fɑ], and one [sɑ] stimuli. Participants had to correctly identify eight of the sounds as ''pa'' or ''not pa'' consecutively to continue in the study. All participants were able to do so without requiring additional stimuli to be presented. Participants were also given practice listening to stimuli like those on the test. Listeners were presented with three [pɑ] stimuli from each condition and six foils (two [bɑ], two [fɑ], and two [sɑ]).

Listeners indicated their response by pushing a button labeled ''pa'' on a response box if they felt the syllable was a [pɑ] and a button labeled ''not pa'' on a response box if they felt the syllable was not a [pɑ]. No feedback regarding correctness was provided. Participants could listen to the syllable as many times as they wished by pressing the ''r'' button on the computer keyboard. The program waited for a response, followed by a 1-s pause, before presenting the next signal.

b. Experimental trials. Listeners were presented with single [pɑ] syllables from each of the conditions and single syllables of the foil stimuli ([bɑ], [fɑ], and [sɑ]). A set of [pɑ] productions was taken from each speaker, one stimulus from the NT condition, two stimuli from the TNB condition, and two stimuli from the TWB condition. For the NT condition, the stimulus was taken from the second syllable of the last trial. For the tube stimuli (TNB and TWB), one stimulus was taken from the second syllable of the first trial and one from the second syllable of the last trial. Six [bɑ], seven [fɑ], and seven [sɑ] stimuli were randomly chosen, one stimulus from each subject. Each stimulus was presented twice to establish intrarater reliability, for a total of 240 stimuli ($1+2+2=5$ productions$\times 20$ speakers$=100+20$ foils$=120\times 2$ presentations$=240$). The procedures for presentation and response were the same as described for the practice stimuli.

4. Statistics

Participant means and standard deviations were computed for the percent correct scores for each condition. To establish intrarater reliability, *t*-tests were computed between the two ratings of the presentations of the stimuli. There was no significant difference in the responses from listeners the first and the second time they rated a stimulus [$t(39)$

$=0.265$, $p=0.79$]. The two sets of responses were strongly correlated [$r=0.98$]. These statistics indicate good intrarater reliability. There were no differences in any of the conditions between ratings of the first and last trials of the ''tube'' conditions [TNB: $t(9)=0.845$, $p=0.420$ and TWB: $t(9)=-1$, $p=0.343$]. Due the lack of statistical difference between the first and last trial productions, they were treated as trials of the same condition and averaged together.

Differences in perception across conditions were determined by computing a MANOVA with the conditions as multiple dependent variables. Tukey A tests were used *post-hoc* for any significant effects in the MANOVAs. An alpha level of $p<0.05$ was used to determine significance. A summary of the statistical results from the MANOVA is given in Table I.

III. RESULTS

The average V_{leak} data demonstrate that the bleed did perturb the oral cavity. Average V_{leak} was significantly higher during the TWB condition (0.596 L/s) than during the other two conditions. The V_{leak} during the NT and the TNB conditions was similar (NT=0.018 L/s, TNB=0.022 L/s). There was no significant difference between women and men in V_{leak} and no sex by condition interaction.

The first hypothesis of this study was if pressure or resistance is controlled in speech production, oral air pressure or resistance will be maintained, at least to some extent, in the TWB condition. Mean P_o during the TWB condition (3.93 cm H_2O) was significantly lower than in the NT and TNB conditions (NT=5.74 cm H_2O, TNB=5.66 cm H_2O), a decline of almost 2 cm H_2O in the TWB condition. Also, vocal tract resistance was significantly lower in the TWB condition (15.61 cm H_2O/L/s) as compared to the NT and TNB conditions (NT=40.52, TNB=39.36 cm H_2O/L/s). Neither P_o nor vocal tract resistance was significantly different between the NT and TNB conditions. There were no significant sex or sex by condition interactions for either pressure or resistance.

The second hypothesis of this study was if acoustics and perceptual accuracy are controlled in speech production, measures of acoustic and perceptual accuracy will be maintained, at least to some extent, in the TWB condition. Looking first at oral acoustic measurements, a burst was present on significantly fewer productions in the TWB condition (57.1%) as compared to the NT (99.6%) and TNB (98.1%) conditions. Noise was present during the closure interval on significantly more productions during the TWB condition (93.1%) as compared to the NT (0.8%) and the TNB (5.5%) conditions. There were no significant differences between the NT and TNB conditions for either measurement. There were no significant sex or sex by condition effects for either measurement.

For the more global acoustic measurements, there were no significant differences in SPL (NT=99.8 dB, TNB=100.1 dB, TWB=99.9 dB) or F0 (NT=174.5 Hz, TNB=171.6 Hz, TWB=172.2 Hz) across the three conditions. The expected main effect for sex was present for F0, with women having a higher average F0 than men (women=224.1 Hz, men=121.5 Hz). There was no difference in SPL between women and men and no sex by condition interaction for either SPL or F0.

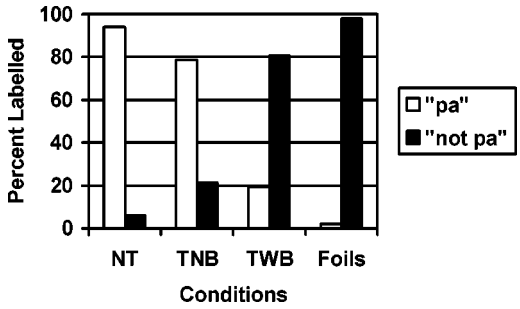


FIG. 2. Percent of productions correctly labeled as ''pa'' and ''not pa'' during the perceptual study.

For perceptual measurements, there was a significant difference among the three conditions in percent correctly identified as ''pa.'' On average, 94% of the stimuli in the NT condition were correctly identified. In the TNB condition, a significantly smaller percent were correctly identified, 78.5%. The percent correctly identified in the TWB condition (19.25%) was significantly lower than both the NT and the TNB conditions. Figure 2 illustrates this effect. In the NT condition, [pɑ] target stimuli were almost always labeled as ''pa.'' In the TNB condition [pɑ] target stimuli were also more often labeled as ''pa.'' However, in the TWB condition, the [pɑ] target stimuli were more often labeled as ''not pa'' (80.75%). The foils (''sa,'' ''fa,'' and ''ba'') were almost always labeled as ''not pa'' (98%). Thus, the presence of an open bleed tube caused productions targeted by the speaker to be [pɑ] to be perceived significantly more often as sounds other than [pɑ]. Qualitatively, the listeners rated male and female speakers equally well.

IV. DISCUSSION

The first hypothesis of this study was if pressure or resistance is controlled in speech production, oral air pressure or resistance will be maintained, at least to some extent, in the TWB condition. The data from the current study do not support the hypothesis that *vocal tract resistance* is controlled during speech. If resistance had been controlled, the oral port resistance during the TWB condition would have been equivalent or closer to the R_{law} during the TNB and NT conditions (Warren, 1986). To maintain vocal tract resistance, speakers could have increased pressure by increasing respiratory drive more than they did, or by reducing airflow during the [p], possibly by narrowing the glottis. However, there is no evidence that either of these mechanisms were used by speakers in the TWB condition to control resistance.

The P_o data suggest that speakers were not trying to maintain P_o at a level comparable to the NT condition. Speakers completed ten trials of the syllable train in the TWB condition which took approximately 7 to 10 min, allowing for rest breathing between each trial. If the primary goal was to maintain P_o, it would be expected that P_o would increase, reaching a level more comparable to the NT level by the end of the trials. Subjects could have achieved higher P_o by utilizing a larger increase in respiratory drive than they did in the current study (Huber and Stathopoulos, 2003).

TABLE II. Summary of differences in oral pressure (P_o), lung volume excursion (LVE), rib cage volume termination (RCVT), and rib cage excursion (RCE) between the first and last trials in the "tube with bleed" condition. (%VC=percent vital capacity; %RCC=percent rib cage capacity). Data taken from Huber and Stathopoulos (2003).

Measure	Mean difference between trials	*t* value (df=38)	*p* value
P_o	−0.17 cm H_2O	−0.412	0.6825
LVE	−0.87 %VC	−0.107	0.9152
RCT	−0.30 %RCC	−0.096	0.9243
RCE	−3.89 %RCC	−1.114	0.2721

However, there was no significant change across the ten trials in either P_o or respiratory kinematic measurements (see Table II). Since P_o and respiratory drive did not increase across the trials, these data suggest that the primary goal was not the maintenance of P_o at a level similar to the NT level.

On the other hand, on average, P_o was indeed maintained above 3 cm H_2O in the TWB condition, although that is still significantly lower than the P_o in the NT and TNB conditions. Maintenance of a P_o of 3 cm H_2O has been posited by Warren as the minimum adequate P_o during the production of [p] (i.e., Warren *et al.*, 1989, 1990) and has been used in previous air pressure bleed studies as a criterion level demonstrating successful compensation for a bleed (i.e., Putnam *et al.*, 1986). The simplest answer to why the P_o only fell to 3 cm H_2O on average in the TWB condition is that the bleed tube limited how much P_o was lost and that a larger tube would have resulted in a larger P_o loss. However, previous studies have used larger tubes and not seen lower P_o values. Putnam *et al.* (1986) used bleed sizes up to 28.3 mm^2 and found that subjects still maintained P_o above 3 cm H_2O on average. Therefore, it is not likely that the bleed tube size is the reason for the P_o findings in this study. It is more likely that the maintenance of P_o at 3 cm H_2O was a result of the speakers' active compensations. It is clear from these data and the data reported in Huber and Stathopoulos (2003) that the speakers were trying to compensate for the bleed. Examination of the speakers' compensations provides some support for the hypothesis that pressure is a controlled variable and some support for the second hypothesis, that acoustic and perceptual accuracy are controlled variables in speech production.

If Warren is correct about 3 cm H_2O being the minimum adequate P_o for [p] production, speakers may have been compensating in order to maintain P_o at 3 cm H_2O. In fact, 77% of the time, P_o was maintained above 3 cm H_2O in the TWB condition. However, five women and three men had P_o below 3 cm H_2O for some or all of their trials in the TWB condition. These data suggest that some speakers may have controlled P_o, supporting the hypothesis that pressure is a controlled variable in speech production.

Dalston *et al.* (1988) suggest a P_o of 3 cm H_2O is required for production of a burst. Production of a burst may have been important to the speakers given that the acoustic data from the NT condition shows bursts were present in normally produced [p] consonants in this syllable train task in 996/1000 productions. If the purpose of the P_o maintenance was to ensure that a burst would be present, this would support the second hypothesis as well, that speakers were compensating to maintain acoustic-perceptual accuracy for the [p].

Another important consideration for why P_o was maintained at 3 cm H_2O is to maintain voicing. A pressure of 3 cm H_2O has been posited to be the minimum threshold P_s required to maintain vocal fold vibration (Titze, 1991). The P_o during the TWB condition was dependent on the amount of subglottal pressure (P_s) being generated since P_s and P_o equalize during voiceless stop production (Netsell, 1969). Since at least 3 cm H_2O of P_s is required for voicing for the vowel, an amount close to that would be likely to be present both subglottally and in the oral cavity during voiceless plosive production, assuming that P_s was held relatively constant across the utterance (Ladefoged, 1967). These results suggest that P_o was not specifically controlled, but perhaps P_s was, still supporting the hypothesis that pressure is a controlled variable in speech production. If P_s was controlled to ensure voicing for the [ɑ], it suggests speakers were compensating in order to maintain an important acoustic-perceptual characteristic of vowels as well.

Additionally, Perkell and colleagues state that maintenance of prosodically correct loudness and pitch are important to speakers in controlling perceptual adequacy (Perkell *et al.*, 1997). SPL and F0 did not change significantly across conditions. Not only did F0 remain constant on average across the conditions, it did not differ more than 25 Hz across syllables within a trial for any subject, and the average difference across syllables did not differ among the conditions (NT=3.60 Hz, TNB=4.76 Hz, TWB=4.67 Hz). The data from Huber and Stathopoulos (2003) demonstrated much larger changes in the speech production patterns in the respiratory subsystem as compared to the laryngeal subsystem in the TWB condition. This may be because large changes in laryngeal vibratory patterns would have altered F0. The SPL and F0 data suggest that, in compensating for the bleed, subjects made active efforts to control perceptual accuracy of the vowel, by preserving a constant SPL and F0.

There are clear acoustic-perceptual consequences of maintaining the pressure as the speakers did, suggesting that pressure maintenance may have been one way speakers maintained acoustic-perceptual accuracy. Speakers were not, however, able to completely compensate for the noise during the closure interval in the TWB condition since they could not close their lips around the bleed tube, to shut off the bleed. The lack of a burst, coupled with the noise during the closure interval, may have led listeners to perceive a continuant consonant, like [s] or [f]. This is substantiated by the perceptual data since the TWB productions were more often labeled as "not pa" by the listeners, in a manner similar to the foils (see Fig. 2). Listeners commented that they were unsure whether to call some of the sounds "pa" or "not pa" since they started out noisy, but had a "pa" like sound in them, almost like a "s-pa."

Oral pressure may have contributed to the addition of noise during the closure interval for [p]. In fact, most productions with noise present in the TWB condition also had a P_o greater than 3 cm H_2O (80%). It may be that the subjects who allowed P_o to fall below 3 cm H_2O did so to improve

perceptual accuracy, specifically to reduce the noise during the closure interval. Of the five speakers who had a TWB production labeled correctly, as ''pa,'' by at least half of the listeners, three had a P_o of less than 3 cm H_2O.

The SPL and F0 data suggest that while speakers could not fully compensate for the bleed in production of the [p], they did compensate in order to maintain acoustic and perceptual accuracy of the vowel. This difference in compensating in production of consonants versus production of vowels has been reported previously. McFarland and Baum (1995) demonstrated that consonants require more compensation time due to the precise nature of their articulation. Subjects may have learned to compensate more for the [p] if given more time to practice with the bleed in place. For four of the five speakers who had a TWB production labeled correctly, as ''pa,'' by at least half of the listeners, the production which was labeled correctly was their last production, rather than their first.

V. SUMMARY

The data from the current study support the idea that there are multiple control variables in speech production, and suggest that there may be layers of control where one variable may be controlled in order to achieve another. These levels of control are not new to our understanding of motor control. An analogous example from the limb system is that the strength of a movement is controlled in order to achieve the ultimate goal of accurately and smoothly picking up an object. It is clear that pressure was controlled as is suggested by Warren and colleagues. Oral pressure was maintained by over half of the speakers above 3 cm H_2O. However, the purpose of pressure maintenance appears to be in order to control acoustic-perceptual accuracy, as suggested by Perkell and colleagues. For example, P_o may have been maintained above 3 cm H_2O in order to ensure a burst was produced during the [p]. Subglottal pressure may have been maintained to ensure voicing for the vowel or to preserve prosodic contours by maintaining SPL and F0. Therefore, these data suggest that controlling pressure may be one way speakers meet the ultimate goal of acoustic and perceptual accuracy.

ACKNOWLEDGMENTS

We would like to thank Kris Tjaden, Jeffrey Higginbotham, Anthony Duggan, and Angela M. Haller for their assistance with this study.

ANSI. (**1996**). *Specifications for audiometers* (Vol. ANSI S3.6-1996) (ANSI, New York).

Bailey, P. J., and Summerfield, Q. (**1980**). ''Information in speech: observations on the perception of [s]-stop clusters,'' J. Exp. Psychol. Hum. Percept. Perform. **6**, 536–563.

Bode, F. R., Dosman, J., Martin, R. R., Ghezzo, H., and Macklem, P. T. (**1976**). ''Age and sex differences in lung elasticity, and in closing capacity in nonsmokers,'' J. Appl. Physiol. **41**, 129–135.

Dalston, R. M., Warren, D. W., Morr, K. E., and Smith, L. R. (**1988**). ''Intraoral pressure and its relationship to velopharyngeal inadequacy,'' Cleft Palate J. **25**, 210–219.

Expert Panel on the Identification, E., and Treatment of Overweight and Obesity in Adults. (**1998**). ''Executive summary of the clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults,'' Arch. Int. Med. **158**, 1855–1867.

Folkins, J. W., and Abbs, J. H. (**1975**). ''Lip and jaw motor control during speech: Responses to resistive loading of the jaw,'' J. Speech Hear. Res. **18**, 207–220.

Guenther, F. H. (**1995**). ''Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production,'' Psychol. Rev. **102**, 594–621.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). ''A theoretical investigation of reference frames for the planning of speech movements,'' Psychol. Rev. **105**, 611–633.

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., and Perkell, J. S. (**1999**). ''Articulatory tradeoffs reduce acoustic variability during American English /r/ production,'' J. Acoust. Soc. Am. **105**, 2854–2865.

Halle, M., Hughes, G. W., and Radley, J.-P. A. (**1957**). ''Acoustic properties of stop consonants,'' J. Acoust. Soc. Am. **29**, 107–116.

Huber, J. E., and Stathopoulos, E. T. (**2003**). ''Respiratory and laryngeal responses to an oral air pressure bleed during speech,'' J. Speech Lang. Hear. Res. **46**, 1207–1220.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. A. (**1984**). ''Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures,'' J. Exp. Psychol. Hum. Percept. Perform. **10**, 812–832.

Ladefoged, P. (**1967**). ''Stress and respiratory activity,'' in *Three Areas of Experimental Phonetics* (Oxford U.P., London), pp. 1–49.

Liberman, A. M., Delattre, P. C., and Cooper, F. S. (**1952**). ''The role of selected stimulus-variables in the perception of the unvoiced stop consonants,'' Am. J. Psychol. **65**, 497–516.

Lindblom, B., Lubker, J., and Gay, T. (**1979**). ''Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation,'' J. Phonetics **7**, 147–161.

Mansell, A. L., Bryan, A. C., and Levison, H. (**1977**). ''Relationship of lung recoil to lung volume and maximum expiratory flow in normal children,'' J. Appl. Physiol.: Respir., Environ. Exercise Physiol. **42**, 817–823.

McFarland, D. H., and Baum, S. R. (**1995**). ''Incomplete compensation to articulatory perturbation,'' J. Acoust. Soc. Am. **97**, 1865–1873.

McFarland, D. H., Baum, S. R., and Chabot, C. (**1996**). ''Speech compensation to structural modifications of the oral cavity,'' J. Acoust. Soc. Am. **100**, 1093–1104.

Moon, J. B., and Folkins, J. W. (**1991**). ''The effects of auditory feedback on the regulation of intraoral air pressure during speech,'' J. Acoust. Soc. Am. **90**, 2992–2999.

Munhall, K. G., Lofqvist, A., and Kelso, J. A. S. (**1994**). ''Lip-larynx coordination in speech: Effects of mechanical perturbations to the lower lip,'' J. Acoust. Soc. Am. **95**, 3605–3616.

Netsell, R. (**1969**). ''Subglottal and intraoral air pressures during the intervocalic contrast of /t/ and /d/,'' Phonetica **20**, 68–73.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., and Jordan, M. I. (**1995**). ''Goal-based speech motor control: A theoretical framework and some preliminary data,'' J. Phonetics **23**, 23–35.

Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J. and Guiod, P. (**1997**). ''Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models,'' Speech Commun. **22**, 227–250.

Putnam, A. H. B., Shelton, R. L., and Kastner, C. U. (**1986**). ''Intraoral air pressure and oral air flow under different bleed and bite-block conditions,'' J. Speech Hear. Res. **29**, 37–49.

Repp, B. H. (**1984**). ''The role of release bursts in the perception of [s]-stop clusters,'' J. Acoust. Soc. Am. **75**, 1219–1230.

Smitheran, J. R., and Hixon, T. J. (**1981**). ''A clinical method for estimating laryngeal airway resistance during vowel production,'' J. Speech Hear Disord. **46**, 138–146.

Stathopoulos, E. T., and Sapienza, C. M. (**1997**). ''Developmental changes in laryngeal and respiratory function with variations in sound pressure level,'' J. Speech Lang. Hear. Res. **40**, 595–614.

Stathopoulos, E. T., and Weismer, G. (**1985**). ''Oral airflow and air pressure during speech production: A comparative study of children, youths, and adults,'' Folia Phoniatr. **37**, 152–159.

Stein, R. B. (**1982**). ''What muscle variable(s) does the nervous system control in limb movements?'' Behav. Brain Sci. **5**, 535–577.

Titze, I. R. (**1991**). ''Phonation threshold pressure: A missing link in glottal

aerodynamics,'' NCVS Status and Progress Report **1**, 1–14.
Warren, D. W. (**1986**). ''Compensatory speech behaviors in individuals with cleft palate: A regulation/control phenomenon?'' Cleft Palate J. **23**, 251–260.
Warren, D. W., Dalston, R. M., and Dalston, E. T. (**1990**). ''Maintaining speech pressures in the presence of velopharyngeal impairment,'' Cleft Palate J. **27**, 53–58.
Warren, D. W., Hall, D. J., and Davis, J. (**1981**). ''Oral port constriction and pressure-airflow relationships during sibilant productions,'' Folia Phoniatr. **33**, 380–394.
Warren, D. W., Dalston, R. M., Morr, K. E., Hairfield, W. M., and Smith, L. R. (**1989**). ''The speech regulating system: Temporal and aerodynamic responses to velopharyngeal inadequacy,'' J. Speech Hear. Res. **32**, 566–575.

A neural network model of the articulatory-acoustic forward mapping trained on recordings of articulatory parameters

Christopher T. Kello[a)]
Department of Psychology, George Mason University, Fairfax, Virginia 22030

David C. Plaut
Department of Psychology, Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

(Received 2 October 2003; revised 24 February 2004; accepted 1 March 2004)

Three neural network models were trained on the forward mapping from articulatory positions to acoustic outputs for a single speaker of the Edinburgh multi-channel articulatory speech database. The model parameters (i.e., connection weights) were learned via the backpropagation of error signals generated by the difference between acoustic outputs of the models, and their acoustic targets. Efficacy of the trained models was assessed by subjecting the models' acoustic outputs to speech intelligibility tests. The results of these tests showed that enough phonetic information was captured by the models to support rates of word identification as high as 84%, approaching an identification rate of 92% for the actual target stimuli. These forward models could serve as one component of a data-driven articulatory synthesizer. The models also provide the first step toward building a model of spoken word acquisition and phonological development trained on real speech.
 [DOI: 10.1121/1.1715112]

PACS numbers: 43.70.Bk, 43.72.Ja, 43.70.Ep, 43.70.Jt [AL] Pages: 2354–2364

I. INTRODUCTION

A necessary component of any complete model of speech acquisition or speech production is the physical relationship between the shape of the vocal tract, and the acoustic energy emitted from the vocal tract. This relationship is often referred to as the *forward* mapping from articulatory states to acoustic outputs, whereas the *inverse* mapping would recover articulatory states from the speech signal (Jordan and Rumelhart, 1992). The forward mapping is integral to speech production because the primary proximal stimulus used by the listener is the acoustic speech signal. Therefore, to produce comprehensible speech, the talker must somehow take into account the forward mapping from articulatory commands to acoustic outputs.

The articulatory-acoustic mapping has been studied primarily for two purposes. One is to better understand how speech is perceived and produced by humans (e.g., Rubin, Baer, and Mermelstein, 1981), and the other is to develop articulatory-based techniques for automatic speech recognition (e.g., Blackburn and Young, 2000a) and speech synthesis (e.g., Greenwood, Goodyear, and Martin, 1992). In the service of these purposes, computational models have been developed to simulate the forward mapping from articulation to acoustics (e.g., Baer *et al.*, 1991; Beautemps, Badin, and Laboissiere, 1995). These forward models have been based upon articulatory and acoustic dimensions that are known to convey phonetic information, and upon physical principles of the vocal tract. For instance, place of contact between the tongue and the upper surface of the oral cavity is an articulatory dimension known to play a role in distinguishing some consonants from each other (Ladefoged, 1993). Formant frequencies are acoustic dimensions known to play a role in distinguishing vowels from each other. In forward models of the articulatory-acoustic mapping, functions have been derived in order to relate these and other articulatory and acoustic dimensions to a physical model of the vocal tract (e.g., Baer *et al.*, 1991; Goodyear, 2000). Models of the vocal tract used for this purpose are commonly divided into a source of acoustic energy, and a filter through which the source is passed. One of the best known examples is the Kelly and Lochbaum (1962) model of the vocal tract in which the filter is modeled as series of tubes with varying lengths and diameters.

Most forward models developed thus far can be thought of as *theory-driven* because they are, in large part, derived from physical principles of the vocal tract (for an exception in automatic speech recognition, see Blackburn and Young, 2000a). These theory-driven models have served as valuable research tools for relating the underlying theories to empirical data on speech production. The theory-driven approach has also proven instrumental in the development of articulatory speech synthesizers because it reduces the complexity of the vocal tract down to a manageable number of functions.

Here we present a forward model of the articulatory-acoustic mapping that was thoroughly *data-driven* by design. The model was an artificial neural network trained on the articulatory and acoustic recordings from one speaker in the multi-channel articulatory (MOCHA) speech database (Wrench and Hardcastle, 2000), recorded at the Edinburgh speech production recording facility. Inputs to the model were electromagnetic articulograph (EMA), electropalatograph (EPG), and laryngograph (LYG) measurements, each windowed over 64 ms slices of time. The output of the model was a power spectrum of the speech acoustics at the center of each 64 ms slice. The inputs and outputs were coded in the

[a)]Electronic mail: ckello@gmu.edu

model as patterns of activity over sets of connectionist processing units. The mapping from inputs to outputs was governed by a single set of weights on the connections between input and output units, some of which were mediated by hidden units (see Sec. II). Thus, a single, unified set of model parameters had to represent the entire mapping from articulatory inputs to acoustic outputs. The weights were determined by gradient descent learning, which was driven to minimize error between acoustic outputs and their corresponding targets. Acoustic targets were derived from the acoustic recordings.

The model was purely data-driven in that values for the model parameters (i.e., the weights) were learned solely on the basis of articulatory and acoustic data from recorded speech tokens. In other words, the parameters were not *a priori* set on the basis of physical principles of the vocal tract. Moreover, the articulatory and acoustic dimensions were ''raw'' in the sense that they did not directly code articulatory or acoustic features known to convey phonetic information. For instance, articulatory features such as place or manner of articulation were not extracted *a priori* from the articulatory recordings, nor were acoustic features such as formant frequencies. Instead, the articulatory and acoustic data streams were presented to the model in a largely unprocessed format. It is true that some assumptions were built into the model architecture, e.g., that acoustic states could be determined on the basis of a certain window of articulatory data, and that acoustic targets are unimodal (see Sec. II). However, these assumptions were minimal, and in some cases, they were forced by constraints of the articulatory recordings.

The data-driven approach to forward modeling is different from the theory-driven approach in that all empirical data on the vocal tract and the corresponding speech acoustics can be made available to the model. It is the learning procedure and the computational capacity of the model that determines what information is and is not extracted from the data and represented in the model parameters. By contrast, the model parameters in a theory-driven forward model are determined more explicitly by the modeler.

Motivation for a data-driven forward model. In the current modeling work, the data-driven approach was motivated by two aims. First, while theory-driven forward models have proven to be useful research tools, they have not yet enabled the development of natural-sounding articulatory speech synthesizers. One reason for this shortcoming is that, in a theory-driven forward model, many details of the vocal tract and speech acoustics are purposely abstracted away. It is presumably these details (among other factors) that impart the quality of a person's voice. Therefore, one way to achieve more natural-sounding speech synthesis would be to capture as much detail as possible about the vocal tract and speech acoustics for a given speaker (e.g., see also Blackburn and Young, 2000; Jiang *et al.*, 2002; Rowels, 1999; Shiga and King, 2003). The data-driven approach to forward modeling has the potential to capture such details. Relatively little detail about the vocal tract was available for use in the current work (see Sec. III A), but the simulations provided an initial test of the viability of a data-driven articulatory speech synthesizer.

The second aim of the current work was to take a first step toward building a computational model of phonological development. A fundamental question in research on speech acquisition is how does the infant language learner acquire knowledge about the phonological structure of his or her language. Moreover, how is that knowledge represented in the mind and brain of the learner, and then used in language tasks such as spoken word comprehension and production? In recent years, computational models of speech acquisition and production have been developed as tools for exploring and testing the underlying theories (Bailly, 1997; Guenther, 1994, 1995; Guenther, Hampson, and Johnson, 1998; Plaut and Kello, 1999). An integral component of these models is the simulation of babbling and early attempts at the production of spoken words. The forward mapping from articulation to acoustics is essential for such simulations.

The forward model reported here is planned to be one component of a computational model of spoken word acquisition and processing. Plaut and Kello (1999) presented a connectionist model of spoken word acquisition and processing in which distributed representations were learned in the service of speech tasks. The central hypothesis tested in that model was that a learned level of representation exists to (1) integrate the speech signal over word-sized units, (2) generate articulatory trajectories over word-sized units, and (3) map between spoken word forms and their meanings. This level of representation was termed ''phonology'' because its structure was hypothesized to be phonological in nature by virtue of the three core speech tasks that it supported. Thus, the model was aimed at simulating how phonological representations emerge over the course of spoken word acquisition.

On the theoretical approach taken by Plaut and Kello (1999), a central factor in the emergence of phonological representations was their dual purpose in supporting both speech perception and speech production (see Hickok, 2001, for neuroimaging evidence of the existence of dual-purpose representations). As a result of this dual purpose, phonological representations were hypothesized to be shaped, in part, by the intersection of acoustic and articulatory structure in speech. The question, then, is, how does the learning that occurs during the early experiences of speech perception combine with the complementary learning that occurs during speech production to form this intersection. One key part of the answer to this question on the approach taken by Plaut and Kello was that the language learner uses her knowledge of the forward mapping from articulation to acoustics as a bridge between learning in speech perception and learning in speech production. This knowledge was embodied as a forward model of the articulatory-acoustic mapping, and the forward model was learned through simulated babbling (see also Perkell *et al.*, 1997; Perkell *et al.*, 2000).

The forward model played a relatively minor, but absolutely necessary, role in the development of phonological representations. It enabled learning on the input side of the system (i.e., perception and comprehension) to drive learning on the output side of the system. It was not part of the

mechanism that integrated inputs over time to form phonological representations, nor was it part of the mechanism that generated outputs over time to produce articulatory trajectories. Therefore, the current work is intended only to investigate whether the bridge between perception and production can be based on real speech. The forward model is fairly small piece of theory proposed by Plaut and Kello (1999), but the use of real speech would be a major improvement over the original modeling work.

In the Plaut and Kello (1999) simulation, the articulatory and acoustic representations were engineered on the basis of knowledge accumulated over years of phonetics research (e.g., Ladefoged, 1993). For example, articulatory representations included tongue height and backness, and acoustic representations included first through third formant frequencies. A forward mapping from articulatory to acoustic representations was also engineered on the basis of phonetics theory and research, and the task of the forward model was to learn this mapping. Thus, the engineered forward mapping was clearly theory-driven in that it was not derived directly from measurements of speech. As a similar example, Guenther's DIVA model (so named because it maps orosensory Directions Into Velocities of Articulators) of speech acquisition and production also includes a forward model that is based on engineered representations of speech articulations and acoustics (Guenther, 1994, 1995; Guenther *et al.*, 1998).

The simulations reported by Plaut and Kello (1999) and Guenther (1994; Guenther, 1995; Guenther *et al.*, 1998) have been successful in accounting for certain phenomena in speech acquisition and production. For instance, the Plaut and Kello model was able to learn representations that functioned to support the tasks of spoken word comprehension, production, and imitation. The DIVA model has accounted for phenomena of coarticulation, motor equivalence, and speaking rate (among others). Part of what made these successes possible were the simplifying assumptions of the models. Most relevant to the current discussion are the simplifications that were made in the articulatory and acoustic representations, and in the forward mapping between them. These simplifications made the models tractable, and they removed extraneous details that would have made it difficult to relate simulation results to the theoretical principles embodied in the models.

While recognizing the value of theory-driven models, it is fair to ask whether the models offered by Plaut and Kello (1999) and Guenther (1994; Guenther, 1995; Guenther *et al.*, 1998) would scale to handle all of the complexities inherent in the development and processing of real speech. The simulations are meant to serve as evidence for theories of speech acquisition and speech production. However, it is unclear how the models would perform when implemented with more veridical representations of speech articulations and acoustics. If the models were to fail under more veridical conditions, one would have to ask whether the theories were fundamentally flawed in some or way, or whether the failures were only due to shortcomings in the computational machinery.

The use of simplified articulatory and acoustic representations also raises questions about the successes of the models, especially with respect to the Plaut and Kello (1999) model. The most relevant question for the current discussion is the following: does the simulated learning of phonological representations stand as support for Plaut and Kello's theory of phonological development, or did this success depend crucially on the simplifications in the articulatory and acoustic representations? For instance, a major issue in phonological development is how sensitivity to the segmental structure of speech emerges from language experience (e.g., see Bernhardt and Stemberger, 1998; Jusczyk, 1997). On the theory proposed by Plaut and Kello, sensitivity to segmental structure is primarily a product of the articulatory and acoustic structure of speech, and the statistical regularities in the speech inputs that come from adults and other children (of course, neuroanatomy, neurophysiology, and mechanisms of learning also play their respective roles). However, in the simulation reported by Plaut and Kello, segmental structure was partially engineered into the articulatory and acoustic representations. This engineering may have been key to the learning of phonological representations in that simulation. Thus, the simulation results left open the question of whether phonological representations can be learned from articulatory and acoustic representations in which no segmental structure is imposed.

The forward model reported here is a first step toward addressing these and other questions. The articulatory inputs and acoustic outputs used in the current model were derived directly from articulatory and acoustic recordings of a female speaker of British English. The procedures for preprocessing the recordings were designed such that segmental structure was not pre-extracted from the data streams. This is not to say that segmental structure is unimportant to speech; a key test of any model of phonological development would to be show that it is sensitive to the segmental structure of speech in the same way that humans are. The point here is that a complete model would need to explain *how* the language learner becomes sensitive to segmental structure in the native language, given only the raw speech signal as input. The forward models reported here do not explain this aspect of learning; on the Plaut and Kello (1999) theory, the learning of segmental structure is explained by other mechanisms. What the current work provides is a necessary first step toward building a more complete model of phonological development based on real speech.

In addition to this long-term purpose, the current work also served two more immediate purposes. One immediate purpose was to test the viability of a data-driven articulatory speech synthesizer. To the extent that the reported forward models are successful, they will output acoustics that are identical to that of the targets derived from the speaker. However, it is important to note that a forward model does not constitute a speech synthesizer because it does not specify how to control the articulatory dimensions (in the current work, articulatory states always came from the speech database). Nonetheless, a data-driven forward model that can output natural-sounding speech may be an important step toward building a complete, data-driven, articulatory speech synthesizer.

The second immediate purpose of the current work was to generate a lower-bound estimate on the amount of phonetic information that is captured by the articulatory recordings in the MOCHA speech database. The task of the forward models reported herein was to generate the acoustic outputs of articulatory inputs as veridically as possible, defined as the minimization of squared error. If the models could perform this task perfectly, it would mean that the articulatory recordings had captured enough information about the vocal tract to generate all of the acoustic detail in the recordings of the speaker's voice. There was no expectation of perfection, but any phonetic information conveyed by the models had to originate in the articulatory recordings. Therefore, the forward models provided a lower bound on the phonetic information available in the articulatory recordings. The forward models could not provide an upper bound because it is possible that the articulatory recordings contained more phonetic information than measured in the acoustic outputs; there is no guarantee that all phonetic information was extracted by the models, or conveyed by our measures of phonetic information.

II. MODEL

Three forward models were trained on the recordings for one speaker in the MOCHA database. The *all* model was trained on all 460 sentences recorded by the speaker, the *even* model was trained on the even-numbered sentences, and the *odd* model was trained on the odd-numbered sentences. The odd/even split was arbitrary, and was used to test the generalization of the learned model parameters to inputs that were not presented during training. Specifically, the odd-numbered sentences were used to test generalization of the even model, and vice versa for the odd model. Tests of generalization served to ensure that the model parameters captured the general relationship of the vocal tract and the resulting acoustics, rather than individual input/output pairings or some unknown peculiarities in the speech database.

A. Speech database and pre-processing

Speech tokens were drawn from one female speaker of British English (subject ID ''fsew,'' southern dialect) in the MOCHA speech database, recorded at the Edinburgh speech production recording facility. The speech corpus consisted of one token each of 460 phonetically compact sentences designed to provide a good coverage of pairs of phones, with extra occurrences of phonetic contexts thought to be either difficult or of particular interest. The corpus included all 450 phonetically compact TIMIT (sx) sentences, plus ten additional sentences designed to include phonetic pairs and contexts that are particular to British English.

Articulatory recordings in the MOCHA database consists of electromagnetic articulograph (EMA), electropalatograph (EPG), and laryngograph (LYN) recordings. The EMA recordings consisted of eight sensors placed in the midsagittal plane of the vocal tract, attached to the following locations: the vermilion border of the upper lip, the vermilion border of the lower lip, the upper incisor, the lower incisor, the tongue tip (5–10 mm from the tip), the tongue blade (approximately 2–3 cm posterior to the tongue tip sensor), the tongue dorsum (approximately 2–3 cm posterior to the tongue blade sensor), and the soft palate (approximately 10–20 mm from the edge of the hard palate). $[X,Y]$ positions were recorded from each sensor, sampled at 500 Hz.

The positions of these eight sensors were used to calculate nine $[X,Y]$ pairs of articulatory dimensions, as follows. One $[X,Y]$ pair coded the position of the lower incisor (i.e., jaw movement) relative to the upper incisor. This relative coding removed head movement because the position of the upper incisor sensor was fixed relative to the head. Two $[X,Y]$ pairs coded the positions of the upper and lower lips, relative to the positions of the upper and lower incisors, respectively. These pairs coded lip movement independent of head and jaw movement. One $[X,Y]$ pair coded movement of the soft palate relative to the upper incisor. Two $[X,Y]$ pairs coded the overall position of the tongue as the average of the three tongue sensors, one pair in absolute coordinates, and one pair relative to the upper incisor. Finally, three $[X,Y]$ pairs coded each of the three tongue positions, relative to the absolute average tongue position. These three pairs coded local movements of the individual sensors independent of more global movements of the entire tongue.

EPG sensors were placed in 48 normalized positions on the hard palate defined by landmarks on the upper maxilla. Contact between the tongue and each EPG sensor (binary values) was sampled at 200 Hz. LYN recordings provided voicing information at the larynx as a wave form sampled 16 kHz, stored with 16 bit precision, and low-pass filtered at 400 Hz. Acoustic recordings were also sampled at 16 kHz and stored with 16 bit precision, but they were low-pass filtered at 8 kHz instead of 400 Hz.

The acoustic and LYN recordings were transformed from the time domain to the frequency domain with the use of Matlab's fast Fourier transform (FFT) routine. FFTs were calculated over hamming windows 64 ms wide, taken at 32 ms intervals. We explored a range of widths and found 64 ms to produce the most intelligible reconstructed speech signal (see Sec. III). Given the sample rate of 16 kHz, this procedure resulted in 511 frequency bins of log magnitude per window after discarding the dc offset. Phase information in the acoustic signal was discarded in the FFT conversion because the articulatory recordings were not expected to carry phase information (the loss of phase information was partly responsible for the need for relatively wide processing windows). Only the lower 25 bins were used for the LYN recordings because the signal was low-pass filtered at 400 Hz. The rear 24 EPG sensors were discarded because they were not activated in the recordings for the chosen speaker.

For each dimension in the acoustic, EMA, and LYN data streams, the observed values across the entire data set were rank-ordered, and the smallest 100 values were set equal to the 100th smallest value, and the largest 100 values were set equal to the 100th largest value. This procedure normalized very extreme outliers in each dimension of the data streams, thereby restricting their range. The restricted range for each dimension was then normalized to [0,1]. This normalization procedure was not necessary for the EPG data because those dimensions were already normalized in the range [0,1]. Finally, the EMA and EPG data streams were down-sampled to

31.25 Hz, and aligned with the FFT windows calculated over the acoustic and LYN data streams.

B. Articulatory and acoustic representations

Outputs of the forward models were vectors of real numbers in the range [0,1] that represented the acoustic power spectrum at a given 64 ms slice of time. The vectors were 1022 dimensions in size. For each of the 511 FFT bins, one dimension represented the values in the range [0,0.5), and another dimension represented values in the range [0.5,1]. Values outside of a given unit's range were set to zero on that unit. This output format allowed for better resolution in the model's representations, and separate parameters for learning between the upper and lower ranges (i.e., separate sets of connection weights fed into the upper-range and lower-range output units; see Sec. III C).

Inputs to the forward model were vectors of real numbers in the range [0,1] that represented the previous (32 ms in the past), current, and next (32 ms in the future) articulatory states, relative to the acoustic outputs. The input vectors were 588 dimensions in size, with one third each representing the previous, current, and next articulatory states. Each point in time consisted of 72 dimensions dedicated to EMA positions, 24 dimensions dedicated to EPG contact, and 100 dimensions dedicated to FFT values from the LYN recordings. The EPG dimensions directly coded the average amount of tongue contact in a given slice of time for each of the front 24 EPG sensors. Four dimensions were assigned to each of the 18 EMA dimensions (i.e., nine pairs of $[X,Y]$ positions), and each of the 25 bins of FFT magnitude (up to 400 Hz in frequency) for the LYN recordings. For each quadruple assigned to value x, one dimension coded x directly, one coded the value $1-x$, one coded the x values in the lower range [0,0.5), and one coded x values in the upper range [0.5,1]. Analogous to the output format, the split range format provided the model with a separate set of parameters for the lower and upper ranges of input values. To complement, the $x|1-x$ inverse coding provided two sets of parameters that spanned the full range of input values. The inverse coding was used to ensure that learning occurred on every training example, for each dimension, regardless of each dimension's value. In backpropagation, no learning will occur on a unit's sending weights when the activation value of that unit is zero. Thus, the x and $1-x$ units served to provide model parameters learned on either side of each dimension.

C. Forward model training and results

All three forward models had the neural network architecture depicted in Fig. 1. Each model consisted of 588 input units, 1022 output units, and 100 hidden units. Acoustic representations corresponded to patterns of activity over the output units, and articulatory representations corresponded to patterns of activity over the input units. Patterns of activity over the hidden units corresponded to internal representations that were learned over the course of training (see the following). The activation value for each hidden unit and each output unit was calculated as the sigmoid of the dot product of the unit's incoming weights and the outputs of the pre-synaptic units. The activation values of the input units were set directly equal to the articulatory representation at a given point in time in one of the trained sentence tokens (i.e., composed of previous, current, and next articulatory states as described earlier). Every articulatory input unit was connected to every acoustic output unit (i.e., full connectivity). In addition, articulatory inputs were fully connected to the hidden units, and the hidden units were fully connected to the output units. Direct connections between inputs and outputs were included to increase the rate of learning by facilitating the extraction of any linear relationships between the articulatory and acoustic dimensions. The hidden units served to capture nonlinear relationships between the input and output dimensions, although they were free to capture linear relationships as well. A total of 100 hidden units was chosen on the basis of trial and error; pilot work indicated that model performance was worse with fewer hidden units, and no better with more hidden units.

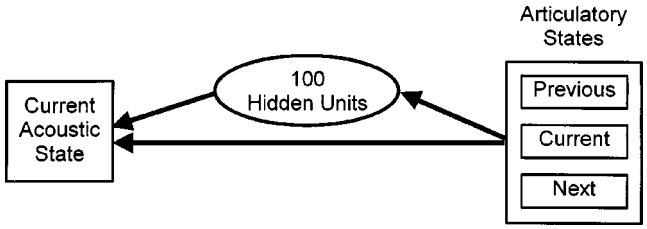


FIG. 1. Architecture of the forward models. Arrows indicate full connectivity between groups of processing units.

At the start of training, the weights on all connections in the network were drawn randomly with replacement from a rectangular distribution in the range $(-0.1,0.1)$. Weights were learned via the backpropagation of error signals generated on the outputs units (Rumelhart, Hinton, and Williams, 1986). In particular, time slices from the sentence tokens were presented to the network in batches of 100, sampled at random from the training set. For each time slice, the activation values of the input units were set to the corresponding articulatory representation, and those activation values were propagated forward through the network connections to generate a pattern of activation on the output units. For each output unit, squared error was calculated between the unit's activation value, and its target activation, which was determined by the acoustic representation for the time slice in question. The error signals were then backpropagated along the network connections to calculate weight derivatives. Each weight's derivatives were summed across each batch of 100 training examples (i.e., time slices). After each batch, the summed derivatives were used to update each weight according to

$$\Delta w_{ij}^{[b]} = \eta_N \eta_{ij} \frac{\partial E}{\partial w_{ij}} + \alpha(\Delta w_{ij}^{[b-1]}), \tag{1}$$

where η_N was the overall network learning rate (decreased from $5e-4$ to $5e-5$ over the course of training), η_{ij} was a weight-specific learning rate, α was a momentum term (fixed at 0.8), and b was the Nth batch over the course of training. Weight-specific learning rates were adjusted on the basis of

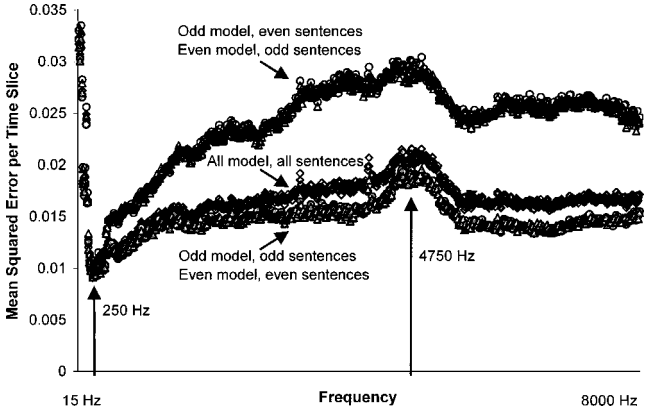


FIG. 2. Mean squared error per output unit per time slice, as a function of frequency for the *all* model, the *even* model, and the *odd* model.

the consistency of weight derivatives across batches (Jacobs, 1988).

The 460 sentences in the training set for the *all* model contained a total of 41 791 slices of time to be trained. The *even* model contained 20 794 slices, and the *odd* model contained 20 997 slices. Each model was trained on 100 000 batches of training examples, which is about the point at which the reduction in error became miniscule. All three models were stopped at exactly 100 000 batches to control for amount of training. The training sets did not appear to be overfit because, for the odd and even models, error on the untrained sentences decreased throughout training. At the end of training, the average squared error between the targets and outputs was calculated per frequency per time slice. These averages are shown in Fig. 2 for each of the three forward models, separated by sentence type (odd- or even-numbered). As can be seen, the mean squared error never exceeded 0.035 by the end of training. By comparison, mean squared error at the beginning of training was 0.193.

Figure 2 shows that there was a clear rank ordering in the overall amount of model error. The models trained on half of the sentences in the speech database (*even* and *odd* models) produced the least amount of error on their respective training sets, and the most amount of error on sentences outside their training sets. The *all* model produced slightly more error than the *even* and *odd* models for their respective training sets, but substantially less error than those models for sentences outside their training sets. This rank-ordering of error indicates that some learning did not generalize beyond the training sets in the *odd* and *even* models, and that error from these models was reduced somewhat by learning that was specialized to the training sets.

The pattern of error as a function of frequency was mostly similar across the different models and training sets. There was a sharp dip in error at about 250 Hz, followed by a fairly steady climb in error to a peak at about 4750 Hz. Error then dropped to a middling baseline level that was maintained out to 8000 Hz. This pattern was somewhat different for the *odd* and *even* models tested outside their training sets in that, for those models, an extra plateau of error can be seen prior to the peak at about 4750 Hz. The dip at 250 Hz is due to the fact that the LYN recordings contain fairly direct information about acoustic energy around the pitch of the speaker's voice. The peak at about 4750 Hz might have been due to the models' inability to determine the spectral details of acoustic energy generated by fricative and plosive speech sounds (but this conjecture needs further investigation). The reasons for the particular characteristics of the rise in error up to its peak, and its drop off after the peak, are currently unknown.

The error scores gave a detailed picture of which frequency bands were processed more or less accurately by the models, but these scores do not give an interpretable measure of intelligibility of the target and model tokens. To provide a more standard measure, an energy histogram method (Hirsch, 1995) was used to estimate the signal-to-noise ratio (SNR) in the target and model tokens, as well as in the original, unprocessed tokens. The mean SNR for each category of tokens was as follows: 28.2 dB for the original tokens, 24.4 dB for the target tokens, 25.8 dB for the *all* model tokens, 26.0 for the *even* model tokens, and 25.7 dB for the *odd* model tokens. These means show that, although the original tokens were distinguished from the processed ones, the energy histogram method did not reflect the overall differences in error scores plotted in Fig. 2. A SNR measure that uses the target signal as a baseline would be more appropriate, but the removal of phase information in the target and model tokens prohibited such a measure. To allow other researchers to experiment with various measures of intelligibility, the wave forms all of the stimuli in all four conditions can be downloaded at http://archlab.gmu.edu/~ckello/forward-models.html.

The error scores and SNRs are quantitative, objective measures of intelligibility. A more qualitative way to assess the modeling results is to view spectrograms of the target utterances, and compare them against spectrograms of the corresponding model outputs. In Figs. 3 and 4, spectrograms are shown for one odd-numbered and one even-numbered example sentence, each chosen arbitrarily from the speech database. At a glance, the model spectrograms are quite similar to the target spectrograms. Some of the spectral and temporal details in the targets appear to be washed out in the model outputs, particularly above 3000 Hz where the harmonics appear to be completely washed out. These spectrograms are informative visualizations, but ultimately, the forward models must be assessed by measuring the amount of phonetic information contained in their outputs. Such an assessment is reported in Sec. III.

III. INTELLIGIBILITY TESTS

The error results shown in Fig. 2 provide a quantitative measure of performance for the forward models, and Figs. 3 and 4 provide a more qualitative measure. However, it is difficult to interpret these measures in terms of the amount of phonetic information that was captured in the forward mapping learned by the models. To better estimate the phonetic information captured in the models, the model outputs and targets were submitted to empirical tests of intelligibility. Intelligibility of the targets served as a baseline comparison. The percentage of words identified correctly was used as a coarse measure of the overall amount of phonetic information captured by the models, relative to the phonetic infor-

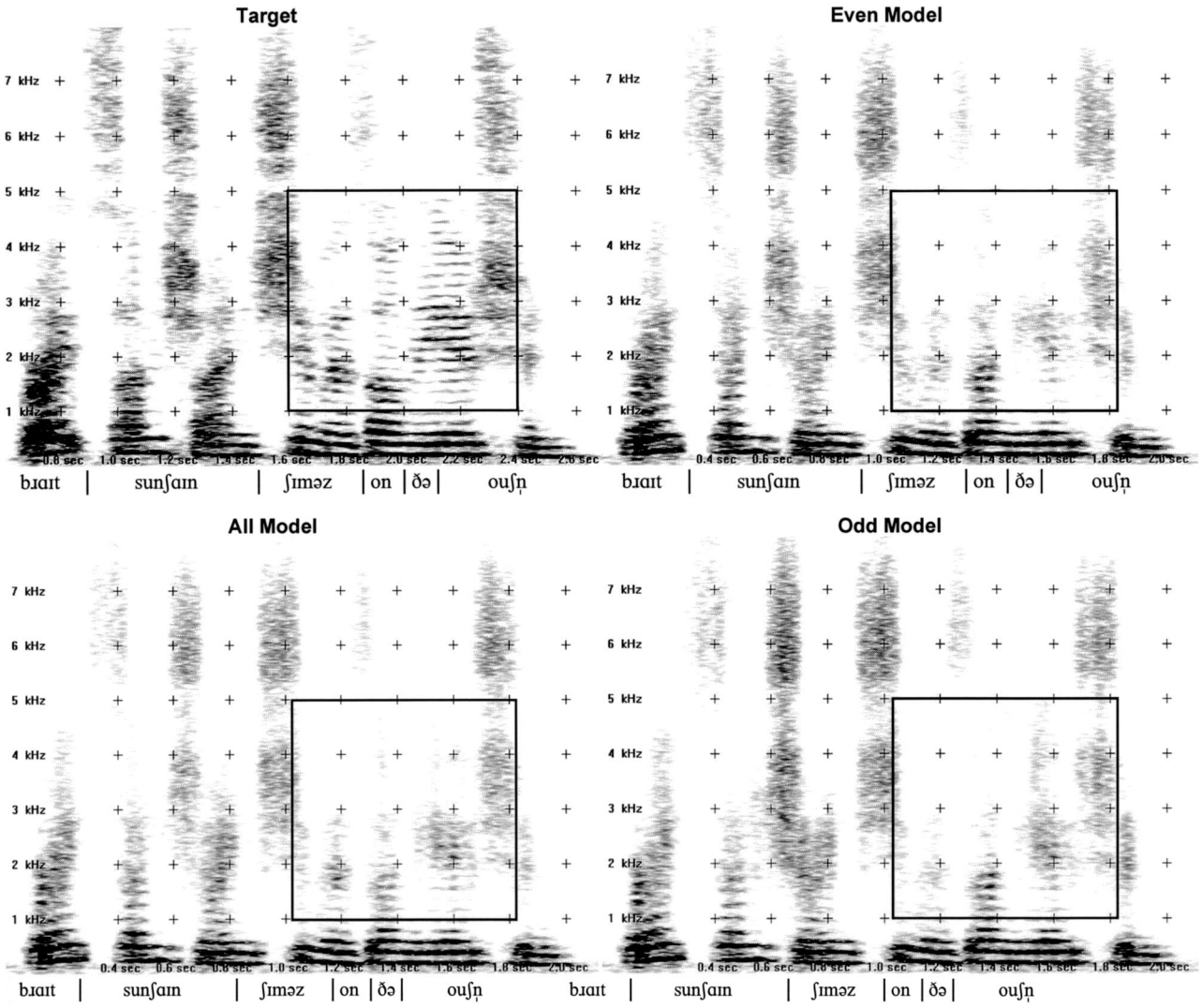


FIG. 3. Spectrograms of the target acoustics, and the acoustics output by each of the three model types for an even-numbered sentence token, ''bright sunshine shimmers on the ocean.'' The outlined region shows where the models can be seen to have lost some of the spectral detail in the target utterance.

mation available in the targets. To provide a rough measure of the kinds of phonetic information that was lost by the models, phoneme confusions were identified in the responses (when possible), and tabulated.

A. Methods

1. Participants

Eight undergraduates participated as listeners in the speech intelligibility tests for course credit. All participants reported being native speakers of American English, none reported a hearing impairment, and none were familiar with the TIMIT speech database.

2. Stimuli

All 460 sentence tokens were passed through each of the three forward models to generate a series of acoustic outputs for each token, and from each model. The Matlab inverse FFT routine was used to convert the acoustic outputs into an acoustic wave form for each sentence token, from each model. The same procedure was also applied to the targets, resulting in four stimulus tokens for each sentence: one from each of the three model types, and one from the target. As noted earlier, some information in the original acoustic recordings was lost because it was necessary to discard phase information in the FFT procedure used to generate the acoustic targets for the models. Phase information was replaced by inserting random phases into the inverse FFT procedure. Pilot tests indicated that random phases produced more intelligible wave forms compared with phases fixed at values such as zero. However, loss of the original phase information caused some distortion in the generated wave forms.

3. Procedure

Participants were seated in a quiet booth and instructed that they would be listening to grammatically correct and semantically plausible English sentences. For each sentence, they were instructed to transcribe what they heard to the best of their ability. They were told that some sentences were garbled and therefore difficult to hear. They were asked to type into the computer as many words as they heard for each

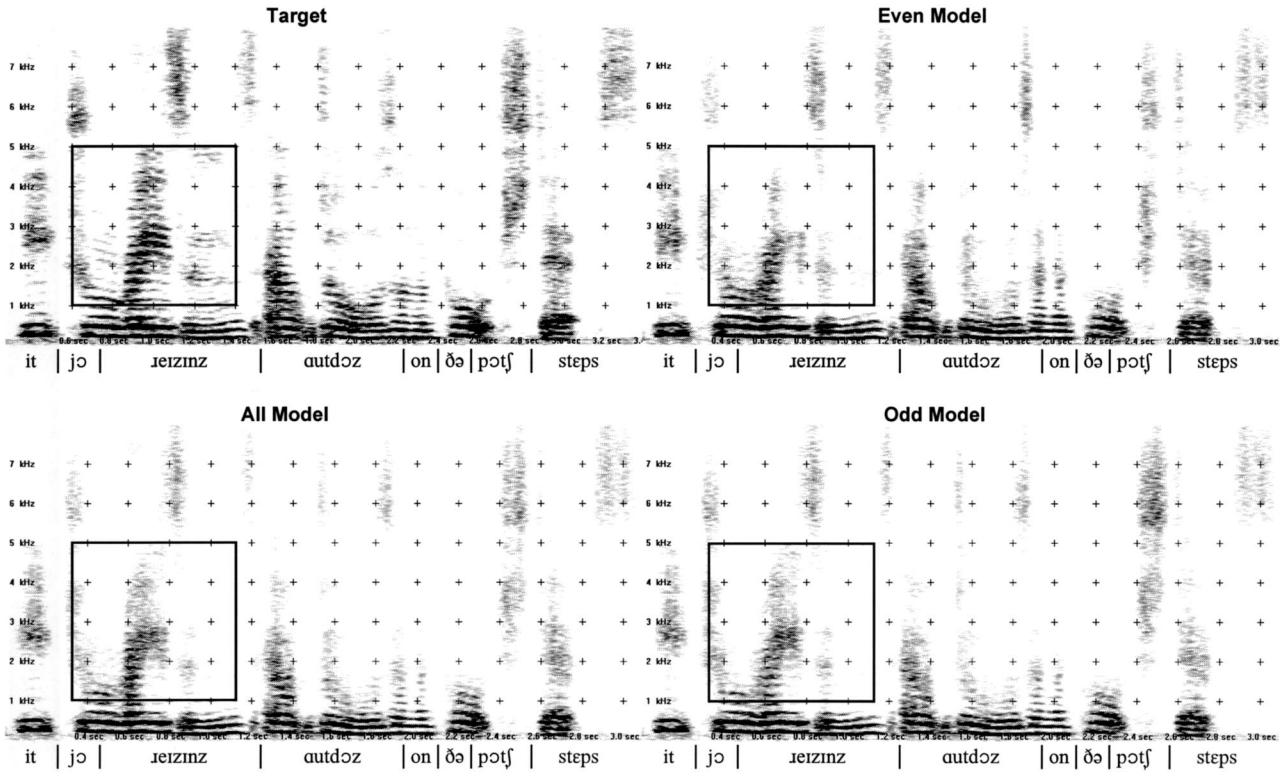


FIG. 4. Spectrograms of the target acoustics, and the acoustics output by each of the three model types for an odd-numbered sentence token, ''eat your raisins outdoors on the porch steps.'' The outlined region shows where the models can be seen to have lost some of the spectral detail in the target utterance.

sentence, in the order that they heard them, and they were encouraged to guess at words whenever necessary and possible.

Stimulus presentation and data collection was controlled through a graphical user interface, and stimuli were presented over Sennheiser MH80 headphones at a comfortable listening level that was held constant across participants. Each trial began with the participant clicking on a button to listen to the current sentence. Participants were forced to click on this button three times in order to listen to each sentence three times before responding. Participants typed each response into a text entry field, and clicked on another button to enter the response and begin the next trial. Participants were not given feedback at any time.

Participants were given four practice sentences at the beginning of the experiment, followed by one-fourth (115) of the 460 sentence tokens. Tokens were rotated across subjects to cover all 460 sentences evenly, and each sentence appeared in two of the four token conditions. The token conditions were rotated across subjects such that each condition was sampled an equal number of times.

B. Results

All responses were corrected for spelling errors, and in the few cases where the participant responded with a homophone of the correct word response (e.g., responding with TACKS when the correct word is TAX), the homophone was replaced with the correct word. The percentage of words transcribed correctly for the even-numbered and odd-numbered sentences is graphed in Fig. 5 for each of the four token conditions. The graph shows that the target outputs were transcribed most accurately, with no noticeable difference in accuracies for the even-numbered versus odd-numbered sentences. There are at least three possible reasons why the intelligibility of the targets was less than perfect: (1) the tokens were generated by a speaker of British English, but the listeners were speakers of American English, (2) some of the TIMIT sentences contain words likely to be unfamiliar to the participants (e.g., ''neoclassic,'' ''Nan,'' ''statuesque,'' etc.), and (3) phase information in the original recordings was lost in the FFT procedure.

The graph also shows that accuracy for the outputs of the *all* model was 11 percentage points lower on average than that for the targets, $t(7)=7.4$, $p<0.001$. Compared with the *all* model, similar levels of accuracy were found for

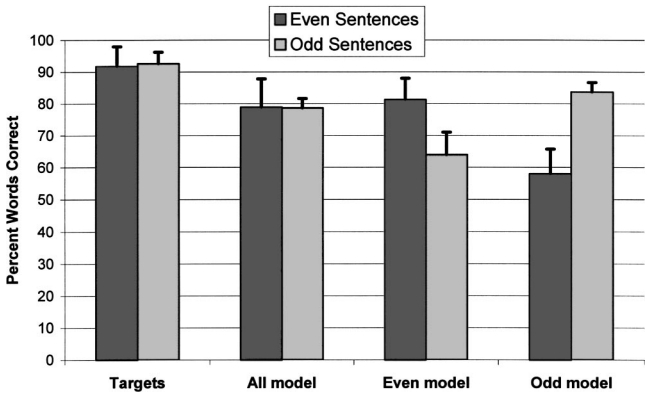


FIG. 5. Mean percent words correct in the intelligibility tests, as a function of token type (target, *all* model, *even* model, or *odd* model) and sentence type (even-numbered or odd-numbered sentences). Error bars show standard deviations of the subject means.

TABLE I. Counts of phoneme confusions summed across all model conditions and all subjects, with the number of times that each confused phoneme appeared in the corpus included as a baseline. Counts were collapsed across order of confusion, and counts under three were not included.

Confusion	Articulatory dimensions confused	No. Observed	No. in corpus
/l, ɹ/	Place	9	1279
/p, b/	Voicing	6	677
/m, b/	Manner	4	717
/n, d/	Manner, nasal	4	1343
/n, l/	Nasal, lateral	4	1361
/k, p	Place	4	901
/n, m/	Place	4	1301
/j, d/	Place, manner	4	669
/k, h/	Place, manner	4	706
/k, f/	Place, manner	4	797
/j, t/	Place, manner, voicing	4	1008
/n, r/	Place, nasal	4	1548
/k, b/	Place, voicing	4	846
/k, t/	Place	3	1402
/b, f/	Place, manner, voicing	3	573
/ç, j/	Voicing	3	253
/d, t/	Voicing	3	1395

the *even* model tested on the even-numbered sentences and the *odd* model tested on the odd-numbered sentences. These model results are reported primarily as points of comparison for the tests of generalization. In particular, accuracy was 20 percentage points lower for the *even* model tested on the odd-numbered sentences, compared with the same model tested on the even-numbered sentences, $t(6)=3.5$, $p<0.05$. Similarly, accuracy was 18 percentage points lower for the *odd* model tested on the even-numbered sentences, compared with the same model tested on the odd-numbered sentences, $t(6)=6.1$, $p<0.001$.

To provide a rough measure of the kinds of phonetic information that were lost by the models, responses to all model outputs were inspected for phoneme confusions. Responses were identified in which a target word in the sentence was clearly replaced with a different word in the response. On this strict criterion, replacements were identified for only 57% of the responses with errors. The difficulty was that participants often left target words out of their responses, or occasionally inserted words that were not in the target sentences. Deletions and insertions often made it difficult to align a given response with its target. To avoid experimenter bias in alignment decisions, no word replacements were identified when the alignment was ambiguous.

The counts of phoneme confusions are shown in Table I. These counts are collapsed across confusion order (i.e., phoneme A replaced with phoneme B, or B with A), and only counts greater than two are shown. In the full set of confusions, vowels were never confused with consonants, and of the few vowel–vowel confusions that were identified, no particular vowel–vowel confusion occurred more than twice. With respect to consonants, the phoneme pairs /l, ɹ/ and /p, b/ were confused most often. This was true in terms of raw counts, and counts relative to the number of times these phonemes appeared in the corpus. Otherwise, features that denote place of articulation were confused most often, but features denoting voicing and manner were also confused with some regularity. Confused features were often similar to each on their respective dimension of confusion; for example, the feature plosive was often confused with the feature affricate, and both manners of articulation are characterized by a burst release. Beyond these general statements, it is difficult to be more specific without results from an experiment aimed more directly at phoneme identification. The MOCHA database contained only sentence stimuli, which made it prohibitively difficult to conduct a phoneme identification experiment.

IV. DISCUSSION

Three neural network models were trained on the articulatory-acoustic mapping for one speaker in the MOCHA speech database. Results indicated that this mapping was well-approximated in the models. Spectrograms and analyses of model error showed that the acoustic outputs in lower frequency range (below 2000 Hz) closely matched the target outputs, whereas acoustic outputs in the upper frequency range (above 3000 Hz) were less accurate. Intelligibility tests showed that listeners could identify a large percentage of words in sentences that were generated by passing the recorded articulatory trajectories through the models. These tests also showed that the model parameters generalized, to some degree, to novel articulatory inputs. On the one hand, intelligibility of the untrained sentences (61% words correct on average) demonstrated that the model learned something about the general relationship between articulatory and acoustic parameters for the speaker's vocal tract. On the other hand, reduced intelligibility of the untrained sentences compared with the trained sentences (81% words correct on average) indicated that some aspects of the general articulatory-acoustic relationship were not learned sufficiently.

The intelligibility tests provided coarse measures of phonetic information in that phoneme confusions provided only a rough measure of the kinds of phonetic information that were and were not contained in the acoustic outputs. Other measures of phonetic information, such as those derived from tests of phoneme identification (e.g., Bernstein, Demorest, and Tucker, 2000), would provide more detail about phonetic information in the model outputs. Unfortunately, only the sentence recordings were available for intelligibility tests, and it would have been difficult to specifically test phoneme identification with sentence stimuli.

Nonetheless, the results in hand demonstrate that the forward mapping from articulations to acoustics can be learned, at least to a reasonable extent, via a heuristic of gradient descent (i.e., backpropagation) in an acoustic error space. They also place a lower bound on the amount of phonetic information captured by the articulatory recordings in the MOCHA database. In particular, articulatory recordings were comprised of 8 mid-sagittal $[X,Y]$ positions at key locations in the vocal tract, 24 positions of tongue contact with the hard palate, and FFTs of the acoustic energy generated at the larynx. These articulatory recordings were sufficient to generate much of the spectral and temporal information in the resulting speech acoustics. The phonetic information in the

models' outputs had to come from the articulatory inputs because the models were data-driven, i.e., they did not contain any *a priori* information about speech.

In fact, it is possible that the articulatory recordings actually contained more phonetic information than indicated by the reported models. Some information was lost in preprocessing the articulatory recordings in order to format them for the models (e.g., phase information was lost in the FFT procedure), and some of this lost information may have been phonetic in nature. Even if no information was lost in pre-processing, it is possible that the mapping defined by the model parameters (connection weights) did not capture all of the phonetic information in the articulatory inputs. This shortcoming can occur because of an inadequacy in backpropagation, in the use of sigmoidal processing units, or in the representational scheme used on the inputs or outputs. Gradient descent learning can settle into a local minimum in error. While any differentiable function can be approximated using sigmoidal hidden units (Cybenko, 1989), some functions are better suited than others for this particular basis function, and the generalization of learning can be influenced by the choice of activation function (Rumelhart *et al.*, 1995). Finally, it is well known that the design of input and output representations is critical to learning and performance in all neural networks (e.g., Plaut *et al.*, 1996). Thus, it is possible that an alternate method of modeling would have resulted in a forward mapping that captured more phonetic information than the models reported herein.

One reason to improve the fidelity of the current forward models is for the purpose of an articulatory speech synthesizer. Acoustic outputs of the reported models were natural-sounding in that they captured the quality of the speaker's voice, although limitations of the models caused their outputs to sound as if they were masked by noise of some kind. Thus, the models might contribute to the development of a natural-sounding articulatory synthesizer if their fidelity was improved. However, a formidable hurdle in such an effort would be to manipulate the articulatory dimensions such that any desired utterance could be produced. All sequences of model outputs reported in the current work were generated from articulatory sequences in the speech database. However, a speech synthesizer must be able to synthesize any given sequence of phones. Modeling articulatory trajectories is known to be a difficult problem (e.g., Kaburagi and Honda, 2001), and the large number of articulatory dimensions used in the current models are likely to exacerbate this problem. Traditionally, articulatory degrees of freedom are reduced and made independent by means of theoretical (Mermelstein, 1973) or empirical (e.g., Badin *et al.*, 2002; Beautemps, Badin, and Bailly, 2001; Blackburn and Young, 2000b) methods. Such methods could be applied to the current forward models, or alternatively, a concatenative method (see Chappell and Hansen, 2002) could be applied to articulatory trajectories recorded specifically for phones or diphones. In any case, further work is necessary to determine whether the forward models reported here could be used in an articulatory speech synthesizer.

Just as a forward model is only one possible component of an articulatory speech synthesizer, it is also only one possible component of a full-scale model of speech acquisition and production. As argued in Sec. I, it would be informative to test whether models of speech acquisition and production can handle the complexities of real speech. The incorporation of a data-driven forward model, similar to the models reported here, would be a significant step toward such a test. However, some difficult problems would need to be addressed before a complete model could be implemented.

For instance, the problem of articulatory control that confronts the development of articulatory speech synthesizers would also confront the development of models of speech acquisition and production. Guenther's DIVA model (Guenther, 1994, 1995; Guenther *et al.*, 1998) has accounted for a number of phenomena that are relevant to the issue of articulatory control, but it is currently unknown whether the DIVA model would scale to handle the control of a real human vocal tract. The Plaut and Kello (1999) approach is well-suited to forward models such as the ones reported here, given that both share the same mechanisms of neural network learning and processing. However, it is currently unknown whether such mechanisms are capable of learning phonological representations on the basis of real speech input. Another issue that would have to be confronted is variability in the speech signal (e.g., see Perkell and Klatt, 1986; Pisoni, 1981). For instance, on any given occasion, the speech signal that corresponds to a given word will be shaped by factors such as the linguistic and nonlinguistic context, and the talker's dialect and voice quality. The resultant variability poses a significant challenge for any effort to build a computational model of speech acquisition and production. The modeling work reported here is one step toward meeting these and other challenges inherent to the research and engineering of speech.

ACKNOWLEDGMENTS

We would like to thank Brandon Beltz, Laura Leach, and Dana Morgan for conducting the intelligibility tests. We would like to also thank Dana Morgan for coding the data from the intelligibility tests. This work was funded in part by NIMH Grant No. MH55628, and NSF Grant No. 0239595.

Badin, P., Bailly, G., Reveret, L., Baciu, M., Segebarth, C., and Savariaux, C. (**2002**). "Three-dimensional linear articulatory modeling of tongue, lips, and face, based on MRI and video images," J. Phonetics **30**, 533–553.

Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W. (**1991**). "Analysis of vocal-tract shape and dimensions using magnetic-resonance-imaging-vowels," J. Acoust. Soc. Am. **90**, 799–828.

Bailly, G. (**1997**). "Learning to speak. Sensori-motor control of speech movements," Speech Commun. **22**, 251–267.

Beautemps, D., Badin, P., and Bailly, G. (**2001**). "Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling," J. Acoust. Soc. Am. **109**, 2165–2180.

Beautemps, D., Badin, P., and Laboissiere, R. (**1995**). "Deriving vocal-tract area functions from midsagittal profiles and formant frequencies—A new model for vowels and fricative consonants based on experimental-data," Speech Commun. **16**, 27–47.

Bernhardt, B. H., and Stemberger, J. P. (**1998**). *Handbook of Phonological Development: From the Perspective of Constraint-based Nonlinear Phonology* (Academic, San Diego).

Bernstein, L. E., Demorest, M. E., and Tucker, P. E. (**2000**). "Speech perception without hearing," Percept. Psychophys. **62**, 233–252.

Blackburn, C. S., and Young, S. J. (**2000**). ''A self-learning predictive model of articulator movements during speech production,'' J. Acoust. Soc. Am. **107**, 1659–1670.
Blackburn, C. S., and Young, S. (**2000a**). ''Enhanced speech recognition using an articulatory production model trained on X-ray data,'' Comput. Speech Lang. **15**, 195–215.
Blackburn, C. S., and Young, S. (**2000b**). ''A self-learning predictive model of articulator movements during speech production,'' J. Acoust. Soc. Am. **107**, 1659–1670.
Chappell, D. T., and Hansen, J. H. L. (**2002**). ''A comparison of spectral smoothing methods for segment concatenation based speech synthesis,'' Speech Commun. **36**, 343–374.
Cybenko, G. (**1989**). ''Approximation by superpositions of a sigmoid function,'' Math. Control, Signals, Syst. **2**, 303–314.
Goodyear, C. C. (**2000**). ''Incorporating lip protrusion and larynx lowering into a time domain model for articulatory speech synthesis,'' Comput. Speech Lang. **14**, 211–226.
Greenwood, A. R., Goodyear, C. C., and Martin, P. A. (**1992**). ''Measurements of vocal-tract shapes using magnetic-resonance-imaging,'' IEEE Proc.-I: Commun. Speech Vision **139**, 553–560.
Guenther, F. H. (**1994**). ''A neural-network model of speech acquisition and motor equivalent speech production,'' Biol. Cybern. **72**, 43–53.
Guenther, F. H. (**1995**). ''Speech sound acquisition, coarticulation, and rate effects in a neural-network model of speech production,'' Psychol. Rev. **102**, 594–621.
Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). ''A theoretical investigation of reference frames for the planning of speech movements,'' Psychol. Rev. **105**, 611–633.
Hickok, G. (**2001**). ''Functional anatomy of speech perception and speech production: Psycholinguistic implications,'' J. Psycholinguist. Res. **30**, 225–235.
Hirsch, H. G., and Ehrlicher, C. (**1995**). ''Noise estimation techniques for robust speech recognition,'' Paper presented at the Proc. ICASSP.
Jacobs, R. A. (**1988**). ''Increased rates of convergence through learning rate adaptation,'' Neural Networks **1**, 295–307.
Jiang, J. T., Alwan, A., Keating, P. A., Auer, E. T., and Bernstein, L. E. (**2002**). ''On the relationship between face movements, tongue movements, and speech acoustics,'' Eurasip J. Appl. Signal Process. **2002**, 1174–1188.
Jordan, M. I., and Rumelhart, D. E. (**1992**). ''Forward models—Supervised learning with a distal teacher,'' Cogn. Sci. **16**, 307–354.
Jusczyk, P. W. (**1997**). *The Discovery of Spoken Language* (MIT, Cambridge).
Kaburagi, T., and Honda, M. (**2001**). ''Dynamic articulatory model based on multidimensional invariant-feature task representation,'' J. Acoust. Soc. Am. **110**, 441–452.
Kelly, J. L., and Lochbaum, C. C. (**1962**). ''Speech Synthesis,'' Proceedings of the Fourth International Congress of Acoustics, paper G42, 1–4, in *Speech Synthesis*, edited by J. L. Flanagan and L. R. Rabiner (Dowden, Hutchinson & Ross, Stroudsburg, PA), pp. 127–130.
Ladefoged, P. (**1993**). *A Course in Phonetics* (Harcourt Brace, Orlando, FL).
Mermelstein, P. (**1973**). ''Articulatory model for the study of speech production,'' J. Acoust. Soc. Am. **53**, 1070–1082.
Perkell, J., and Klatt, D. (Eds.). (**1986**). *Invariance and Variability in Speech Processes*. (Lawrence Erlbaum, Hillsdale, NJ).
Perkell, J. Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., *et al.* (**1997**). ''Speech motor control: Acoustic goals, saturation effects, auditory feedback, and internal models,'' Speech Commun. **22**, 227–250.
Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J. *et al.* (**2000**). ''A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss,'' J. Phonetics **28**, 233–272.
Pisoni, D. B. (**1981**). ''Some current theoretical issues in speech perception,'' Cognition **10**, 249–259.
Plaut, D. C., and Kello, C. T. (**1999**). ''The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach,'' in *The Emergence of Language*, edited by B. MacWhinney (Erlbaum, Mahweh, NJ), pp. 381–415.
Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (**1996**). ''Understanding normal and impaired word reading: Computational principles in quasi-regular domains,'' Psychol. Rev. **103**, 56–115.
Roweis, S. (**1999**). ''Data driven production models for speech processing,'' Ph.D. thesis, University of Toronto, Toronto.
Rubin, P., Baer, T., and Mermelstein, P. (**1981**). ''An articulatory synthesizer for perceptual research,'' J. Acoust. Soc. Am. **70**, 321–328.
Rumelhart, D. E., Durbin, R., Golden, R., and Chauvin, Y. (**1995**). ''Backpropagation: The basic theory,'' in *Backpropagation: Theory, Architectures, and Applications Developments in Connectionist Theory*, edited by Y. Chauvin and D. E. Rumelhart (Erlbaum, Hillsdale, NJ), pp. 1–34.
Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (**1986**). ''Learning representations by back-propagating errors,'' Nature (London) **323**, 533–536.
Shiga, Y., and King, S. (**2003**). ''Estimation of voice source and vocal tract characteristics based on multi-frame analysis.'' Paper presented at Eurospeech.
Wrench, A., and Hardcastle, W. (**2000**). ''A multichannel articulatory speech database and its application for automatic speech recognition,'' Paper presented at the Proceedings of the 5th Seminar on Speech Production.

Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners[a)]

Sarah Hargus Ferguson[b)]
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 30 December 2003; revised 8 July 2004; accepted 14 July 2004)

Several studies have shown that when a talker is instructed to speak as though talking to a hearing-impaired person, the resulting ''clear'' speech is significantly more intelligible than typical conversational speech. While variability among talkers during speech production is well known, only one study to date [Gagné *et al.*, J. Acad. Rehab. Audiol. **27**, 135–158 (1994)] has directly examined differences among talkers producing clear and conversational speech. Data from that study, which utilized ten talkers, suggested that talkers vary in the extent to which they improve their intelligibility by speaking clearly. Similar variability can be also seen in studies using smaller groups of talkers [e.g., Picheny, Durlach, and Braida, J. Speech Hear. Res. **28**, 96–103 (1985)]. In the current paper, clear and conversational speech materials were recorded from 41 male and female talkers aged 18 to 45 years. A listening experiment demonstrated that for normal-hearing listeners in noise, vowel intelligibility varied widely among the 41 talkers for both speaking styles, as did the magnitude of the speaking style effect. While female talkers showed a larger clear speech vowel intelligibility benefit than male talkers, neither talker age nor prior experience communicating with hearing-impaired listeners significantly affected the speaking style effect. [DOI: 10.1121/1.1788730]

PACS numbers: 43.70.Gr, 43.71.Bp, 43.72.Ew, 43.71.Es [RLD] Pages: 2365–2373

I. INTRODUCTION

Several studies have shown that when a talker is instructed to speak as though talking to a hearing-impaired person, the resulting ''clear'' speech is significantly more intelligible than ordinary conversational speech. In studies using sentence materials, a clear speech intelligibility benefit of 17 to 20 percentage points has been observed, both for hearing-impaired listeners and for normal-hearing listeners in noise (e.g., Picheny *et al.*, 1985; Payton *et al.*, 1994). Uchanski *et al.* (1996), using words excised from sentences, also showed similar clear speech intelligibility benefits for these two groups.

In contrast with earlier studies, Ferguson and Kewley-Port (2002), using vowels in words excised from clear and conversational sentences, found very different results for normal-hearing and hearing-impaired listeners. While normal-hearing listeners identifying vowels in noise enjoyed a 15-percentage-point clear speech intelligibility benefit, overall vowel intelligibility for elderly hearing-impaired listeners was practically identical in the two styles. In fact, the front vowels (/i, ɪ, e, ɛ, æ/) were actually *less* intelligible in clear speech than in conversational speech for these listeners. Based on acoustic and regression analyses, Ferguson and Kewley-Port (2002) attributed this surprising outcome to an interaction between the sloping sensorineural hearing losses of the elderly hearing-impaired listeners and one of the acoustic features of their single talker's clear speech. Specifically, the talker produced front vowels with significantly higher F2 values in clear speech than in conversational speech. Front vowel F2 values fall in a region where the hearing-impaired listeners had a sloping hearing loss (i.e., 2000–2500 Hz). As a result, Ferguson and Kewley-Port (2002) claimed, this ''raised front vowel F2'' clear speech strategy made the front vowels less audible, and thus less intelligible, than they were in conversational speech.

Ferguson and Kewley-Port's discovery of a clear speech acoustic change that *negatively* impacted intelligibility for hearing-impaired listeners led them to question whether the strategies used by their single talker were unique to him, or whether they represented a typical response to instructions to speak as though talking to a hearing-impaired listener (2002). In the current paper, an important step toward answering that question is described. To determine the range of clear speech intelligibility and acoustic effects that can be expected to occur in a population of untrained talkers, an extensive database of clear and conversational sentences was recorded from a large group of talkers ($n=41$). The pool of talkers was designed so that the effects of individual talker characteristics on the magnitude of any observed clear speech effects could be assessed. Specifically, talkers ranging in age from 18 to 45 were recorded, with roughly equal numbers of talkers from each gender. In addition, talkers were queried regarding prior experience communicating with hearing-impaired listeners, to determine whether talkers with such experience achieved larger clear speech effects than talkers without such experience. In the present experiment, the clear speech intelligibility effect was assessed by presenting vowels in /bVd/ context from all 41 talkers to normal-hearing listeners for identification in noise.

[a)]Portions of these data were presented at the 143rd meeting of the Acoustical Society of America [J. Acoust. Soc. Am. **111**, 2482 (2001)].

[b)]Current affiliation: Department of Speech–Language–Hearing: Sciences and Disorders, University of Kansas, Dole Center, 1000 Sunnyside Avenue, Room 3001, Lawrence, Kansas 66045. Electronic mail: safergus@ku.edu

Only one published study to date has directly assessed individual talker differences in clear and conversational speech, using monosyllabic and bisyllabic words spoken in isolation. Gagné *et al.* (1994) recorded these materials from ten female talkers and presented them to normal-hearing listeners in several modalities (auditory-only, visual-only, and audiovisual). The size of the clear speech intelligibility effect varied among the ten talkers for all three modalities. For example, for monosyllabic words in the auditory-only condition, the clear speech intelligibility effect ranged from −3 percentage points (clear speech *less* intelligible than conversational speech) to +24 percentage points. Similar variance in the clear speech intelligibility effect is also apparent in studies that did not directly assess these differences, such as Gagné *et al.* (1995; $n=6$), Schum (1996; $n=20$), and Gagné *et al.* (2002; $n=6$). The size of the clear-speech intelligibility effect can also be seen to vary among talkers in studies involving only two or three talkers (e.g., Picheny *et al.*, 1985; Bradlow and Bent, 2002; Bradlow *et al.*, 2003).

This variability among talkers calls into question a widely held assumption, implied in most studies of clear and conversational speech, that clear speech is a specific speaking style that talkers automatically produce when they are told to speak as though conversing with a hearing-impaired person. Examining the published data (e.g., Gagné *et al.*, 1994; Bradlow *et al.*, 2003), it appears that talkers respond in various ways to this instruction, with correspondingly varied results, both for intelligibility and in acoustic analyses. The goal of the current project was to begin to establish the extent to which clear-speech intelligibility effects vary within the population of normal talkers, focusing on vowels in words. It also sought to determine whether talker age, gender, or previous experience communicating with hearing-impaired listeners have any influence on clear speech intelligibility effects.

II. CLEAR AND CONVERSATIONAL SPEECH DATABASE

A. Talkers

A total of 41 talkers, all from the South Midland region of the United States (Labov *et al.*, in press), were recorded. Talkers reported normal hearing and denied any training in phonetics or diction, experience with professional speaking, or history of speech or language disorders. Five male and five female talkers were recruited in four age brackets: (1) 18 to 24 years, (2) 25 to 31 years, (3) 32 to 38 years, and (4) 39 to 45 years. One additional female talker was also recruited into the 18–24 age bracket. Talkers were paid for their participation.

B. Materials

Each talker recorded two lists of 188 sentences, one list in each speaking style condition (clear and conversational). These lists contained three types of sentences: (1) /bVd/ sentences, (2) consonant-vowel-consonant (CVC) sentences, and (3) CID sentences. The /bVd/ and CVC sentences were constructed by inserting test words quasirandomly into 16 neutral sentence frames (see the Appendix). For the /bVd/ sentences, the test words consisted of ten vowels (/i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ, u/) in /bVd/ context. Each /bVd/ word was placed in seven unique sentence frames, for a total of 70 /bVd/ sentences. These 70 sentences were used for both speaking style conditions. The test words for the CVC sentences included two 50-item lists selected from the Northwestern University Auditory Test No. 6 (NU-6; Tillman and Carhart, 1966), plus four additional CVC words (two per style) containing the vowel /ʊ/. Within each speaking style, each NU-6 word occurred twice, in two different neutral sentence frames, for a total of 104 CVC sentences per speaking style. Different sets of CVC sentences were thus recorded for the clear and conversational speaking styles. For the CID sentences, two sets of 14 sentences were selected from the Central Institute for the Deaf (CID) Everyday Sentences test (Davis and Silverman, 1978). As was the case for the CVC sentences, a different set of 14 CID sentences was used for each speaking style condition.

The /bVd/ and CVC sentences for each speaking style condition were combined, yielding a 174-sentence test list for each style. Four randomizations of each test list were prepared and assigned randomly to the 41 talkers. Each test list was divided into five blocks of 34 or 35 sentences; a sixth block contained the 14 CID sentences in a single order that was used for all talkers. For each speaking style condition, a test packet was created for each talker containing these six blocks in random order.

C. Recording procedures

Recordings were made in a single-wall, sound-attenuating booth using a Shure SM-10 headset microphone placed approximately 1 in. from the talker's lips. Microphone output was routed to a preamplifier (Shure M267), low-pass filtered (8500 Hz) using a Tucker–Davis Technologies (TDT) programmable filter (PF1), and digitized (22050 Hz sample rate) using a TDT analog-to-digital interface (DD1). During the recordings, the experimenter monitored the output of the preamplifier through headphones.

Recordings were completed in two sessions at least one day apart. In the first session, all talkers were instructed to read the test sentences as they would in everyday, normal conversation. The text of the instructions included the sentence, ''It is important that your speech be as much like your normal conversational style as possible.'' In the second session, they were instructed to say the test sentences as they would if they were talking to a hearing-impaired person. The text of the instructions included the sentence, ''It is important that you speak clearly, so that a hearing-impaired person would be able to understand you.'' Instructions were given orally and also printed on a card placed prominently in the recording booth. Talkers were also given a card containing 16 practice sentences representing the 16 neutral sentence frames.

Talkers read the practice sentences aloud two to three times before beginning the actual recordings. In the conversational speech session, talkers were given feedback on their manner of speaking during this rehearsal. This feedback typically consisted of comments from the experimenter regarding how conversational the speech sounded, along with in-

structions to repeat the practice list until the experimenter judged the speech to be sufficiently conversational. The purpose of this feedback was to help the talkers produce speech approximating their every day conversational speech despite the formal laboratory conditions. Neither such feedback nor any additional instructions about how to produce clear speech were given in the clear speech session, leaving the talker to decide what it means to ''speak clearly.'' Similarly nonspecific clear speech instructions were used by Schum (1996) and by Gagné and colleagues (Gagné *et al.*, 1994; Gagné *et al.*, 1995; Gagné *et al.*, 2002). This nonspecific approach also has clinical relevance. Family members of people with hearing loss are often provided with a list of suggestions to help improve communication; most such lists include a general instruction to speak clearly.

At the end of the second (clear speech) recording session, talkers were interviewed briefly. First, they were asked whether they had any experience communicating with people with hearing loss. If a talker answered affirmatively, he or she was then asked to describe the relationship with the hearing-impaired communication partner(s), the frequency with which the talker communicated with them, and the length of time over which the talker communicated with them.

III. INTELLIGIBILITY TEST METHODS

A. Listeners

Seven young adults (19–30 years) participated in this experiment. All listeners were from the South Midland region of the United States. All had normal hearing (thresholds $\leqslant$20 dB HL at octave intervals from 250–8000 Hz), as confirmed by pure-tone screening. Listeners were paid for their participation.

B. Materials

Vowel intelligibility in clear and conversational speech was assessed using 1640 /bVd/ words excised from the recorded sentence database: two tokens of the 10 /bVd/ words for each of the 41 talkers in each speaking style. The third and fourth of the seven repetitions of each /bVd/ word were arbitrarily chosen as stimuli. If one of these tokens contained extraneous noise or other distortion, or was judged by the experimenter to be a poor instance of the intended vowel, a subsequent repetition of the word was used instead. This occurred 33 times out of the 1640 stimuli. Amplitude differences among vowels and talkers and between the two speaking styles were eliminated by scaling all test words to a peak amplitude of 75 dB (amplitude *re*: 1 bit) using MATLAB (The MathWorks, Inc., 2000).

During the experiment, test stimuli were presented in a background of 12-talker babble. A 30-s sample of babble was low-pass filtered at 8500 Hz and digitized from the noise channel of a recording of the Speech Perception in Noise Test (Kalikow *et al.*, 1977), using digitization parameters identical to those employed during the talker recordings. On each test trial, a segment of babble was selected from a random location within this 30-s sample. The duration of the babble segment exceeded that of the test stimulus by 1000 ms such that the test item was centered temporally within the babble segment.

C. Procedures and apparatus

Listeners performed all testing individually in a single-wall sound treated room, seated in front of a computer monitor and keyboard. On each trial, a test word and a segment of babble were played out from separate channels of a 16-bit D/A converter (TDT DA1). The test word was attenuated (TDT PA4) to achieve the desired speech-to-babble (S/B) ratio. After the test word was attenuated, the test word and babble segment were mixed (TDT SM3) and routed to a programmable filter (TDT PF1) that low-pass filtered (8500 Hz) and attenuated the combined speech and babble so that the overall speech level was 70 dB SPL. The test word and babble segment were then delivered monaurally to the listener via TDH-39 supra-aural earphones. The listener identified the vowel of the test word by typing the number of the response category corresponding to that vowel. The ten response alternatives were displayed on the computer as ten sets of three keywords (selected to be suitable for both the current study and future experiments using other consonant environments, and to encourage listeners to respond based on the vowel sound and not on the spelling of the test word): (1) feet, thief, bead; (2) sit, rib, bid; (3) tape, raid, bade; (4) head, said, bed; (5) back, mass, bad; (6) pot, sod, bod; (7) cup, rug, bud; (8) rode, own, bode; (9) good, should, book; (10) rude, news, boot.

Listeners were tested in ten sessions, each lasting approximately 90 minutes. In the first session, listeners underwent a hearing screening and were familiarized with the test procedures prior to beginning the experimental conditions. For familiarization, a 41-trial block of clear vowel tokens was created using one /bVd/ token from each talker. After being oriented to the vowel identification task and response alternatives by the experimenter, listeners completed the familiarization block three times: once in quiet, once in babble at S/B$=-2$ dB, and once in babble at S/B$=-10$ dB. Listeners were given feedback only during the initial, quiet familiarization block. All listeners identified vowels with 90% or greater accuracy in quiet.

For the experimental conditions, the S/B ratio was -10 dB, an S/B ratio shown in previous studies (e.g., Ferguson and Kewley-Port, 2002) to prevent ceiling effects for the most intelligible vowels. The test stimuli were arranged into 48 blocks of 100 to 120 items. Each listener received the 48 test blocks in random order, and stimuli were randomized within each block for each listener. The test stimulus blocks were created by dividing the set of 1640 stimuli into a 2×2 factorial design of gender by speaking style. Each block contained 20 stimuli from 5 or 6 talkers of the same gender, from a single speaking style. That is, there were 12 blocks of clear stimuli from male talkers (MCL), 12 of conversational stimuli from male talkers (MCO), 12 of clear stimuli from female talkers (FCL), and 12 of conversational stimuli from female talkers (FCO). The test blocks were arranged so that each stimulus was presented three times, each time in a block containing a different set of talkers. That is, gender and

TABLE I. Percent correct vowel intelligibility for each talker in clear (CL) and conversational (CON) speech, and the percentage point difference between the two styles (DIFF). Mean values were calculated over all talkers.

Code	CL	CON	DIFF
F01	94.3	61.0	33.3
F02	83.1	78.3	4.8
F03	80.0	79.5	0.5
F04	73.8	75.7	−1.9
F05	76.2	59.8	16.4
F06	90.2	82.9	7.3
F07	83.3	71.0	12.3
F08	50.0	49.1	0.9
F09	87.9	72.4	15.5
F10	65.5	58.6	6.9
F11	92.1	64.5	27.6
F12	56.2	41.7	14.5
F13	75.2	50.5	24.7
F14	86.9	76.0	10.9
F15	77.6	81.0	−3.4
F16	87.1	82.6	4.5
F17	90.2	72.1	18.1
F18	66.4	38.6	27.8
F19	85.5	77.6	7.9
F20	57.6	57.9	−0.3
F21	80.7	74.3	6.4
M01	52.9	65.0	−12.1
M02	79.1	61.0	18.1
M03	76.7	58.6	18.1
M04	61.7	56.9	4.8
M05	83.3	72.6	10.7
M06	80.7	76.9	3.8
M07	28.8	25.0	3.8
M08	69.3	70.0	−0.7
M09	68.6	60.5	8.1
M10	65.7	62.1	3.6
M11	71.2	60.2	11.0
M12	67.4	59.8	7.6
M13	52.6	44.3	8.3
M14	61.9	52.9	9.0
M15	71.0	69.1	1.9
M16	68.3	71.7	−3.4
M17	79.3	79.3	0.0
M18	71.4	70.2	1.2
M19	75.0	71.0	4.0
M20	75.0	61.2	13.8
MEAN	73.2	64.7	8.5

speaking style combinations were blocked, with three different randomizations of the list of talkers divided into four groups of talkers.

IV. RESULTS

Each /bVd/ test word was identified 21 times (7 listeners × 3 presentations per listener). Overall vowel intelligibility for each talker in each speaking style was determined by averaging the percent correct scores of the 20 test words. These data, and the clear speech vowel intelligibility effect for each talker (determined by subtracting the conversational speech score from the clear speech score for each vowel), are shown in Table I. Prior to statistical analysis, percent correct scores for each of the 1640 test words were converted to rationalized arcsine units (RAUs; Studebaker, 1985). Individual vowel RAU scores for each speaking style and talker were determined by averaging across the RAU scores for the two tokens of each vowel. Clear and conversational intelligibility data for each vowel, averaged across the 41 talkers, are shown in Fig. 1. In this figure it is apparent that the signal-to-babble ratio employed here (−10 dB) was effective in eliminating any ceiling effects for individual vowels.

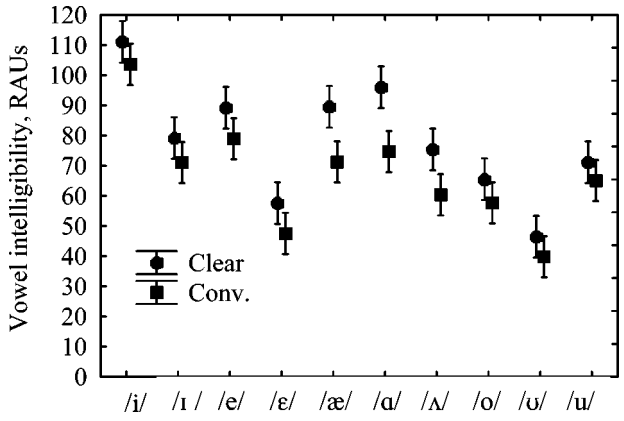


FIG. 1. Intelligibility of individual vowels in clear and conversational speech in RAUs, averaged across all 41 talkers. Error bars indicate 95% confidence intervals.

A. Talker and speaking style effects

For the first analysis, individual vowel RAU scores for each talker and speaking style were submitted to a two-way mixed-model analysis of variance (ANOVA) using Statistica (StatSoft, 2003). The main effect of speaking style was significant [$F(1,40)=38.807$, $p<0.001$]. Vowels in clear speech were more intelligible than vowels in conversational speech, with an overall clear speech vowel intelligibility benefit of 8.5 percentage points. The main effect of talker was also significant [$F(40,369)=4.0599$, $p<0.001$]. Vowel intelligibility varied widely among the talkers in both speaking styles, ranging from 29% to 94% in clear speech and from 25% to 83% in conversational speech.[1] The interaction between talker and speaking style was also significant [$F(40,369)=3.49$, $p<0.001$]. As is apparent in Table I, the difference between clear and conversational vowel intelligibility scores (i.e., the clear speech vowel intelligibility effect) varied widely among the talkers, ranging from 33 percentage points (talker F01) to −12 percentage points (talker M01).

The talker×speaking style interaction is graphically represented in Fig. 2, in which RAU vowel intelligibility scores are shown for the 41 talkers, who are ordered by their conversational vowel intelligibility scores. Comparing vowel intelligibility in clear speech with the ordered conversational scores, the size of the clear speech effect appears to be unrelated to conversational vowel intelligibility. There are talkers with large and small clear speech vowel intelligibility effects throughout much of the range of the conversational scores, though effects tend to be smaller for talkers with higher conversational vowel intelligibility (a possible ceiling effect). These impressions were confirmed by correlational analyses of each talker's average RAU clear and conversational vowel intelligibility scores, along with the RAU difference scores for each talker. Clear and conversational scores were highly correlated ($r=0.74$, $p<0.001$). While difference scores were slightly negatively correlated with

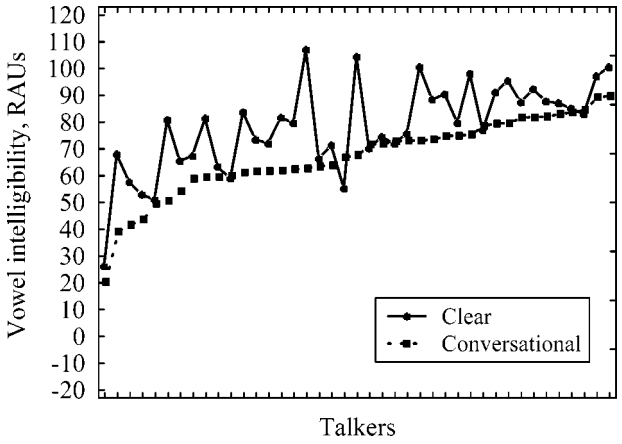


FIG. 2. Overall RAU vowel intelligibility in clear and conversational speech for individual talkers. Talkers are shown in order of ascending conversational vowel intelligibility.

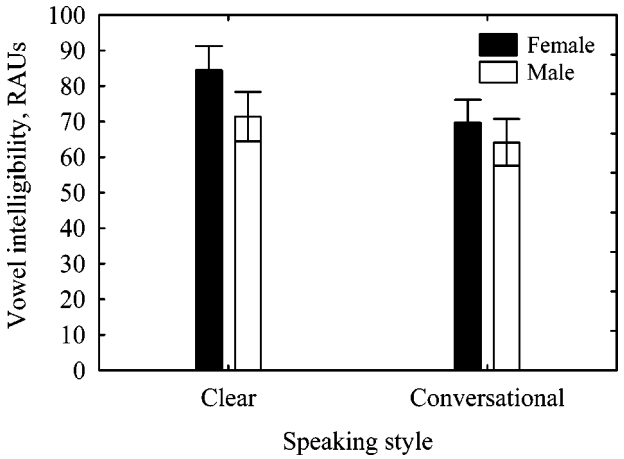


FIG. 4. Overall RAU vowel intelligibility in clear and conversational speech for male and female talkers.

conversational intelligibility ($r=-0.22$), this correlation was not significant ($p>0.05$). In contrast, the correlation between difference scores and clear intelligibility was significant and positive ($r=0.49$, $p<0.05$). Thus, talkers with better vowel intelligibility in clear speech tended to have larger clear-speech vowel intelligibility effects.

B. Age and gender effects

Recall that the talkers were recruited into four age brackets: (1) 18 to 24 years, (2) 25 to 31 years, (3) 32 to 38 years, and (4) 39 to 45 years. It was hypothesized that talkers in the higher age brackets would have more frequent contact with hearing-impaired people, and that this experience would make them more effective "clear speakers." To test this hypothesis, the effect of talker age was examined using a one-way ANOVA. Overall talker RAU difference scores were the dependent variable; these were computed for each talker by subtracting the overall conversational RAU vowel intelligibility score from the overall clear RAU vowel intelligibility score. Contrary to the hypothesis, the main effect of age bracket was not significant [$F(3,37)=0.21$, $p=0.89$]. Figure 3 shows RAU difference scores and 95% confidence intervals for each age bracket. While Fig. 3 seems to suggest a trend toward larger difference scores in the higher age brackets, the correlation between talker age and RAU difference score (calculated using actual ages rather than age brackets) was not significant ($r=0.17$, $p>0.05$).

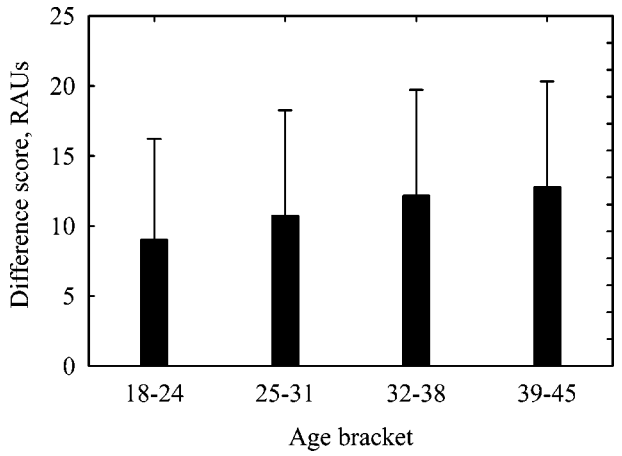


FIG. 3. Average clear-minus-conversational RAU difference scores for each age bracket. Error bars indicate 95% confidence intervals.

The effect of talker gender was then investigated. A two-way repeated-measures ANOVA was run with clear and conversational RAU overall vowel intelligibility scores as dependent variables, speaking style as a within-subject factor, and gender as a between-subject factor. As expected, the main effect of speaking style [$F(1,39)=41.82$, $p<0.001$] was still significant. The main effect of talker gender was also significant [$F(1,39)=4.53$, $p<0.04$], as was the talker gender×speaking style interaction [$F(1,39)=4.82$, $p<0.04$]. This interaction is illustrated in Fig. 4. Univariate tests revealed that while women had significantly greater vowel intelligibility than men in the clear speech condition [$F(1,39)=7.37$, $p<0.01$], vowel intelligibility for men and women did not differ significantly when speaking conversationally [$F(1,39)=1.47$, $p=0.23$]. In addition, a one-way ANOVA with overall talker RAU difference scores as the dependent variable showed that the magnitude of the clear speech vowel intelligibility effect was significantly larger for female talkers than for male talkers. The average difference score for female talkers was 11 percentage points, while for male talkers the average difference score was 5.6 percentage points.

C. Experience communicating with hearing-impaired listeners

Based on their responses in the post-recording interview, talkers were placed into one of four categories of experience communicating with hearing-impaired listeners: none (no prior experience; $n=14$), little (only a few experiences; $n=8$), occasional (a relative or friend with hearing loss, but less than one interaction per week; $n=9$), or frequent (at least weekly contact with one or more hearing-impaired people; $n=10$). Mean RAU difference scores and 95% confidence intervals for talkers in each experience category are shown in Fig. 5. It had been hypothesized that talkers with more experience communicating with hearing-impaired people would have larger clear speech vowel intelligibility

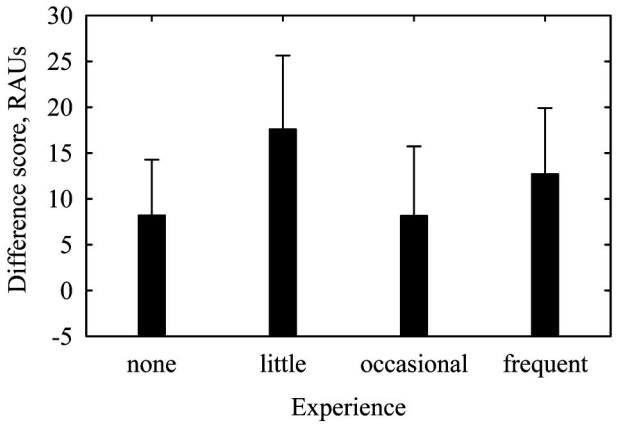


FIG. 5. Clear-minus-conversational RAU difference scores averaged across talkers in four categories of experience communicating with hearing-impaired listeners. Error bars indicate 95% confidence intervals.

effects than talkers with less experience. Nonetheless, the main effect of experience was not significant in a one-way ANOVA on the overall talker difference scores [$F(3,37) = 1.48$, $p = 0.24$]. This suggests that the amount of experience a talker has communicating with people with hearing loss has no bearing on the extent to which a talker is able to improve his or her vowel intelligibility by speaking clearly.

V. DISCUSSION

A. Vowels in clear and conversational speech

Averaged across talkers, vowel intelligibility for normal-hearing listeners in this experiment was significantly higher in clear speech than in conversational speech. This result is consistent with that observed by Ferguson and Kewley-Port (2002), who also excised vowels in /bVd/ context from clear and conversational sentences and presented them to normal-hearing listeners for identification. Taken together, these two studies demonstrate that vowels presented in noise to normal =hearing listeners are generally more intelligible when talkers are given general instructions to speak clearly than when talkers are instructed to speak in a typical conversational style.

As noted above, this experiment was motivated in part by curiosity about whether the clear speech production strategies employed by the one talker in Ferguson and Kewley-Port (2002) were typical. A direct answer to this question requires acoustic analysis, and so cannot be provided at this time. It may be informative, however, to compare the size of the clear speech vowel intelligibility advantage observed in Ferguson and Kewley-Port (2002) with the current data. The average clear speech vowel intelligibility benefit for normal-hearing listeners for the single talker in Ferguson and Kewley-Port (2002) was 15 percentage points, somewhat higher than the 8.5-percentage-point average benefit observed here for 41 talkers. However, the clear speech vowel intelligibility effects in the current experiment ranged widely, from −12 to 33 percentage points. Furthermore, nine of the 41 talkers had larger clear speech vowel intelligibility benefits for normal-hearing listeners than the talker in Ferguson and Kewley-Port (2002). This indicates that although the talker in Ferguson and Kewley-Port (2002) was better than average at making his vowels more intelligible in clear speech, he was not an extreme case. Acoustic analyses of the current database will help to determine how common his clear speech production strategies were.

B. Talker differences in clear and conversational speech

This experiment demonstrated that vowel intelligibility for normal-hearing listeners in noise varies widely among individual talkers in both conversational and clear speech. In conversational speech, vowel intelligibility scores for the 41 talkers ranged from 25% to 83%; in clear speech the range was from 29% to 94%. The intelligibility difference between clear and conversational speech, or the clear speech vowel intelligibility effect, also differed greatly among the talkers, with a range of −12 to 33 percentage points. If a somewhat lax criterion of 5 percentage points is selected as corresponding to clinical significance, only 23 of the 41 talkers (56%) showed a significant clear speech vowel intelligibility benefit. The remaining 18 talkers showed clear-versus-conversational differences of less than 5 percentage points, with one talker showing a substantial *negative* clear speech vowel effect, i.e., vowels *less* intelligible in clear speech than in conversational speech, of −12 percentage points. Thus, it appears that in a population of normal talkers, slightly more than half would be expected to show significantly improved vowel intelligibility in clear speech, and a few would show a substantial loss of intelligibility.

The talker variability shown in the current experiment is consistent with earlier studies on clear and conversational speech that either directly studied talker differences (Gagné *et al.*, 1994) or showed individual data for a relatively large group of talkers (Schum, 1996). Detailed comparisons of the current data with these previous studies are complicated, however, by methodological differences. For example, although Gagné *et al.* (1994) used normal-hearing subjects, intelligibility was assessed for whole words rather than for words varying only in their vowel nuclei. Schum's study (1996) differed from the current experiment not only in terms of speech materials (sentences) but also in terms of the listener population (elderly hearing-impaired listeners). Nonetheless, there is an interesting parallel between Schum (1996) and the current experiment. In both cases, the range of speaking style effects was 45 percentage points: from −12 to 33 percentage points here, and from approximately 5 to approximately 50 percentage points in Schum (1996). Thus, despite methodological differences, one can conclude that the intelligibility improvement achieved by talkers when they are instructed to speak clearly varies widely among normal talkers.

C. Effects of talker gender

While vowel intelligibility was similar for male and female talkers in conversational speech (Fig. 4), differences were observed in clear speech. Specifically, the female talkers had significantly higher vowel intelligibility in clear speech than the male talkers (78% vs 68%, respectively), as

well as a significantly larger clear speech effect (11 vs 5.6 percentage points). A similar pattern, albeit for a very limited set of talkers, was observed in Bradlow and Bent (2002) and in Bradlow *et al.* (2003): the two talkers had comparable sentence intelligibility in conversational speech, but the female talker showed a significantly greater clear speech intelligibility benefit than the male talker. No other previous study of clear and conversational speech has compared male and female talkers.

Among studies of talker differences using only a single speaking style (e.g., Cox *et al.*, 1987), most have not examined gender effects. Of the two studies known to us that have examined talker gender, results are conflicting. Gengel and Kupperman (1980) observed no gender differences in word intelligibility for normal-hearing listeners in noise. However, that study included only six talkers, making the data difficult to generalize. In contrast, a much larger study (with 20 talkers) by Bradlow *et al.* (1996) found that for sentences identified by normal-hearing listeners in quiet, female talkers were more intelligible than male talkers. This is consistent with the current clear speech data. As Bradlow *et al.* (1996) noted, the tendency for greater intelligibility in female talkers agrees well with phonetic data indicating that women produce more ''careful'' speech than men. Byrd (1994) found that male talkers in the TIMIT database (which contains speech from 630 talkers) used a faster speaking rate and released final stop consonants less often than female talkers.

The lack of a gender effect in conversational speech in the current vowel intelligibility data and in the recent Bradlow papers (Bradlow and Bent, 2002; Bradlow *et al.*, 2003) suggests that when speaking conversationally, women and men produce speech of comparable intelligibility. In contrast, when speaking more formally, as in a lab setting without specific instructions to be conversational (as in Bradlow *et al.*, 1996, and in Byrd, 1994) or with instructions to speak clearly (as in the current experiment, in Bradlow and Bent, 2002, and in Bradlow *et al.*, 2003), it appears that women tend to be more careful in their speech production than men.

D. Effects of age and prior experience

The amount of clear speech vowel intelligibility benefit achieved by different talkers in the current experiment was unrelated to either talker age or prior experience communicating with hearing-impaired listeners. It had been hypothesized that talkers in the higher age brackets (those aged 32 to 45 years) would have more frequent contact with hearing-impaired people than younger talkers, and that this experience would make them more effective ''clear speakers.'' The second part of this hypothesis was tested using an ANOVA, as described above. To test the first part of the hypothesis, a Spearman rank-order correlation was performed. The correlation between age bracket and experience was very low ($r = 0.1$), indicating that there was no relationship between talker age and prior experience. It is possible that a stronger relationship between age and experience would emerge if the sample were expanded to include talkers from a larger age range than the one used here, specifically elderly talkers. However, data from Schum (1996) suggest that any age differences in the clear speech effect would still be minimal. Schum's 20 talkers included 10 young and 10 elderly talkers, and the clear speech effect did not differ significantly between these two groups.

The fact that prior experience had no bearing on the clear speech effect was surprising, as it runs counter to a general expectation that people who communicate with hearing-impaired people on a regular basis would be ''better'' at speaking clearly (i.e., achieve a larger intelligibility benefit) than people without such experience. This expectation has as its basis an assumption that when communication breakdowns occur, talkers automatically adopt strategies that will make their speech more intelligible. This assumption is a key provision of Lindblom's H&H Theory (hyper- versus hypospeech; Lindblom, 1990), which states that talkers adjust their speech production to the immediate needs of their listener. The lack of any effect of prior experience on the clear speech vowel intelligibility effect, however, can be interpreted as evidence against this notion. It is, of course, possible that some talkers might improve their overall clear speech intelligibility without actually improving intelligibility for vowels, and that this overall intelligibility would be affected by prior experience. This possibility could be explored further using the NU-6 materials and CID sentences that were recorded along with the current vowel stimuli.

VI. CONCLUSIONS

In a listening experiment, vowel intelligibility in clear and conversational speech for normal-hearing listeners in noise varied widely among 41 talkers, as did the magnitude of the clear speech vowel intelligibility benefit. The size of this clear speech effect was apparently unrelated to talker age or experience communicating with hearing-impaired listeners. Although vowel intelligibility in conversational speech was similar for male and female talkers, the female talkers showed a significantly greater clear speech vowel intelligibility benefit than the men.

While the current experiment explored clear speech talker differences for vowel intelligibility only, data from previous studies using multiple talkers (Gagné *et al.*, 1994; Schum, 1996) suggest that variance in the clear speech intelligibility effect similar to that observed here for vowels would also be obtained for more meaningful materials. The fact that talkers told to speak clearly vary so much in their effectiveness has important clinical implications. Family members of people with hearing loss are often provided with a list of strategies to help improve communication. These strategies include avoiding background noise and making sure the hearing-impaired listener can see the talker's face. Communication partners are also given general instructions to speak clearly, similar to those given to the subjects in the current experiment. As shown here and by earlier multitalker studies using similar general instructions (i.e., Gagne *et al.*, 1994), not all talkers actually achieve improved speech intelligibility when given such general instructions. The current study also showed that even after extensive experience communicating with hearing-impaired listeners, talkers do not necessarily learn to produce clear speech that results in improved vowel intelligibility. These results suggest that certain

talkers might need more specific instructions on how to improve their intelligibility than others.

Surveying the research literature on the acoustic characteristics that underlie the superior intelligibility of clear speech, however, it is difficult to determine what to include in these specific instructions. While Picheny *et al.* (1986) described numerous clear speech acoustic characteristics, the question of exactly which characteristics enhance intelligibility, or ''What makes clear speech clear?'' remains open. To date, studies taking a ''signal-processing approach'' to this question have determined that slowing conversational speech does not by itself yield the kinds of intelligibility improvements typically found in naturally produced clear speech (e.g., Uchanski *et al.*, 1996), and that while increasing consonant-vowel ratio improves consonant intelligibility, increasing consonant duration does not (Gordon-Salant, 1986). Another approach to the question is to relate observed intelligibility differences in natural speech to observed acoustic differences. Bradlow *et al.* (2003) took such a ''talker-differences approach'' when they discussed the differing clear speech intelligibility effects produced by their two talkers in terms of the clear speech acoustic changes made by each talker. For example, they reported that the talker with the largest clear speech intelligibility benefit showed much more slowing in clear speech than the other talker, suggesting (in contrast with earlier studies) that reduced speaking rate may in fact play a role in improving intelligibility.

The database described here was designed with this talker-differences approach in mind, and acoustic analyses of vowel stimuli from the current experiment are ongoing. Correlating the current intelligibility data with acoustic data from all 41 talkers would go a long way toward answering the question ''What makes clear speech clear?'' by revealing which clear speech characteristics lead to improvements in vowel intelligibility for normal-hearing listeners in noise. A much more important analysis, however, would involve the acquisition of intelligibility data from the listener population for whom clear speech is actually intended: hearing-impaired listeners. The results of Ferguson and Kewley-Port (2002) suggest that talkers who produced a significant clear speech vowel intelligibility benefit for normal-hearing listeners might not necessarily produce significant clear speech improvements for listeners who are hearing=impaired. It may well be the case that certain clear speech acoustic characteristics that enhance intelligibility for normal-hearing listeners may have no effect on or be detrimental to intelligibility for hearing-impaired listeners. Extending the current experiment to listeners with hearing loss is therefore an important future research need.

The database was also designed to assist in the exploration of acoustic cues underlying clear speech intelligibility effects for materials other than vowels. In addition to the /bVd/ words used in the current experiment, both meaningful sentences and monosyllabic words in neutral sentence context were recorded. Intelligibility data for these materials would reveal the extent to which the clear speech intelligibility effect varies within a population of talkers for meaningful stimuli, a crucial next step in this line of research. Such data would also shed light on the extent to which vowel intelligibility predicts overall speech intelligibility for normal talkers. Subsequent acoustic analyses of the monosyllabic words would help identify the clear speech characteristics that are relevant for the variety of consonantal contexts included in these materials, while sentence materials could be used to examine possible relationships between the more global prosodic features of clear speech and features that occur at the phoneme level.

ACKNOWLEDGMENTS

This research was conducted as part of a doctoral dissertation completed under the guidance of Diane Kewley-Port and supported by NIHDCD-02229. The author gratefully acknowledges the contributions of Charles S. Watson, Larry E. Humes, David B. Pisoni, and Kenneth de Jong, as well as the helpful comments of Allard Jongman, Hugh Catts, Michael Hammer, Randy Diehl, and two anonymous reviewers.

APPENDIX: SENTENCE FRAMES

Neutral sentence frames used to construct sentences for recording.

1. Vera put the ____ on the table.
2. His mother calls him ____ at home.
3. Jean bought a ____ at the store.
4. They spelled the word ____ the wrong way.
5. Say the word ____ into the microphone.
6. He picked up the ____ and put it away.
7. He looked for a ____ but couldn't find one.
8. Look up the word ____ on the Internet.
9. I think the word ____ is hard for kids to say.
10. Her friends called her ____ back in school.
11. Please put the ____ next to the skis.
12. You might find a ____ in the garage.
13. Write the word ____ on the chalkboard.
14. Use the word ____ in a sentence.
15. I looked up the word ____ in the dictionary.
16. He said ____ but he meant box.

[1]In contrast, overall percent correct scores for individual listeners, averaged across the 41 talkers, varied over a considerably smaller range: from 66% to 77% in clear speech and from 60% to 72% in conversational speech.

Bradlow, A. R., and Bent, T. (**2002**). ''The clear speech effect for non-native listeners,'' J. Acoust. Soc. Am. **112**, 271–283.

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). ''Speaking clearly for children with learning disabilities: Sentence perception in noise,'' J. Speech Hear. Res. **46**, 80–97.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). ''Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics,'' Speech Commun. **20**, 255–272.

Byrd, D. (**1994**). ''Relations of sex and dialect to reduction,'' Speech Commun. **15**, 39–54.

Cox, R. M., Alexander, G. C., and Gilmore, C. (**1987**). ''Intelligibility of average talkers in typical listening environments.'' J. Acoust. Soc. Am. **81**, 1598–1608.

Davis, H., and Silverman, S. R. (**1978**). *Hearing and Deafness*, 4th ed. (Holt, Rinehart, and Winston, New York).

Ferguson, S. H., and Kewley-Port, D. (**2002**). ''Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **112**, 259–271.

Gagné, J.-P., Masterson, V. M., Munhall, K. G., Bilida, N., and Querengesser, C. (**1994**). ''Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech,'' J. Acad. Rehab. Audiol. **27**, 135–158.

Gagné, J.-P., Querengesser, C., Folkeard, P., Munhall, K. G., and Masterson, V. M. (**1995**). ''Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: An investigation of conversational and clear speech,'' Volta Rev. **97**, 33–51.

Gagné, J.-P., Rochette, A.-J., and Charest, M. (**2002**). ''Auditory, visual, and audiovisual clear speech,'' Speech Commun. **37**, 213–230.

Gengel, R. W., and Kupperman, G. L. (**1980**). ''Word discrimination in noise: Effect of different speakers,'' Ear Hear. **1**, 156–160.

Gordon-Salant, S. (**1986**). ''Recognition of natural and time-altered CVs by young and elderly subjects with normal hearing,'' J. Acoust. Soc. Am. **80**, 1599–1607.

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.

Labov, W., Ash, S., and Boberg, C. (**in press**). *Atlas of North American English: Phonetics, Phonology and Sound Change*. (Mouton/DeGruyter, Berlin).

Lindblom, B. (**1990**). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). (Kluwer Academic, Netherlands.)

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). ''Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **95**, 1581–1592.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). ''Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech,'' J. Speech Hear. Res. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). ''Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech,'' J. Speech Hear. Res. **29**, 434–446.

Schum, D. J. (**1996**). ''Intelligibility of clear and conversational speech of young and elderly talkers,'' J. Amer. Acad. Audiol. **7**, 212–218.

StatSoft, Inc. (**2003**). STATISTICA (data analysis software system), version 6. http://www.statsoft.com.

Studebaker, G. A. (**1985**). ''A ''rationalized'' arcsine transform,'' J. Speech Hear. Res. **28**, 455–62.

The MathWorks, Inc. (**2000**). MATLAB (The MathWorks, Inc., Natick, MA).

Tillman, T., and Carhart, R. C. (**1966**). ''An expanded test for speech discrimination utilizing CNC monosyllabic words: N.U. Auditory Test No. 6,'' USAF School of Aerospace Medicine Report No. SAM-TR-66-55.

Uchanski, R. M., Choi, S. S., Braida, L. D., and Durlach, N. I. (**1996**). ''Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate,'' J. Speech Hear. Res. **39**, 494–509.

Clear speech perception in acoustic and electric hearing[a)]

Sheng Liu[b)]
Hearing and Speech Research Laboratory, Department of Biomedical Engineering, University of California, Irvine

Elsa Del Rio
Hearing and Speech Research Laboratory, Department of Biology, University of California, Irvine

Ann R. Bradlow
Phonetics Laboratory, Department of Linguistics, Northwestern University, Evanston, Illinois

Fan-Gang Zeng[b)]
Hearing and Speech Research Laboratory, Departments of Otolaryngology, Biomedical Engineering and Cognitive Sciences, University of California, Irvine

(Received 22 May 2003; revised 17 June 2004; accepted 6 July 2004)

When instructed to speak clearly for people with hearing loss, a talker can effectively enhance the intelligibility of his/her speech by producing ''clear'' speech. We analyzed global acoustic properties of clear and conversational speech from two talkers and measured their speech intelligibility over a wide range of signal-to-noise ratios in acoustic and electric hearing. Consistent with previous studies, we found that clear speech had a slower overall rate, higher temporal amplitude modulations, and also produced higher intelligibility than conversational speech. To delineate the role of temporal amplitude modulations in clear speech, we extracted the temporal envelope from a number of frequency bands and replaced speech fine-structure with noise fine-structure to simulate cochlear implants. Although both simulated and actual cochlear-implant listeners required higher signal-to-noise ratios to achieve normal performance, a 3–4 dB difference in speech reception threshold was preserved between clear and conversational speech for all experimental conditions. These results suggest that while temporal fine structure is important for speech recognition in noise in general, the temporal envelope carries acoustic cues that contribute to the clear speech intelligibility advantage. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1787528]

PACS numbers: 43.71.Es, 43.71.Gv, 43.71.Ky [PFA] Pages: 2374–2383

I. INTRODUCTION

Speech recognition in noise is an extremely challenging task, particularly for hearing-impaired listeners. One way to alleviate this difficulty is to speak ''clearly'' as opposed to ''conversationally'' to these individuals. Previous studies have reported an advantage of 10–20 percentage points in intelligibility for clear speech over conversational speech for a range of speech materials and listening conditions and for a variety of listener populations. Speech stimuli for which a clear speech intelligibility advantage has been reported include isolated syllables (Chen, 1980; Gagne *et al.*, 2002), words (Gagne *et al.*, 1994), nonsense sentences (Picheny *et al.*, 1985; Payton *et al.*, 1994; Uchanski *et al.*, 1996; Helfer, 1997; Krause, 2001) and meaningful sentences (Bradlow and Bent, 2002; Bradlow *et al.*, 2003). Listening conditions include white noise (Chen *et al.*, 1980; Gagne *et al.*, 1995; Uchanski *et al.*, 1996; Bradlow and Bent, 2002; Gagne *et al.*, 2002; Bradlow *et al.*, 2003), speech-spectrum-shaped noise (Krause, 2001; Krause and Braida, 2002), cafeteria noise (Schum, 1996), multitalker babble (Helfer, 1997; 1998; Ferguson and Kewley-Port, 2002), and reverberation (Payton *et al.*, 1994).

Listeners who demonstrated a benefit from clear speech include normal-hearing young adults (Chen *et al.*, 1980; Gagne *et al.*, 1994; 1995; 2002; Uchanski *et al.*, 1996; Helfer, 1997; Krause, 2001; Krause and Braida, 2002; Bradlow and Bent 2002), elderly adults (Payton *et al.*, 1994; Helfer, 1998; Ferguson and Kewley-Port, 2002), hearing-impaired adults (Picheny *et al.*, 1985; Uchanski *et al.*, 1996; Krause, 2001; Ferguson and Kewley-Port, 2002; Krause and Braida, 2002), children with and without learning disabilities (Bradlow *et al.*, 2003), and non-native listeners with normal hearing although these last three listener groups showed smaller benefits from clear speech than other listener populations (Bradlow and Bent, 2002; Bradlow and Pisoni 1999).

The intelligibility difference between clear and conversational speech is related to specific acoustic-phonetic characteristics (Picheny *et al.*, 1986; 1989; Moon and Lindblom, 1994; Bond and Moore, 1994; Bradlow *et al.*, 1996; Uchanski *et al.*, 1996). In comparison to conversational speech, clear speech has a generally slower speaking rate (Picheny *et al.*, 1985; Moon and Lindblom, 1994; Bradlow *et al.*, 1996) and larger temporal envelope fluctuations, i.e., greater temporal amplitude modulations (Payton *et al.* 1994; Krause and Braida, 2002; Krause and Braida, 2004). Several studies

[a)]Portions of this work were presented at the 26th Midwinter Meeting of the Association for Research in Otolaryngology, Daytona Beach, Florida, 2003.

[b)]Corresponding authors: University of California, 364 Med Surge II, Irvine, CA 92697; electronic mail: fzeng@uci.edu; sliu@uci.edu

have systematically analyzed the role of the decreased speaking rate in producing the clear speech intelligibility advantage (Picheny *et al.*, 1989; Uchanski *et al.*, 1996; Krause, 2001; Krause and Braida, 2002). Picheny *et al.* (1989) used digital signal processing to uniformly increase the clear speech rate or decrease the conversational speech rate in the range of 100 to 200 wpm without changing the voice pitch and found that both manipulations significantly degraded speech intelligibility. They attributed the degraded intelligibility to signal processing artifacts because using the same digital signal processing techniques to restore the processed sentence to the original rate did not restore the original intelligibility. Uchanski *et al.* (1996) used a nonuniform time-scaling method to change the phonetic segments within a sentence to reflect the previously measured segmental durational differences between clear and conversational speech. Although this method was generally less harmful to intelligibility than the uniform-scaling method, Uchanski *et al.* found that the normal-rate speech produced by nonuniformly altering segment duration of the original slow-rate clear speech had lower intelligibility than the unprocessed conversational speech. Taking a different approach, Krause and Braida (2002) instructed talkers to produce natural clear and conversational speech at various rates and were able to demonstrate the clear speech advantage even at the fast speaking rate. They concluded that the slow speaking rate in clear speech is not necessary for maintaining the high intelligibility.

Increasing the speaking rate also altered other inherent acoustic properties, one of which was the temporal amplitude modulation index. Payton *et al.* (1994) found that clear speech at a slow speaking rate has greater temporal amplitude modulations than conversational speech. It was unknown whether these greater temporal amplitude modulations were a result of the slower rate in the clear speech. However, Krause and Braida (2004) found greater temporal amplitude modulations in naturally produced clear speech at a normal speaking rate, suggesting that these greater temporal modulations do not have to be associated with a change in the speaking rate and may directly contribute to the clear speech intelligibility advantage.

Other evidence suggests that temporal amplitude modulation may play an important role in speech intelligibility in general. First, studies have shown that speech remains intelligible when the temporal fine structure is removed but the temporal amplitude modulations are preserved in a small number of broad frequency bands. The preserved temporal modulations were used to modulate band-limited white noise or biphasic impulses and delivered to normal-hearing or cochlear-implant subjects. Both groups of the subjects were able to achieve a high level of speech perception at least in quiet (Van Tasell *et al.*, 1987; Wilson *et al.*, 1991; Rosen, 1992; Shannon *et al.*, 1995; Dorman *et al.*, 1997). Second, the temporal amplitude modulation index has been used to predict speech intelligibility in terms of the Speech Transmission Index or STI (Houtgast and Steeneken, 1985; Payton and Braida, 1999). The speech-based STI has been shown to be an accurate predictor of intelligibility and used to predict the clear speech intelligibility advantage (Krause and Braida, 2004), suggesting that physical differences in temporal amplitude modulation exist between clear and conversational speech. Third, several studies have shown that speech intelligibility is reduced when temporal amplitude modulations are decreased. The modulations were digitally decreased by compressing the amplitude of peaks and/or expanding the amplitude of troughs of the extracted temporal envelope, or naturally reduced when speech signals passed through auditory systems of patients with auditory neuropathy, a disorder associated with impaired processing of amplitude modulations. Either the digital or natural means of decreasing temporal amplitude modulations degraded speech intelligibility (e.g., Hou and Pavlovic, 1994; Noordhoek and Drullman, 1997; Zeng *et al.*, 1999b)

To extend previous studies of the perception and acoustic analysis of clear speech, this study used a different group of subjects (cochlear-implant subjects) and different speech processing strategies. The first goal of this study was to measure clear and conversational speech perception as a function of signal-to-noise ratio in order to derive psychometric functions that could provide a complete characterization of the clear speech advantage over a range of signal-to-noise ratios. We used measures such as the speech reception threshold (Dirks *et al.*, 1982) and speech dynamic range (Zeng *et al.*, 2002) to quantify the clear speech advantage. The second goal was to measure the relative contributions of temporal envelope and fine structure to the clear speech advantage. We extracted the temporal envelope from a number of frequency bands and replaced speech fine-structure with noise fine-structure. We conducted four experiments to achieve these goals. Experiment I measured clear and conversational speech perception as a function of signal-to-noise ratio in normal-hearing listeners. Experiment II measured the speech recognition performance in quiet with the speech processed to contain temporal envelope cues in 2, 4, 8, or 16 frequency bands. Experiment III measured the speech recognition performance with an eight-band processor as a function of signal-noise-ratio in normal-hearing listeners. Finally, Experiment IV measured the performance as a function of signal-to-noise ratio in cochlear-implant listeners.

II. METHODS

A. Subjects

Twenty-seven normal-hearing listeners were recruited from the Undergraduate Social Science Subject Pool at the University of California, Irvine. Eleven subjects (divided into two groups) participated in Experiment I, five in Experiment II, and eleven in Experiment III. To evaluate the effect of individual difference in clear speech production across talkers, Experiment I consisted of five subjects tested on sentences produced by a female talker and six subjects tested on sentences produced by a male talker. Experiments II and III only presented speech material produced by the female talker. More subjects were tested in Experiment III in the case of the female talker because of the greater variability in their performance. None of the subjects reported any speech and/or hearing impairment. All were native English speakers and received course credit for their participation.

TABLE I. Biographical and audiological information for cochlear implant subjects.

Subject	Age (yrs)	Onset age	Deaf dur. (yrs)	Etiology	CI use (yrs)	Device	Strategy
S1	67	46	17	unknown	2	CII	CIS
S2	49	4	0	unknown	4	CI	MPS
S3	70	40	2	unknown	4	CI	CIS
S4	68	63	1	sudden	4	N24	ACE
S5	40	18	3	otosclerosis	1	N24	SPEAK
S6	45	35	0	trauma	10	N22	SPEAK
S7	25	23	unknown	unknown	3	Med-El	CIS+
S8	39	9	28	unknown	2	Med-El	CIS

Eight cochlear-implant subjects participated in Experiment IV. Experiment IV only presented speech material produced by the female talker. Table I details the implant subjects' biographical and audiological information. These subjects were 25–70 years old and all were post-lingually deafened with 1–10 years of implant use. Three used the Clarion device, another three used the Nucleus device, and the remaining two used the Med-El device. All used envelope-based strategies including continuous interleaved sampling (CIS), advanced combined encoder (ACE), multiple pulsatile stimulation (MPS), and spectral peak (SPEAK).

B. Stimuli

Stimuli consisted of a total of 144 sentences recorded in clear and conversational styles. These sentences were modified from the original Bamford-Kowl-Bench (BKB) sentences used for British children (Bench and Bamford, 1979). A male and a female adult talker recorded these sentences with a sampling rate of 16 kHz in a sound-treated room in the Phonetics Laboratory of the Department of Linguistics at Northwestern University, Evanston, Illinois (Bradlow *et al.*, 2003). Except for the four original lists consisting of a total of 64 sentences from the female talker, the remaining sentences from the female talker and all the sentences from the male talker were processed in the Hearing and Speech Research Laboratory at University of California, Irvine. The breath noise in the original recordings was removed by a 150-Hz, tenth-order Butterworth high-pass filter (Cool Edit Pro™ 2.0). All sentences were normalized to have the same long-term rms level and then stored in a Microsoft Windows PCM wav file. A total of 144 sentences was processed to result in 18 lists each consisting of 8 sentences. The sentences in each list were either clear or conversational speech.

In Experiment I, the original sentences were individually mixed with a speech-spectrum shaped noise at signal-to-noise ratios from −20 to 20 dB in 5 dB steps. The speech-spectrum shaped noise was produced independently for the female and male talkers by filtering white noise with a tenth-order linear predictive coding (LPC) spectral envelope derived from combined clear and conversational speech sentences from each talker. Sentences from the female and male talkers were presented to two groups of listeners, five listeners in the female condition and six in the male condition. In the remaining experiments, only the female sentences were used. In Experiment II, the original sentences were processed to preserve the temporal envelope cues (Fig. 1, see also Shannon *et al.*, 1995). The stimuli were first divided into several spectral bands (2, 4, 8, or 16) via band-pass filters with their cut-off frequencies calculated from the Greenwood map which purportedly maps each equally distanced cochlear partition into a corresponding physical frequency range (Greenwood, 1990). The output of each band-pass filter was then full-wave rectified and low-pass filtered at 400 Hz to extract the temporal envelope. The envelope was multiplied by a white noise, and then band-pass filtered again by a filter that was identical to the analysis filter in the first stage. The filtered band-limited signals were finally summed to form a synthesized signal that contained the original sentence's temporal envelope but no fine structure cues. In Experiment III, the same original sentences in quiet and noise as in Experiment I were processed via an eight-band processor to test speech perception in noise. In Experiment IV, the same stimuli as in Experiment I were used for cochlear implant users.

C. Speech analysis

Due to a computer memory limitation, only 64 of the 144 sentences were concatenated to produce a long running

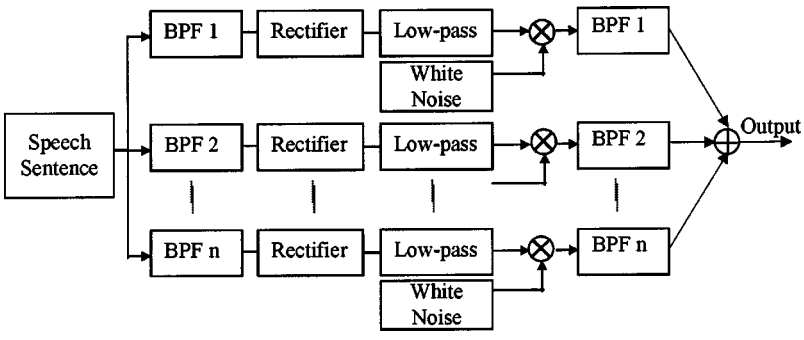


FIG. 1. Digital signal processing used to process original sentences to simulate cochlear implants in experiments II and III.

speech file for each of the four conditions, namely, the female clear and conversational speech, and the male clear and conversational speech. A 1/3-octave modulation spectrum was calculated as a function of modulation frequency for these concatenated sentences (Steeneken and Houtgast, 1980; Payton and Braida, 1999; Krause, 2001; Krause and Braida, 2004). The modulation spectrum was derived by first filtering the concatenated sentences into octave bands centered at 125 to 4000 Hz in octave steps, then squaring the band-limited signals, and finally low-pass filtering at 60 Hz to obtain the slow-varying intensity envelope. The intensity envelope signal was down-sampled to 200 Hz and a 1/3-octave modulation spectrum representation was calculated by summing all spectral components over a 1/3-octave interval with the center frequencies (or modulation frequencies) ranging from 0.4 to 20 Hz. The modulation spectrum representation was normalized by the averaged intensity within the original octave band to obtain an index reflecting the amount of temporal modulation. Therefore, for each different octave band with center frequencies between 125 and 4000 Hz, the modulation indexes were produced as a function of the modulation frequencies ranging from 0.4 to 20 Hz.

D. Procedures

Normal-hearing subjects listened to the stimuli monaurally presented via a Sennheiser HDA 200 headphone in an independent atomic center (IAC) sound-treated booth. The speech presentation level was always at 65 dBA. The noise level was varied to produce different signal-to-noise ratios. The cochlear-implant subjects listened to the stimuli monaurally presented via direct connection. The speech presentation level was adjusted to fit the individual subject's most comfortable level.

To avoid a sentence repetition effect on intelligibility, sentences were used only once for a given subject over the course of the entire experiment. To avoid a presentation order effect, in each testing session clear and conversational speech sentences were mixed together and presented in random order. To counteract a task-learning effect, the experimental conditions were conducted in the order of decreasing signal-to-noise ratios from 20 to −20 dB and the number of bands from 16 to 2. To minimize the effect of differences in inherent difficulty among the sentence lists, each subject was presented with 7 conversational speech lists and 7 clear speech lists that were randomly selected from a total of 18 sentence lists. Finally, to familiarize the subjects with the test materials and procedures, a short session with 5 sentences in quiet was conducted for each experiment.

For formal data collection, the subjects were asked to type the sentence presented via a keyboard and were instructed to double-check the spelling before entering the answer. A computer program automatically calculated the recognition accuracy score based on the number of the key words correctly identified. Each experimental condition had 8 sentences containing three or four keywords each and took about 5 min to finish. The reported result was the averaged score from these 8 sentences. Experiment I had a total of 28 conditions including 2 talkers, 2 speaking styles, and 7 signal-to-noise ratios, Experiment II had 8 conditions (1 talker X 2 speaking styles X 4 bands), Experiment III had 12 conditions (1 talker X 2 speaking styles X 6 signal-to-noise ratios), and Experiment IV had 14 conditions (1 talker X 2 speaking styles X 7 signal-to-noise ratios).

E. Data analysis

The percent correct scores (PC) as a function of signal-to-noise ratio (SNR) from experiments I, III, and IV were fitted with a three-parameter sigmoid function (Zeng and Galvin, 1999a)

$$\mathrm{PC} = \frac{S}{1 + e^{-(\mathrm{SNR} - a/b)}}, \tag{1}$$

where S indicates the asymptotic performance, a is the intercept corresponding to the SNR at which performance is 50% of the asymptotic performance, and b is a parameter related to the slope. The actual slope at the 50% of the asymptotic performance can be derived

$$\mathrm{Slope} = \frac{S}{4b}. \tag{2}$$

In addition, the speech reception threshold (SRT) corresponding to the 50% correct score can be derived

$$\mathrm{SRT} = a - b \ln\left(\frac{S}{50} - 1\right). \tag{3}$$

Finally, the dynamic range (DR), defined as the dB difference between the signal-to-noise ratios producing 10 and 90% of the asymptotic performance, can be derived

$$\mathrm{DR} = b\left[\ln\left(\frac{S}{10\%*S} - 1\right) - \ln\left(\frac{S}{90\%*S} - 1\right)\right]. \tag{4}$$

III. RESULTS

A. Speech analysis

Figure 2 shows wave-form examples from the female (left panels) and male (right panels) talkers in both clear (top panels) and conversational (bottom panels) speech styles. First, note that, regardless of the talker, clear speech has longer overall duration and contains longer and more frequent interword pauses than conversational speech. Second, note that this difference in the temporal patterns of clear and conversational speech appears to be smaller for the male talker than for the female talker.

Table II shows the mean and standard deviation of the overall duration for clear and conversational speech from both talkers. On average, for the female talker, clear speech was 2.2 times longer than conversational speech, whereas for the male talker, clear speech was only 1.5 times longer than conversational speech. Similar ratios for the standard deviation were also observed (see also Bradlow *et al.*, 2003). We should point out that the observed differences between the two talkers may reflect individual differences rather than a gender difference per se.

Figure 3 shows the modulation spectra to further quantify the differences between clear and conversational speech. First, note that clear speech generally has a larger modulation

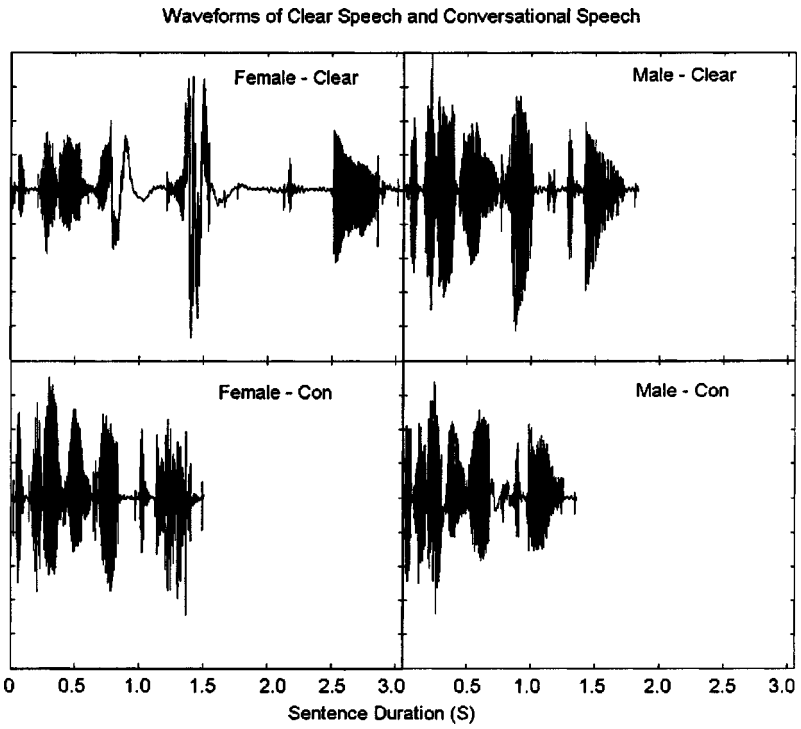


FIG. 2. Sample wave forms of clear (top panels) and conversational (bottom panels) speech by a female (left panels) and male (right panels) talker. The sentence was: *The children dropped the bag*.

index than the conversational speech, particularly for frequency bands with higher center frequency and for the female talker. A simple pairwise *t*-test across 64 sentences shows that the larger modulation index in clear speech is significant in all subbands ($p<0.05$) for the female talker but only in high-frequency bands (center frequencies $>=2000$ Hz) for the male talker. Second, note that regardless of the speech style, the bands with high center frequencies produce larger modulation indexes than the bands with low center frequencies (pairwise *t*-test, $p<0.05$), possibly reflecting the acoustic differences between consonants and vowels. Third, the overall modulation spectra have a lower global peak for clear speech (1–3 Hz) than conversational speech (2–6 Hz). Considering the overall duration difference between clear and conversational speech (Fig. 2 and Table I), these peaks most likely reflect the overall syllable rate. On the other hand, the small but distinct peaks at lower modulation frequencies (e.g., 0.5–0.8 Hz) most likely reflect the overall sentence rate. For example, the averaged sentence duration for the male conversational speech was 1.3 s, corresponding nicely to the 0.8-Hz peak in the modulation spectrum. Unfortunately, the averaged sentence duration for the female clear speech was too long (3.3 s) to be displayed as a peak (0.3 Hz) in the present modulation spectrum since it was beyond the range of the analyzed modulation frequency.

TABLE II. Average duration and standard deviation of 144 clear and 144 conversational speech sentences in both female and male talkers (These data are also reported in Bradlow *et al.* 2003).

Talker	Clear speech (s)	Conversational speech (s)
Female	3.32±0.45	1.47±0.19
Male	1.97±0.27	1.31±0.14

B. Experiment I: Speech perception in normal-hearing subjects

Figure 4 shows percent correct scores as a function of signal-to-noise ratio obtained with both female (left panel) and male (right panel) talkers for clear (open circles) and conversational (closed triangles) speech. The most significant finding is that clear speech produced higher intelligibility than conversational speech for both female and male talkers [$F(1,11)=105.09$, $p<0.05$]. The clear speech advantage can be viewed by examining both the percent correct scores at a given signal-to-noise ratio and the SRT difference. For example, the percent correct score at -5 dB was 86.0 and 60.2% for the female clear and conversational speech, respectively. Similarly, the score was 80.7 and 55.5% for the male clear and conversational speech, respectively. The SRT difference between the clear and conversational speech was 3.1 and 2.2 dB for the female and male talker, respectively. There was no significant difference between the female and male talkers [$F(1,11)=0.24$, $p>0.05$].

As expected, the percent correct score increased as a function of signal-to-noise ratio [$F(6,66)=430.05$, $p<0.05$]. A significant interaction between speech style and signal-to-noise ratio was also observed [$F(6,66)=19.43$, $p<0.05$]. The interaction reflected a significant difference in performance between clear and conversational speech at the intermediate signal-to-noise ratios, but no significant difference at low and high signal-to-noise ratios due to the floor and ceiling effect, respectively.

C. Experiment II: Speech perception with reduced spectral cues in quiet

Figure 5 shows percent correct scores as a function of the number of bands for clear (open circles) and conversa-

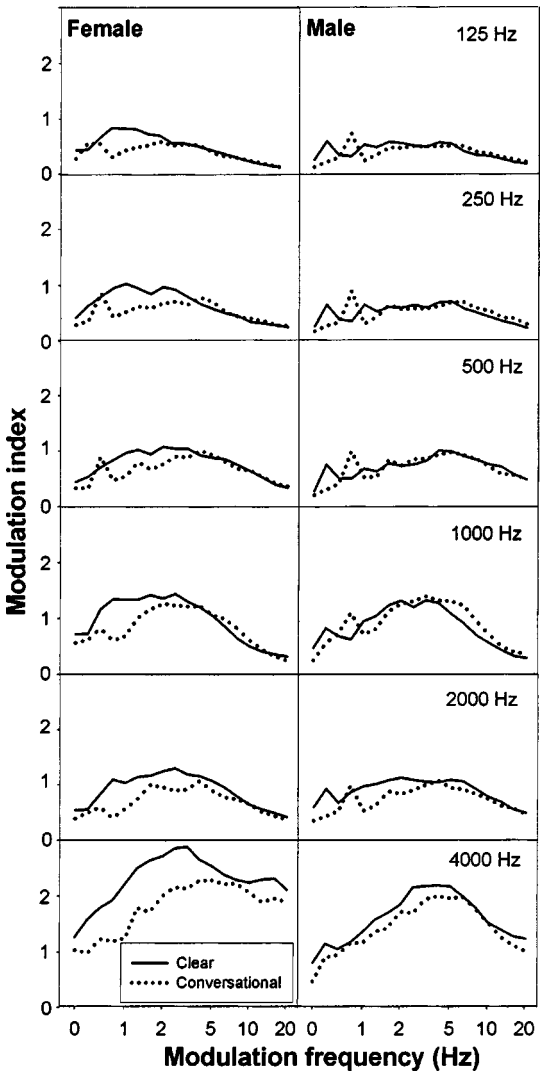


FIG. 3. Modulation spectra showing modulation index (*y*-axis) as a function of modulation frequency (*x*-axis) for the female (left panels) and male (right panels) talkers. The modulation spectra were measured in octave bands from 125 to 4000 Hz with the solid line representing clear speech and the dotted line representing conversational speech.

tional (closed triangles) speech with the female talker in quiet only. Both clear and conversational speech perception increased from essentially 0% with 2 bands to 100% with 8 and 16 bands [$F(3,15)=10.21$, $p<0.05$], but clear speech produced significantly better overall performance than conversational speech [$F(1,5)=539.95$, $p<0.05$]. A post-hoc analysis indicated that this overall difference was due to a 35 percentage point advantage for clear speech over conversational speech in the 4-band condition only [$F(1,5)=81.39$, $p<0.05$].

D. Experiment III: Speech perception with reduced spectral cues in noise

Figure 6 shows percent correct scores for the eight-band clear (open circles) and conversational (closed triangles) speech as a function of signal-to-noise ratio in normal-hearing subjects. Similar to the natural stimuli (Fig. 4 in Experiment I), both the speech style [$F(1,11)=351.82$, $p<0.05$] and the signal-to-noise ratio [$F(6,66)=60.72$, $p<0.05$] were significant factors. The eight-band clear speech had a 32.4 percentage point advantage over the eight-band conversational speech (70.9% vs. 38.5%) at the −5 dB signal-to-noise ratio. The corresponding difference in SRT was 3.2 dB, essentially identical to the 3.1 dB difference in the natural signals. Different from the natural signals, no significant interaction was observed between speech style and signal-to-noise ratio [$F(6,66)=3.96$, $p>0.05$], indicating a more or less parallel shift in the overall performance from the clear speech to the conversational speech. In fact, no floor or ceiling effect was observed as there were still about 10 and 15 percentage point differences between clear and conversational speech at the −10 and 15 dB signal-to-noise ratios, respectively.

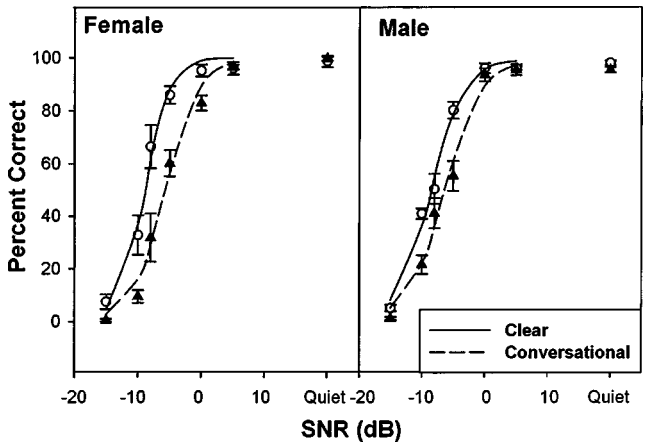


FIG. 4. Results of Experiment I showing percent correct scores as a function of signal-to-noise ratios in 11 normal-hearing subjects for the original clear (open circles) and conversational (closed triangles) speech produced by the female (left panel) and the male (right panel) talker, respectively, five subjects for female and six for male talker experiments. The solid line represents the best fitting of a sigmoid psychometric function for clear speech and the dashed line represents the best fit for conversational speech.

E. Experiment IV: Speech perception in cochlear-implant listeners

Figure 7 shows percent correct scores for clear (open symbols) and conversational (closed symbols) speech as a function of signal-to-noise ratio in eight cochlear-implant subjects. The left panel shows averaged data and standard deviations from five good users whose intelligibility scores were 75% or higher for conversational speech in quiet. The right panel shows averaged data from three relatively poor users whose scores were 60% or lower for conversational speech in quiet. The reason for dividing them into two groups was that we intended to derive globally useful parameters such as speech reception threshold; for example, if a user's score was less than 50%, it would be meaningless and theoretically impossible to derive the 50%-correct speech reception threshold.

Despite large individual variability, Fig. 7 shows an apparent clear speech advantage. Even taking all cochlear implant subjects into account, the pattern of results obtained with cochlear-implant subjects was remarkably similar to that obtained with the eight-band simulation in normal-hearing subjects: both the speech style [$F(1,8)=50.37$,

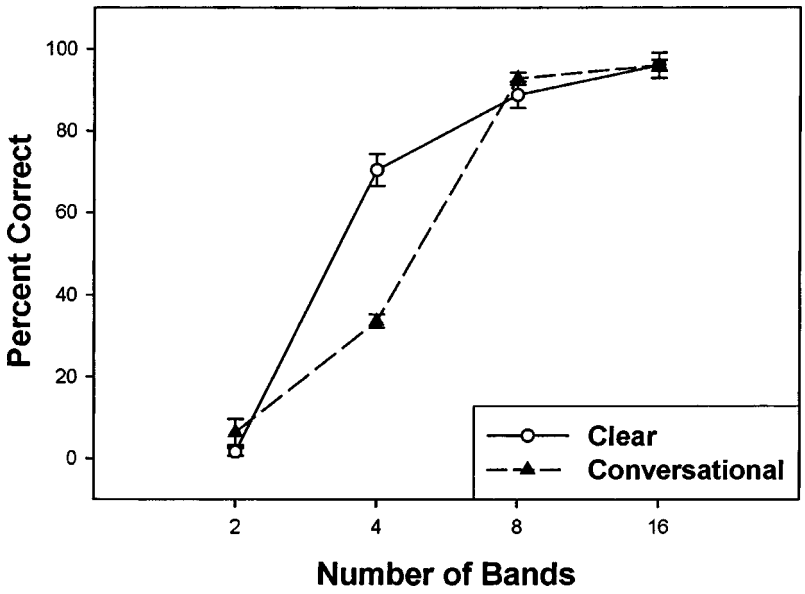


FIG. 5. Results of Experiment II showing percent correct scores as a function of the number of frequency bands (*x*-axis) in five normal-hearing subjects for the processed clear (open circles) and conversational (closed triangles) speech produced by the female talker only. The solid line represents the best fitting of a sigmoid psychometric function for clear speech and the dashed line represents the best fit for conversational speech.

$p<0.05$] and signal-to-noise ratio [$F(6,48)=81.99$, $p<0.05$] were significant factors with no significant interaction between them [$F(6,48)=2.89$, $p>0.05$]. The averaged asymptotic performance from the five good users was 95.4% correct for clear speech and 88.8% for conversational speech. The averaged asymptotic performance from the three poor users was 62.8% correct for clear speech and 49.6% for conversational speech. For the good users, the intelligibility difference between clear and conversational speech was the smallest (five percentage points) at the −10 dB signal-to-noise ratio and the largest (35 percentage points) at −5 dB. In contrast, the smallest intelligibility difference for the poor users was zero due to the floor effect at −10 and −5 dB and the largest was 30 percentage points at 0 dB.

IV. DISCUSSION

A. Macroanalysis of the perceptual data

Table III summarizes three fitting parameters and two derived parameters for the perceptual data from Experiments I, III, and IV [see Eqs. (1)–(4)]. For Experiment IV, only the averaged data from five good users were included for discussion. Except for the male conversational speech condition in Experiment I where the asymptotic performance approached perfect level, clear speech always produced higher asymptotic performance (*S*), lower speech reception thresholds (SRT), and a steeper slope (*b*) than conversational speech. Note, however, that the relative difference in SRT between clear and conversational speech in the natural condition (3.1 dB) was closely preserved in both simulated (3.2 dB) and actual (4.2 dB) cochlear implant conditions. Note also that both simulated and actual cochlear-implant listeners produced similar asymptotic performance, slope, and dynamic range.

Detailed comparisons revealed several additional differences between clear and conversational speech perception. First, a simple multiplication of the slope and the SRT difference would convert the SRT difference into the traditionally measured clear speech advantage in percentage points.

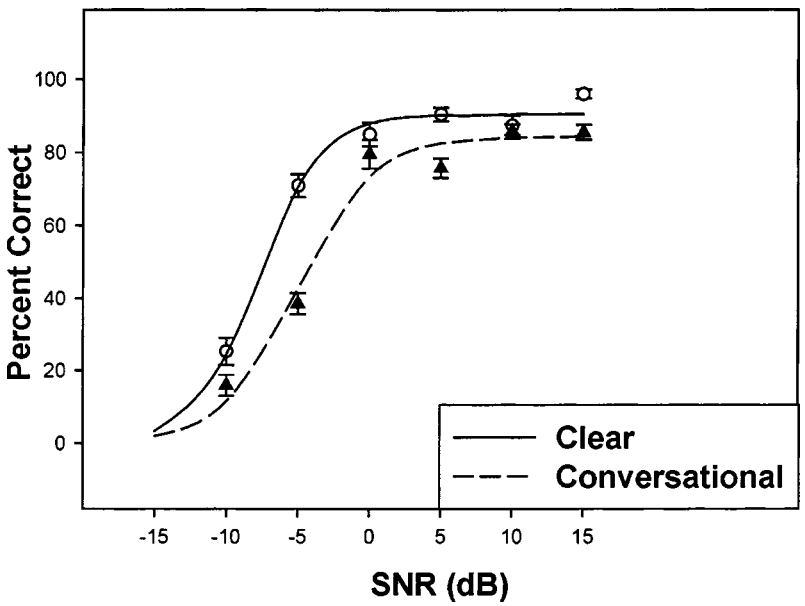


FIG. 6. Results of Experiment III showing percent correct scores as a function of signal-to-noise ratios in 11 normal-hearing subjects for the eight-band processed clear (open circles) and conversational (closed triangles) speech. The solid line represents the best fitting of a sigmoid psychometric function for clear speech and the dashed line represents the best fit for conversational speech.

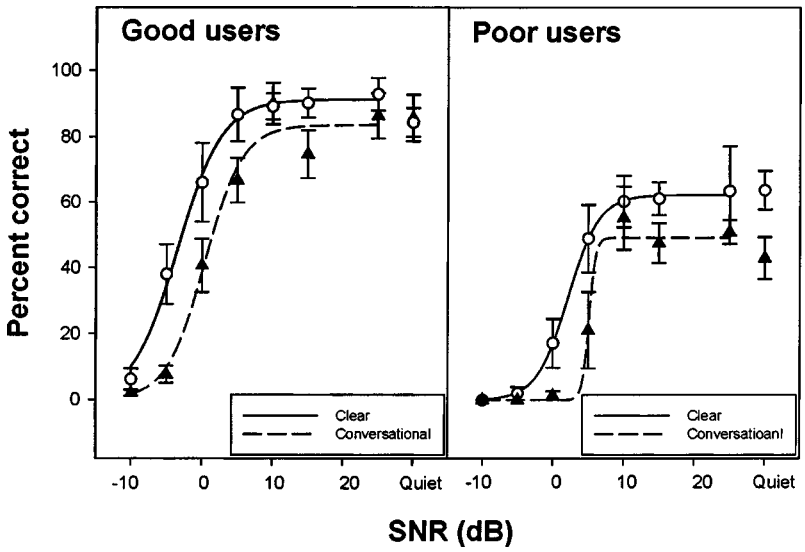


FIG. 7. Results of Experiment IV showing percent correct scores as a function of signal-to-noise ratios in cochlear-implant subjects for the original clear (open circles) and conversational (closed triangles) speech. The left panel represents data from five good cochlear implant users and the right panel represents data from three relatively poor users. The solid line represents the best fitting of a sigmoid psychometric function for clear speech and the dashed line represents the best fit for conversational speech.

The estimated clear speech advantage in percentage point difference was 29.2, 28.8, and 37.8 for the normal hearing, simulated, and actual cochlear implant good users conditions, respectively. This result implies that cochlear-implant listeners may benefit more from clear speech than normal-hearing listeners. Second, compared with the natural speech in normal-hearing listeners, the overall psychometric function was shifted toward a higher signal-to-noise ratio for the simulated (+2 dB) and actual (+6 dB) cochlear-implant experiments. The greater shift in the actual cochlear implant condition was consistent with previous studies using the hearing in noise test (HINT) sentences (e.g., Dorman *et al.*, 1997; Friesen *et al.*, 2001). Finally, note the slightly better performance [$F(1,5)=11.31$, $P<0.05$] with the 25 dB signal-to-noise ratio condition than the quiet condition in the good cochlear implant user group, indicating that a low-level noise might improve speech performance (Zeng *et al.*, 2000; Collins., 1999).

B. Microanalysis of the perceptual data: Individual differences

Here, the clear speech advantage is discussed at a more detailed level by examining who benefited from clear speech and where the benefit occurred. Figure 8 shows the improvement in percentage points for clear speech relative to conversational speech as a function of the percent correct score for conversational speech by normal-hearing (top panel), simulated (middle panel), and actual (bottom panel) cochlear-implant conditions. The closed symbols represent individual data obtained in noise and the open circles represent data obtained in quiet. The minus 45° diagonal line represents the theoretical maximum of the clear speech advantage. For example, if conversational speech perception had already reached a 100% performance level, then the largest improvement that clear speech could reach would be at the same level, resulting in a zero percentage point improvement. This was true for the natural condition in quiet (see the right most open circle on the top panel).

Note first that the overall trend of the improvement in all conditions had an inverted ''U'' shaped curve, indicating that clear speech provided the maximal benefit in terms of percentage points when conversational speech scored moderately. When conversational speech scored too high or too low, the benefit that clear speech provided reached a minimal point. In addition, the individual variability increased significantly from the normal-hearing condition to the simulated and actual cochlear-implant conditions. However, both good and poor cochlear implant users clearly derived a significant clear speech advantage. In the quiet condition, the two implant users who scored the lowest with the conversational speech (about 40%, the two leftmost open circles in bottom panel) had an improvement of 19 and 36 percentage points with the clear speech. In noise conditions, many implant users reached or approached the maximal benefits when their conversational speech had scores of 40% correct or above. The greatest benefit of 64 percentage points (the highest filled square) was achieved by a good user who scored merely 10% correct with the conversational speech at −5 dB signal-to-noise ratio. This good user achieved a 96% correct score in conversational speech recognition in quiet (the third open circle from the right).

C. Talker and rate effects

When different talkers are instructed to speak clearly, they may use different strategies to produce clear speech by slowing down the overall speaking rate, by inserting pauses, enhancing consonant intensity, increasing plosive duration, and/or expanding the vowel space. Our acoustic analysis shows that the female and male talker in the present study appeared to use different strategies to produce clear speech. While she had a comparable speaking rate for conversational speech, the female talker had a much slower rate than the male talker in producing clear speech (Table II). Because no statistical difference was observed in intelligibility between the female and male clear speech for the listeners, the present result provides additional evidence for the previously proposed hypothesis that speaking rate is not the most critical acoustic cue responsible for the clear speech advantage (Krause and Braida, 2002). However, Bradlow and Bent (2002) and Bradlow *et al.* (2003) reported that the female

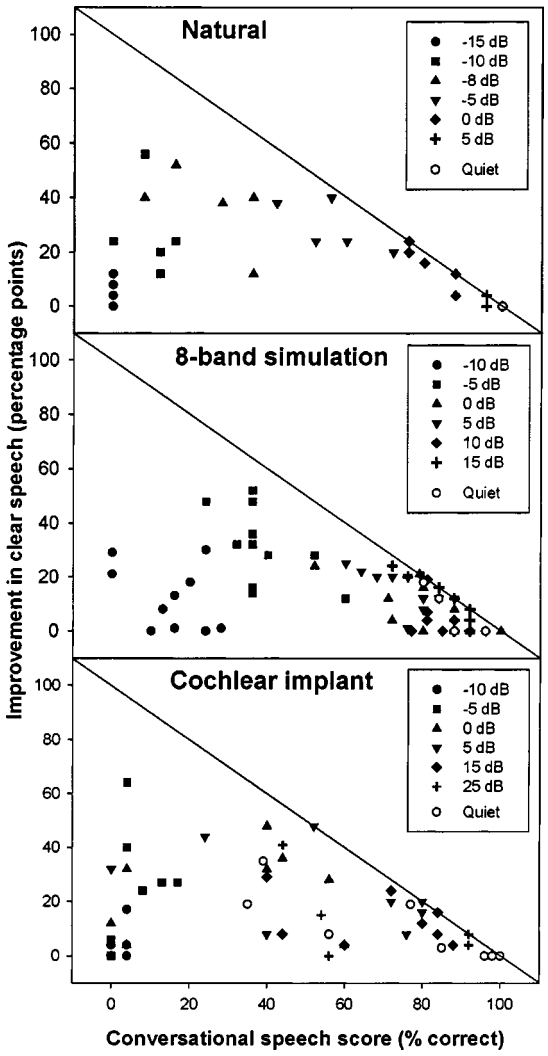


FIG. 8. Individual data showing the clear speech advantage in percentage points (*y*-axis) as a function of the conversational speech score (*x*-axis) in normal-hearing (top panel), simulated (middle panel), and actual (bottom panel) cochlear-implant listeners. The minus 45° line represents the predicted theoretical maximum of the clear speech advantage.

talker's clear speech intelligibility was significantly higher than that of the male talker when the listeners were adult non-native listeners, and children with or without diagnosed learning disabilities. This difference in performance between the present study and the study of Bradlow *et al.* (2003) suggests that different listener populations are sensitive to different clear speech features.

D. Temporal envelope and fine structure

Both the acoustic analyses and the perceptual results in the present study implicate a critical contribution of the temporal envelope to the clear speech advantage. Acoustic analyses of the two speaking styles showed that clear speech had larger temporal modulation indexes than conversational speech for both the female and the male talkers (Fig. 3). This result is consistent with that of Krause and Braida (2002) who showed a similar difference in temporal modulation indexes between clear and conversational speech.

The present study provides direct evidence linking this acoustic difference to the perceptual difference observed between clear and conversational speech. We showed that when the temporal fine structure was removed but the temporal envelope was preserved in the four-band cochlear implant simulation, clear speech produced an intelligibility score that was 35 percentage points higher than the conversational speech in the quiet condition (Fig. 5). More importantly, we showed in both the simulated and actual cochlear implant conditions that the clear speech advantage in noise (3–4 dB SRT difference) was preserved. The consistent clear speech advantage strongly supports the hypothesis that temporal envelope is a major acoustic correlate responsible for the difference in clear and conversational speech perception. On the other hand, the overall shift in the psychometric function in the simulated and actual cochlear implant results indicates that temporal fine structure contributes equally to both clear and conversational speech perception. Together, these data suggest that better encoding of both temporal envelope and fine structure is needed to improve cochlear implant performance in noise.

TABLE III. Comparison of parameters derived from the psychometric function in Experiments I, III, and IV (column 1). The asymptotic performance level "*S*" and the intercept "*a*" were defined by Eq. (1) in the text. The slope, speech-reception-threshold (SRT), and dynamic range (dB) were defined by Eqs. (2), (3), and (4), respectively, in the text.

Experiment	*S* (%)	*a* (dB)	Slope (%/dB)	SRT (dB)	DR (dB)
I-Female-Clear	95.4	−9.1	14.0	−9.0	7.5
I-Female-Conv.	92.9	−6.3	10.6	−5.9	9.6
I-Male-Clear	98.5	−8.6	9.9	−8.5	11.0
I-Male-Conv.	100.0	−6.3	8.6	−6.3	12.8
III-Clear	90.5	−7.8	10.3	−7.3	9.7
III-Conv.	84.3	−5.1	7.8	−4.1	11.9
IV[a]-Clear	95.4	−3.9	9.2	−3.6	11.4
IV[a]-Conv.	88.8	0.0	8.7	0.6	11.2

[a]Because three cochlear implant subjects only achieved an asymptotic performance at 60% correct or below, their data were not used to derive the parameters for experiment IV.

V. CONCLUSIONS

Consistent with previous acoustic studies, the present study shows that clear speech has a slower speaking rate and larger temporal modulation indexes than conversational speech. Also, consistent with previous perceptual studies, the present study finds a significant clear speech advantage in intelligibility, particularly in noise. Through the system atic collection and quantitative analysis of acoustic and perceptual data in both normal-hearing and cochlear-implant listeners, the present study has revealed several findings including:

(1) A quantitative measure of the clear speech intelligibility advantage in terms of the speech reception threshold (SRT) equal to 3.1, 3.2, and 4.2 dB in normal-hearing, simulated, and actual cochlear-implant listeners, respectively (Table III). Taking the slope and dynamic range into account, these SRT differences translate into higher clear speech intelligibility scores of 29.2, 28.8, and 37.8 percentage points, respectively.

(2) A direct relationship between greater temporal modulations and higher intelligibility scores (Figs. 3 and 4). This relationship is validated by the preserved, or even

slightly enhanced, clear speech advantage in both simulated and actual cochlear-implant conditions where primarily temporal envelope cues were available (Figs. 5–7).

(3) A demonstration of the clear speech advantage in both good and poor cochlear-implant users (Figs. 7 and 8). However, a high degree of variability still exists for clear speech perception with some cochlear-implant users achieving the theoretical maximal benefit while others derive a relatively small benefit.

(4) A different role of temporal envelope and fine structure in clear and conversational speech perception. While the temporal fine structure contributes equally to both clear and conversational speech perception, the temporal envelope carries acoustic cues that contribute to the clear speech intelligibility advantage.

ACKNOWLEDGMENTS

The authors thank the Associate Editor, Dr. Peter Assmann and two anonymous reviewers for their helpful comments on earlier drafts of this paper. The authors thank our normal-hearing and cochlear-implant subjects for their time and dedication. This work was supported by a grant from the National Institutes of Health, Department of Health and Human Services (2 RO1-DC02267).

Bench, J., and Bamford, J. (**1979**). *Speech-hearing tests and the spoken language of hearing-impaired children* (Academic Press, London).

Bond, Z. S., and Moore, T. J. (**1994**). ''A note on the acoustic-phonetic characteristics of inadvertently clear speech,'' Speech Commun. **14**, 325–337.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). ''Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics,'' Speech Commun. **20**, 255–272.

Bradlow, A. R., and Pisoni, D. B. (**1999**). ''Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors,'' J. Acoust. Soc. Am. **106**, 2074–2085.

Bradlow, A. R., and Bent, T. (**2002**). ''The clear speech effect for non-native listeners,'' J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., Kraus, N., and Erin, H. (**2003**). ''Speaking clearly for learning-impaired children: Sentence perception in noise,'' J. Speech Lang. Hear. Res. **46**, 80–97.

Chen, F. R. (**1980**). ''Acoustic characteristics and intelligibility of clear and conversational speech at segmental level,'' Unpublished master's dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Collins, L. M. (**1999**). ''SR in cochlear implant speech perception,'' Paper presented at 1999 Conference on Implantable Auditory Prostheses, Pacific Grove, CA. (1999).

Dirks, D. D., Morgan, D. E., and Dubno, J. R. (**1982**). ''A procedure for quantifying the effects of noise on speech recognition,'' J. Speech Hear Disord. **47**, 114–123.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). ''Speech intelligibility as a function of the number of channelsof stimulation for signal processors using sin-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.

Ferguson, S. H., and Kewley-Port, D. (**2002**). ''Vowel Intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **112**, 259–271; Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). ''Speech recognition in noise as a function of the number of spectral channel: comparison of acoustic hearing and cochlear implants,'' *ibid.* **110**, 1150–1163.

Gagne, J. P., Masterson, V., Munhall, K. G., Bilida, N., and Querengesser, C. (**1994**). ''Across talker variability in auditory, visual, and audiovisual speech intelligibility for conversational and clear speech,'' J. Acad. Rehabil. Audiol. **27**, 135–158.

Gagne, J. P., Querengesser, C., Folkeard, P., Munhall, K. G., and Mastern, V. M. (**1995**). ''Auditory, visual, and audiovisual speech intelligibility for sentence-length stimuli: An investigation of conversational and clear speech,'' The Volta Review **97**, 33–51.

Gagne, J. P., Rochette, A. J., and Charest, M. (**2002**). ''Auditory, visual and audiovisual clear speech,'' Speech Commun. **37**, 213–230.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species-29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Helfer, K. S. (**1997**). ''Auditory and auditory-visual perception of clear and conversational speech,'' J. Speech Lang. Hear. Res. **40**, 432–443.

Helfer, K. S. (**1998**). ''Auditory and auditory-visual recognition of clear and conversational speech by older adults,'' J. Am. Acad. Audiol **9**, 234–242.

Hou, Z., and Pavlovic, C. V. (**1994**). ''Effects of temporal smearing on temporal resolution, frequency selectivity, and speech intelligibility,'' J. Acoust. Soc. Am. **96**, 1325–1340.

Houtgast, T., and Steeneken, H. J. M. (**1985**). ''A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria,'' J. Acoust. Soc. Am. **77**, 1069–1077.

Krause, J. C. (**2001**). ''Properties of naturally produced clear speech at normal rates and implications for intelligibility,'' Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Krause, J. C., and Braida, L. D. (**2002**). ''Investigating alternative forms of clear speech: the effects of speaking rate and speaking mode on intelligibility,'' J. Acoust. Soc. Am. **112**, 2165–2173.

Krause, J. C., and Braida, L. D. (**2004**). ''Acoustic properties of naturally produced clear speech at normal speaking rates,'' J. Acoust. Soc. Am. **115**, 362–378.

Moon, S.-J., and Lindblom, B. (**1994**). ''Interaction between duration, context and speaking-style in English stressed vowels,'' J. Acoust. Soc. Am. **96**, 40–55.

Noordhoek, I. M., and Drullman, R. (**1997**). ''Effect of reducing temporal intensity modulations on sentence intelligility,'' J. Acoust. Soc. Am. **101**, 498–502.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). ''Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **95**, 1581–1592.

Payton, K. L., and Braida, L. D. (**1999**). ''A method to determinate the speech transmission index from speech waveforms,'' J. Acoust. Soc. Am. **106**, 3637–3648.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). ''Speaking clearly for the hard of hearing I: intelligibility differences between clear and conversational speech,'' J. Speech Hear. Res. **28**, 96–103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). ''Speaking clearly for the hard of hearing II: intelligibility differences between clear and conversational speech,'' J. Speech Hear. Res. **29**, 434–446.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1989**). ''Speaking clearly for the hard of hearing III: intelligibility differences between clear and conversational speech,'' J. Speech Hear. Res. **32**, 600–603.

Rosen, S. (**1992**). ''Temporal information in speech: Acoustic, auditory and linguistics aspects,'' Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Schum, D. J. (**1996**). ''Intelligibility of clear and conversational speech of young and elderly talkers,'' J. Am. Acad. Audiol **7**, 212–218.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.

Steeneken, H. J. M., and Houtgast, T. (**1980**). ''A physical method for measuring speech-transmission quality,'' J. Acoust. Soc. Am. **67**, 318–326.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). ''Speaking clearly for the hard of hearing IV: Further studies on speaking rate,'' J. Speech Hear. Res. **39**, 494–509.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). ''Speech waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, R. D. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.

Zeng, F. G., and Galvin, J. J. (**1999a**). ''Amplitude mapping and phoneme recognition in cochlear implant listeners,'' Ear Hear. **20**, 60–74.

Zeng, F. G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (**1999b**). ''Temporal and speech processing deficits in Auditory Neuropathy,'' NeuroReport **10**, 3429–3435.

Zeng, F. G., Fu, Q.-J., and Morse, R. P. (**2000**). ''Human hearing enhanced by noise,'' Brain Res. **869**, 251–255.

Zeng, F. G., Grant, G., Niparko, J., Galvin, J., Shannon, R., Opie, J., and Segel, P. (**2002**). ''Speech dynamic range and its effect on cochlear implant performance,'' J. Acoust. Soc. Am. **111**, 377–386.

Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts[a)]

Yukari Hirata[b)]
Department of East Asian Languages and Literatures, Colgate University, Hamilton, New York 13346

(Received 8 August 2003; revised 18 June 2004; accepted 18 June 2004)

This study investigated whether native speakers of American English with no knowledge of Japanese could learn to perceive Japanese vowel and consonant length distinctions through auditory training with immediate feedback. One group of participants was trained to identify the number of moras in Japanese words spoken in isolation (word training), and another group in sentences (sentence training). Trained groups' pretest and post-test scores in the words-in-isolation context (word context) and the words-in-sentences context (sentence context) were compared to those of an untrained control group. The questions addressed were whether there was an overall effect of training, and whether there were differential effects of two types of training. Both trained groups showed similar improvement in their overall test scores. The results suggested that learning in one context generalized to the other. However, an advantage of sentence training over word training was found: at the post-test, there was a greater difference between the scores of the two contexts for the word-training group than for the sentence-training group. The results are discussed in terms of the factors that might contribute to the differences in second language learning between the word and the sentence contexts. [DOI: 10.1121/1.1783351]

PACS numbers: 43.71.Hw, 43.71.Es [RLD] Pages: 2384–2394

I. INTRODUCTION

The present study examined native English (NE) speakers' perception of Japanese vowel and consonant length contrasts presented in isolated words (word context) versus words in carrier sentences (sentence context). How second language (L2) contrasts are perceived in the word versus the sentence contexts is one variable that has not received much attention in previous L2 studies, as most studies examined only perception of isolated syllables or words. It is not known whether the ability to perceive isolated syllables or words automatically predicts the ability to perceive those units in a larger context. Recently, Tajima *et al.* (2002) trained NE speakers on Japanese phonemic length contrasts by providing minimal pairs of words in isolation. These authors found that the participants' accuracy in identifying Japanese length contrasts before and after this training depended on whether words were spoken in isolation or in sentences, and how fast they were spoken. The participants' overall accuracy improved after training, but the effects of context (isolated words versus words in carrier sentences) and rate (slow versus normal versus fast) were greater after training. Tajima *et al.*'s results imply that an ability to perceive L2 contrasts gained from isolated syllable/word training might not generalize well to perception of the L2 contrasts in sentences.

Similarly, Strange *et al.* (1998) found that native Japanese listeners' perceptual assimilation of English vowels significantly differed when words are heard in isolation versus in sentences. Strange *et al.* pointed out ''the need to consider the global context in which the vowels are produced and presented'' (p. 342) when studying perception of vowels, as most previous studies examined only isolated syllables or words. The authors further suggested that L2 contrast training with isolated syllables may not be optimal for improving perception of those contrasts in natural speech utterances.

The present experimental design was modeled after that of Greenspan *et al.* (1988). These authors trained adult NE participants to recognize synthetic English words in isolation versus synthetic sentences, and tested their perceptual abilities in these two contexts. They found that the words-in-isolation training group only increased its ability to recognize synthetic words in isolation, whereas the sentence-training group increased its ability to recognize synthetic words both in isolation and in sentences. These results motivated the present study with NE speakers who were trained to perceive L2 contrasts. If L2 learners show differences between the contexts of words in isolation and words in sentences, theories of L2 learning must account for this and for how the difference diminishes or increases over the course of L2 learning.

Duration of words spoken in sentences tends to be shorter than duration of the same words spoken in isolation (Sagisaka and Tohkura, 1984). It is not known whether an ability to distinguish a vowel duration difference of, for example, 150 ms gained from isolated-word training in an L2 automatically generalizes to distinguishing a vowel duration difference of about 50 ms in sentences. Similarly, no previous study has examined whether L2 learners can learn to perceive difficult L2 contrasts in sentences rather than isolated words. The present study, therefore, examined how the use of sentences as well as isolated words in training affected

[a)]A portion of this work was presented in ''Learning to form new L2 phonetic categories in sentence contexts,'' Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 2003.
[b)]Electronic mail: yhirata@mail.colgate.edu

NE speakers' perception of Japanese vowel and consonant length contrasts.

Vowel and consonant length is phonemic in Japanese, and this length distinction has been found to be difficult for some non-native learners of Japanese to acquire (e.g., Han, 1992; Toda, 1997; Oguma, 2000). A difference in a minimal pair of short and long vowels or singleton and geminate consonants is indicated by the number of moras in the words, e.g., /ko/ ''child'' (1 mora) versus /kó:/ ''this way'' (2 moras), and /káta/ ''shoulder'' (2 moras) versus /kát:a/ ''won'' (3 moras). The primary acoustic correlate of length is the duration of either the vowel or the consonant. Long vowels are 2.2–3.2 times longer in duration than short vowels (Tsukada, 1999; Ueyama, 2000); only small differences have been observed between the formant frequencies of short and long vowels (Kondo, 1995; Tsukada, 1999; Ueyama, 2000). Similarly, a main acoustic difference between single and geminate obstruents is duration (1:2.25 by Beckman, 1982; see also Han, 1994).

In the present experiment, two groups of participants were trained to identify the number of moras from one to six in each word spoken either in isolation (word training) or in sentences (sentence training). The ability of the trained participants to identify the number of moras of Japanese words was measured with a pretest and a post-test (before and after training), and was compared to that of untrained participants who only took these tests. Two questions were addressed by this study. First, can NE speakers who do not know any Japanese learn to accurately perceive Japanese vowel and consonant length distinctions? Although Japanese length distinction has been found difficult for non-native learners of Japanese to acquire (Han, 1992; Toda, 1997; Oguma, 2000), studies of auditory training have demonstrated that adults can learn to perceive differences between a pair of L2 sounds that are not phonemic in their first language (L1) sound system (Morosan and Jamieson, 1989; Logan *et al.*, 1991; Rochet, 1995; Bradlow *et al.*, 1997; Francis and Nusbaum, 2002). Auditory training has been found effective not only for segmental contrasts such as in the above studies, but also for suprasegmental contrasts of tone (Wang *et al.*, 1999, 2002) and duration (Yamada *et al.*, 1995; Tajima *et al.*, 2002). These findings are consistent with Flege's (1995) speech learning model of L2 sound acquisition that the mechanisms and processes used in learning the L1 sound system remain intact in adulthood and can be applied to L2 learning. If adults' perception is indeed malleable in learning difficult L2 contrasts, NE speakers who participate in the present training would show significant improvement in their test scores while the control group would not.

The second question addressed by this study was whether differential effects of word versus sentence training were found. One group of participants participated in ''word training,'' in which target words were always presented in isolation (word context), and another group participated in ''sentence training,'' in which target words were always spoken in carrier sentences (sentence context). The abilities of the two groups to identify the number of moras of Japanese words were tested in these two contexts before and after training. This part of the experiment was designed to investigate whether there was an advantage of one type of training over the other in these two contexts. More specifically, does the word-training group's ability to identify the number of moras of words in isolation generalize to do the same in the sentence context? Does the sentence-training group's ability generalize to the word context?

In a previous study (Hirata, 1999a/1999b), the trained participants tended to show greater improvement than the untrained participants on the test items with long vowels and geminate consonants, while the two groups' improvement tended not to differ for the test items with short vowels and single consonants. However, this finding was statistically weak because different numbers of test items were included in each word type. In order to examine further whether robust differences do exist in perception of separate word types, the present study used equal numbers of test items for each word type.

Four word types were used in the present experiment: ''short segments,'' ''long vowels,'' ''geminates,'' and ''combination'' (i.e., a combination of a long vowel and a geminate consonant, or two long vowels). Counting moras in word type short segments was equivalent to counting syllables (as in English), e.g., /kudámono/ ''fruits'' (4 moras, 4 syllables). Counting moras in the other three word types, however, was always different from counting syllables. For long vowels and geminates, the number of moras was one greater than the number of syllables, e.g., /su:/ ''to suck'' (2 moras, 1 syllable) and /mat:o/ ''a mat'' (3 moras, 2 syllables). For word type combination, the number of moras was two more than the number of syllables, e.g., /tas:e:/ ''achievement'' (4 moras, 2 syllables). The test items in these four word types were randomly ordered in the tests so that participants would not predict which word type they would hear on each trial.

Since counting the number of moras is not an established means of testing non-native speakers' perception of length contrasts (see Beddor and Gottfried, 1995; Flege, 2002, for methods for assessing non-native speakers' perception), it is important to first examine how NE speakers initially perform on this task. Based on the previous results for different word types (Hirata, 1999a/1999b), it was predicted that NE participants who do not know any Japanese would not be very sensitive initially to phonemic length of vowels and consonants in Japanese, and that they would tend to count the number of syllables instead of the number of moras. For example, NE participants would initially tend to incorrectly identify the word /su:/ (2 moras) as having one mora, and the word /mat:o/ (3 moras) as having two moras because the words have one and two syllables, respectively. Because of this bias toward syllable counting, participants were predicted to initially perform well on word type short segments, for which the number of moras and the number of syllables coincided. However, for the other three word types, it was predicted that participants would initially perform worse than for short segments. These predictions were tested by analyzing all participants' pretest scores.

II. METHOD

A. Participants

Fifty-three monolingual native speakers of American English whose age ranged from 19 to 25 participated in the experiment. These participants were students or employees at the University of Alabama at Birmingham. None of the participants had studied or were studying Japanese, or had had long-term exposure to spoken Japanese. Previous foreign languages studied included Spanish (35 participants), French (11 participants), Latin (8 participants), German (5 participants), Russian (3 participants), Italian (3 participants), Greek (2 participants), as well as Portuguese, Swedish, Hebrew, Twi, and American Sign Language (1 participant each). None of these participants reported being native or near-native speakers of these languages. None of the participants reported any hearing problems. They were randomly divided into three groups: word training (group W, $n=18$, mean age $=22.1$), sentence training (group S, $n=17$, mean age$=22.1$), and control (group C, $n=18$, mean age$=22.6$).

B. Stimuli

1. Training

Training materials were 489 words, selected from Noto (1992), which were spoken in isolation and in carrier sentences (see Hirata, 1999a/1999b). Of 489 words, 395 words appeared in training only once, and the other 94 words appeared more than once (Appendix A). Ten training sessions were provided to both the word- and the sentence-training groups. Each session consisted of six blocks of ten trials. For word training, target words had one to six moras, and contained short or long vowels and single or geminate consonants in various positions within the words. For sentence training, the same sets of words used in the word training were spoken in carrier sentences by the same talkers as in the word training (see below). Each block had a different carrier sentence (Appendix B). Each of the ten sessions had six new sentences that did not appear in other sessions. The position of the target words in the carrier sentences was sentence-initial or sentence-medial.

The words and sentences were recorded by two male (talkers MA and MB) and two female (talkers FA and FB) native speakers of Japanese. For both word and sentence training, talkers MA and FA recorded sessions 1, 2, 5, 6, and 10, and talkers MB and FB recorded sessions 3, 4, 7, 8, and 9. Each of the talker pairs (MA and FA, and MB and FB) recorded alternate blocks. The materials were recorded on DAT cassettes and digitized at 48 kHz with WAVES+/ESPS speech analysis software.

2. Testing

Test items were all novel to the participants: they included none of the words and sentences that appeared in training sessions. A pretest and a post-test each contained identical sets of test items, but in a differently randomized order. Each of the tests consisted of a word test and a sentence test. The word test contained 60 test items and included four types of words: ''short segments,'' ''long vowels,'' ''geminates,'' and ''combination'' (left-most column in Tables I–IV). For short segments, 15 words with only short vowels and single consonants were used, and the number of moras equaled the number of syllables, e.g., /kudamono/ ''fruits'' (4 moras, 4 syllables). For long vowels, 15 words with a long vowel were used, and the number of moras was one greater than the number of syllables, e.g., /su:/ ''to suck'' (2 moras, 1 syllable). For geminates, 15 words with a geminate obstruent were used, and the number of moras was, again, one greater than the number of syllables, e.g., /mat:o/ ''a mat'' (3 moras, 2 syllables). For combination, 15 words each included a geminate obstruent and a long vowel, e.g., /tas:e:/ ''achievement'' (4 moras, 2 syllables), or two long vowels, e.g., /bo:so:zoku/ ''young motorcyclist groups'' (6 moras, 4 syllables). For this word type, the number of moras was two more than the number of syllables. Six test items that had two vowel sequence, e.g., /itʃiɾui/ ''first base'' (4 moras, 3 syllables), were exceptions to these mora/syllable counting rules (see Tables I–IV).

For each word type, words were of one to six moras in length. The 60 test words were randomly ordered into six blocks. All test items were recorded by a native speaker of Japanese who did not appear in training sessions (talker FC). The recordings were made on DAT cassettes and digitized at 22.05 kHz using Kay Elemetrics' COMPUTER SPEECH LAB (CSL 4300B).

The 60 words used in the sentence test were identical to those used in the word test, but were spoken in carrier sentences. The sentences were all spoken by talker FC. The position of the target words in the carrier sentences was sentence-initial and sentence-medial (Appendix C). None of these carrier sentences was used in training sessions.

For the word test, three examples of isolated words spoken by talker FC were played at the beginning of two of the six blocks, and the correct answers were shown on the computer screen. For the sentence test, at the beginning of each block, two example sentences spoken by talker FC were played, and the correct answers were given on the computer screen.

C. Procedure

In phase one, all three groups of participants took a pretest that consisted of a word test and a sentence test. In phase two, the word-training group and the sentence-training group participated in ten sessions of training. In phase three, which took place approximately 4 weeks after phase one, all three groups took a post-test that contained materials identical to the pretest.

1. Training

A six-alternative identification task was employed. During word training, the participants in the word-training group identified the number of moras, from one to six, in each word presented through headphones. Feedback was provided immediately after every response. If participants gave a correct response, the training program responded with ''Correct!'' and displayed both answer and test word, e.g., ''ha-de'' (2 moras). If the participants gave an incorrect response, the

TABLE I. Test items used for word type long vowels and all participants' response patterns for each item at the pretest in the word context. Values under ''response percentage (1–6)'' indicate the percentage of participants out of 53 who responded with the corresponding numbers of moras for each test item. Numbers in italic represent the percentage of participants who responded correctly. The percentage for the most frequently given incorrect response is in boldface.

Test items	Number of moras	Number of syllables	Response percentage					
			1	2	3	4	5	6
kju:	2	1	**37.7**	*62.3*	0.0	0.0	0.0	0.0
sú:	2	1	**28.3**	*69.8*	1.9	0.0	0.0	0.0
bóko:	3	2	0.0	**54.7**	*43.4*	1.9	0.0	0.0
buʃó:	3	2	0.0	**73.6**	*26.4*	0.0	0.0	0.0
dʒó:ʃi	3	2	0.0	**35.8**	*64.2*	0.0	0.0	0.0
ke:ki	3	2	0.0	**49.1**	*49.1*	1.9	0.0	0.0
sate:	3	2	0.0	**58.5**	*41.5*	0.0	0.0	0.0
iʃárjo:	4	3	0.0	0.0	**41.5**	*58.5*	0.0	0.0
dʒidó:ʃa	4	3	0.0	1.9	**60.4**	*37.7*	0.0	0.0
dʒokjó:dʒu	4	3	0.0	0.0	**47.2**	*50.9*	1.9	0.0
ku:haku	4	3	0.0	0.0	**67.9**	*30.2*	1.9	0.0
nju:gaku	4	3	0.0	0.0	**64.2**	*35.8*	0.0	0.0
kaike:	4	2	0.0	43.4	**47.2**	*9.4*	0.0	0.0
kakigó:ɾi	5	4	0.0	0.0	0.0	**45.3**	*50.9*	3.8
se:butsúgaku	6	5	0.0	0.0	0.0	13.2	**64.2**	*22.6*

program responded with ''Sorry...,'' gave the same information as above, and required the participants to listen to the word three additional times.

During sentence training, a carrier sentence of each block was written on the screen, e.g., ''zuibun ___ desune'' (''It's very ___, isn't it?''). The participants in the sentence-training group heard the whole sentence, e.g., ''zuibun hade desune,'' and then determined the number of moras in the target word, e.g., ''ha-de'' (2 moras). The same feedback as in the word training was given in sentence training. When the participants gave an incorrect response, the participants were required to listen to the whole sentence three additional times. This method of word and sentence training required participants to pay attention to the length of every segment in the word, not just the length of the segment in a given position.

The ten training sessions took place over 3 1/2 weeks. Each participant came to three sessions per week. When some participants could not come to their scheduled sessions, the sessions were rescheduled so that any two consecutive sessions occurred within 1 to 7 days apart. At the beginning of each session, the experimenter (native speaker of Japanese) pronounced 6–12 examples written in a handout for the participants. After the experimenter and the participants made sure how many moras each word had in those examples, the session with the automated feedback, described above, took place. Each session took about 10–20 min.

TABLE II. Test items used for word type geminates and all participants' response patterns for each item at the pretest in the word context.

Test items	Number of moras	Number of syllables	Response percentage					
			1	2	3	4	5	6
itʃ:i	3	2	0.0	**41.5**	*56.6*	1.9	0.0	0.0
kís:a	3	2	0.0	**7.5**	*88.7*	3.8	0.0	0.0
mát:o	3	2	0.0	**43.4**	*54.7*	1.9	0.0	0.0
ɾip:a	3	2	0.0	**60.4**	*37.7*	1.9	0.0	0.0
ʃís:o	3	2	0.0	**30.2**	*64.2*	5.7	0.0	0.0
ʃus:e	3	2	0.0	**13.2**	*83.0*	3.8	0.0	0.0
dzaʃ:u	3	2	0.0	**26.4**	*69.8*	3.8	0.0	0.0
dagák:i	4	3	0.0	0.0	**30.2**	*69.8*	0.0	0.0
godʒúʃ:a	4	3	0.0	0.0	**47.2**	*45.3*	7.5	0.0
mas:atsu	4	3	0.0	1.9	**71.7**	*26.4*	0.0	0.0
odot:a	4	3	0.0	0.0	**54.7**	*45.3*	0.0	0.0
sek:akú	4	3	0.0	0.0	**54.7**	*45.3*	0.0	0.0
dʒús:ai	4	2	0.0	22.6	**71.7**	*5.7*	0.0	0.0
matʃ:íbako	5	4	0.0	0.0	1.9	**64.2**	*34.0*	0.0
matsubók:uɾi	6	5	0.0	0.0	0.0	1.9	**54.7**	*43.4*

TABLE III. Test items used for word type combination and all participants' response patterns for each item at the pretest in the word context.

Test items	Number of moras	Number of syllables	Response percentage					
			1	2	3	4	5	6
tʃo:jó:	4	2	0.0	**50.9**	35.8	*11.3*	1.9	0.0
tʃu:kju:	4	2	0.0	28.3	**35.8**	*35.8*	0.0	0.0
das:o:	4	2	0.0	15.1	**77.4**	*7.5*	0.0	0.0
gap:e:	4	2	0.0	5.7	**88.7**	*5.7*	0.0	0.0
gaʃ:o:	4	2	0.0	5.7	**66.0**	*28.3*	0.0	0.0
ke:se:	4	2	0.0	**52.8**	35.8	*9.4*	1.9	0.0
mé:ka:	4	2	0.0	39.6	**54.7**	*5.7*	0.0	0.0
ʃótʃ:u:	4	2	0.0	9.4	**73.6**	*17.0*	0.0	0.0
ʃú:kjo:	4	2	0.0	5.7	**86.8**	*7.5*	0.0	0.0
tas:e:	4	2	0.0	28.3	**67.9**	*3.8*	0.0	0.0
ká:pet:o	5	3	0.0	0.0	28.3	**64.2**	*7.5*	0.0
kas:e:ka	5	3	0.0	0.0	17.0	**79.2**	*3.8*	0.0
ʃijó:rjo:	5	3	0.0	1.9	24.5	**60.4**	*13.2*	0.0
bo:só:zoku	6	4	0.0	0.0	1.9	**58.5**	32.1	*7.5*
tʃo:tók:ju:	6	3	0.0	0.0	**52.8**	47.2	0.0	*0.0*

2. Testing

A six-alternative identification task was employed for testing as well as training. The procedure of testing was the same as that of training except that participants did not receive feedback on each response they gave. All participants took the word and the sentence tests of the pretest and the post-test. The order of word and sentence tests was counterbalanced across participants.

The pretest and the post-test were conducted in a 4-week interval. Before the pretest began, the experimenter gave each participant a handout summarizing what the mora is and how it is different from the syllable, with which all participants were familiar. The experimenter then pronounced examples listed in the handout to the participants, and made sure that the participants knew how to count moras in the examples. The trained groups of participants took the post-test within 1 to 7 days after completing their final training session. The control group of participants took the post-test 4 weeks after they took the pretest so that the interval between the two tests was equivalent for the trained versus the control groups. For the participants in the control group, the summary of the mora as described earlier was given before they took the post-test. This procedure was taken to assure that the control participants knew how to count moras clearly 4 weeks after their pretest.

III. RESULTS

First, results of the pretest are presented. The purpose was to examine the participants' initial performance predicted with regard to four word types. Specifically, to examine whether the NE participants were biased to count syllables when they were asked to count moras, their initial

TABLE IV. Test items used for word type short segments and all participants' response patterns for each item at the pretest in the word context.

Test items	Number of moras	Number of syllables	Response percentage					
			1	2	3	4	5	6
já	1	1	*90.6*	**9.4**	0.0	0.0	0.0	0.0
tsubo	2	2	0.0	*90.6*	**9.4**	0.0	0.0	0.0
tʃígaku	3	3	0.0	0.0	*83.0*	**17.0**	0.0	0.0
gakubu	3	3	0.0	0.0	*98.1*	**1.9**	0.0	0.0
dʒigoku	3	3	0.0	1.9	*56.6*	**41.5**	0.0	0.0
kágaku	3	3	0.0	0.0	*86.8*	**13.2**	0.0	0.0
otósu	3	3	0.0	**3.8**	*94.3*	1.9	0.0	0.0
tʃìnomígo	4	4	0.0	0.0	0.0	*94.3*	**5.7**	0.0
kotowáru	4	4	0.0	1.9	**5.7**	*86.8*	**5.7**	0.0
kudámono	4	4	0.0	0.0	0.0	*92.5*	**7.5**	0.0
mágomago	4	4	0.0	0.0	0.0	*100.0*	0.0	0.0
tetsúgaku	4	4	0.0	0.0	1.9	*90.6*	**7.5**	0.0
itʃìrui	4	3	0.0	0.0	**7.5**	*90.6*	1.9	0.0
koinóbori	5	4	0.0	0.0	0.0	**22.6**	*66.0*	11.3
otʃìbahiroi	6	5	0.0	1.9	1.9	1.9	**26.4**	*67.9*

responses to the pretest items in the word context were analyzed for each word type. Second, a four-way analysis of variance (ANOVA) was conducted with overall percent-correct test scores, followed by *post hoc* tests with Bonferroni correction. The questions were whether there was an overall effect of training, and whether there were differential effects of two types of training.

A. Pretest results

Because of NE speakers' initial bias towards counting syllables, the participants were predicted to initially perform well on word type short segments in which the number of moras and the number of syllables coincided. However, for the other three word types, in which the number of moras and the number of syllables always differed, they would initially perform worse than for short segments. This prediction was supported, as shown in the pretest scores: The mean percent-correct scores of the pretest averaged over word and sentence contexts were 65.8% on short segments, 41.3% on geminates, 37.9% on long vowels, and 11.1% on combination. Paired sample t-tests (Bonferroni corrected) indicated that differences between all of these two word types, except for the difference between geminates and long vowels [$tD(6,52)=2.23$, $p>0.05$], were significant [short versus long: $tD(6,52)=13.41$, $p<0.01$; short versus geminates: $tD(6,52)=12.75$, $p<0.01$; short versus combination: $tD(6,52)=28.18$, $p<0.01$; long versus combination: $tD(6,52)=15.44$, $p<0.01$; geminates versus combination: $tD(6,52)=16.48$, $p<0.01$]. Thus, short segments >geminates and long vowels>combination.

Tables I–IV show all 53 participants' response patterns on all test items of the pretest in the word context separately for the four word types. The patterns of responses were similar for word types long vowels (Table I) and geminates (Table II) in that the most frequently given incorrect response (in boldface in the tables) was one less than the correct response. This incorrect response corresponded to the number of syllables. For long vowels (word context), the mean percentage of participants who responded correctly was 43.5%. Of the participants who made errors (56.5%), the percentage of participants who responded with the number of syllables (51.5%) was significantly higher than the percentage who responded with the other incorrect numbers (5.0%) [$tD(3,14)=9.56$, $p<0.01$]. For geminates, 51.3% of the participants responded correctly, but of the 48.7% who made errors, the percentage of participants who responded with the number of syllables (41.5%) was significantly higher than the percentage who responded with the other incorrect numbers (7.2%) [$tD(3,14)=4.42$, $p<0.01$]. These results suggest that participants' initial responses were biased towards counting the number of syllables of these test words.

For word type combination (Table III), only 10.9% of the participants, on average, responded correctly. The highest number of errors for each test item (in boldface) was made on either one or two numbers smaller than the correct number of moras. Note that the correct number of moras for this word type was two greater than the number of syllables, with the exception of /tʃo:tók:ju:/, where it was three greater. The patterns of responses were similar among long vowels, geminates, and combination in that a very small percentage of participants responded with numbers greater than the correct number of moras.

The short segments word type (Table IV) showed response patterns different from the other three word types. First, the correct response made for each test item was high ($M=85.9\%$), one of which reached ceiling (100% on /magomago/), in spite of the fact that none of the participants had heard these Japanese words before. Counting moras for this word type was the same as counting syllables (as in English) for 80% of the test items (20% of the test items had two short vowel sequences), and this is likely to be responsible for the high response accuracy. The number of participants who responded according to the number of syllables (74.7%) was significantly greater than the number of participants who responded with other incorrect numbers (10.3%) [$tD(2,14)=7.45$, $p<0.01$]. The response patterns for this word type were also different from the other three word types in that the highest errors (in boldface, Table IV) were made on both one number greater and one number smaller than the correct numbers of moras.

B. Overall results

The participants' percent-correct scores were submitted to a four-way ANOVA. Group (group W, group S, group C) was a between-subjects factor, and test (pretest, post-test), context (word test, sentence test), and word type (short segments, long vowels, geminates, combination) were repeated-measures factors. *Post hoc* tests were conducted with Bonferroni correction.

There were significant main effects for all four factors. For group [$F(2,50)=14.35$, $p<0.001$], the mean percent-correct scores of the trained groups (group W: 50.1%; group S: 50.1%) did not significantly differ from each other, but each of their scores was significantly better than that of the control group (group C: 39.0%) ($p<0.001$). For test [$F(1,50)=182.37$, $p<0.001$], the mean post-test score (53.7%) was significantly higher than the pretest score (39.0%). For context [$F(1,50)=311.40$, $p<0.001$], the mean test score in the sentence context (35.8%) was significantly lower than that in the word context (57.0%). For word type [$F(3,150)=306.32$, $p<0.001$], the mean score for short segments (68.9%) was significantly higher than the scores for geminates (48.7%) and long vowels (47.0%) ($p<0.001$), while the latter two did not differ from each other ($p=1$). The scores for geminates and long vowels were significantly higher than that for combination (20.9%) ($p<0.001$). This order of the four word types was the same as that reported in the pretest, and was expected from the prediction made earlier.

There was a significant group×test interaction [$F(2,50)=52.63$, $p<0.001$], indicating that the improvement from the pre- to the post-test depended on the group. As expected, there was no difference among the pretest scores of the three groups (group W: 38.9%; group S: 38.6%; group C: 39.5%) [groups W versus C: $tD(6,34)=0.24$, $p>0.1$; groups S versus C: $tD(6,33)=0.73$, $p>0.1$; groups W versus S: $tD(6,33)=0.11$, $p>0.1$]. However, the post-test scores were significantly higher for the trained groups

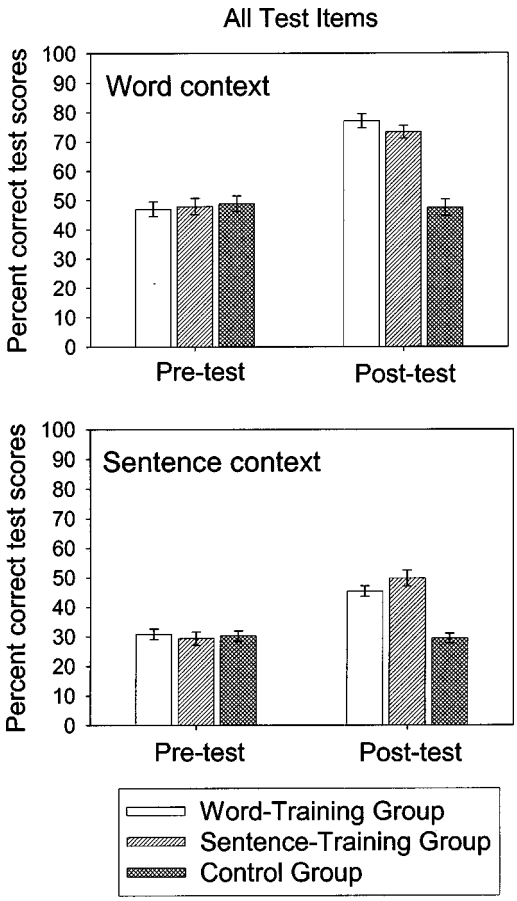


FIG. 1. Mean percent-correct scores on all test items at pretest and post-test for word training ($n=18$), sentence training ($n=17$), and control ($n=18$) groups. The error bars represent one standard error from the mean.

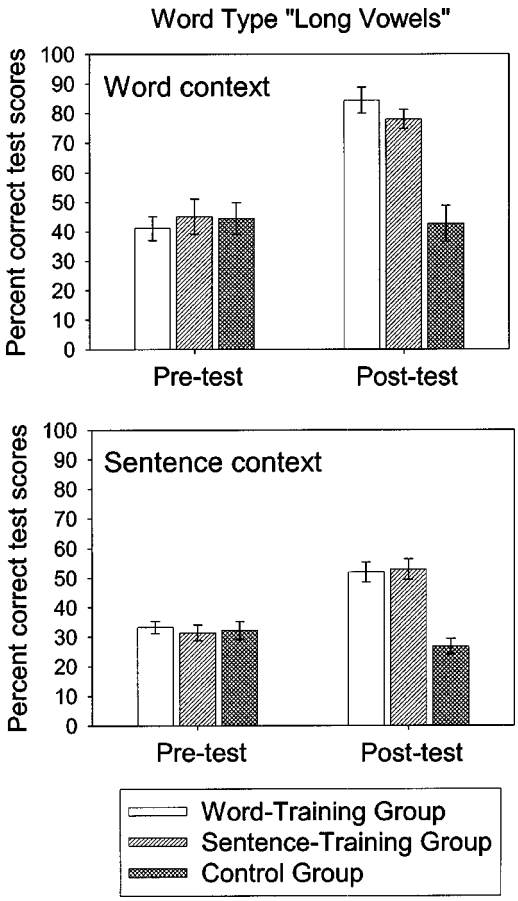


FIG. 2. Mean percent-correct scores on word type long vowels at pretest and post-test for word training ($n=18$), sentence training ($n=17$), and control ($n=18$) groups. The error bars represent one standard error from the mean.

(group W: 61.2%; group S: 61.5%) than for group C (38.5%) [groups W versus C: $tD(6,34)=8.35$, $p<0.01$; groups S versus C: $tD(6,33)=7.88$, $p<0.01$], indicating that training had an effect on the trained groups. The difference between the post-test scores of groups W and S was not significant [$tD(6,33)=0.11$, $p>0.05$], indicating that the two types of training were equally effective for improving the two groups' overall test scores.

There was a significant test×context interaction [$F(1,50)=18.57$, $p<0.001$], indicating that the amount of improvement from the pre- to the post-test, on average, differed in the two contexts. The improvement in the word context ($\Delta=18.1\%$; pre: 47.9%; post: 66.0%) was greater than that in the sentence context ($\Delta=11.4\%$; pre: 30.1%; post: 41.5%).

A group×test×context interaction was also significant [$F(2,50)=9.53$, $p<0.001$], which indicates that the amount of the three groups' improvement from the pre- to the post-test depended on the context. Figure 1 shows the three groups' mean percent-correct scores of the pretest and the post-test separately for the word and the sentence contexts. Group C improved from the pretest to the post-test in neither the word nor the sentence contexts (word: −1.3%; sentence: −0.8%). Groups W and S, in contrast, improved in both contexts, but their improvement was greater for the word context (group W: 30.1%; group S 25.4%) than for the sentence context (group W: 14.5%; group S 20.4%). Since group C did not show any improvement from the pre- to the post-tests, the test×context interaction reported earlier, i.e., the differential amount of improvement in the two contexts, was driven by the improvement of the trained groups. The trained groups' post-test scores in the word context were significantly higher than those in the sentence context [group W: $tD(5,33)=17.72$, $p<0.01$; group S: $tD(5,33)=11.68$, $p<0.01$].

One major question addressed in this study was whether the word training and the sentence training had differential effects in the two contexts. Examining the post-test in the word context, group W's score (77.1%) was higher than group S's score (73.3%), but this difference was not significant [$tD(5,33)=1.17$, $p>0.05$]. In the sentence context, group S's post-test score (49.7%) was higher than group W's score (45.3%), but this difference was not significant [$tD(5,33)=1.39$, $p>0.05$]. However, the difference score of the two contexts (word context minus sentence context) was significantly higher for group W (31.9%) than for group S (23.6%) [$tD(5,33)=3.05$, $p<0.05$]. This indicates that the type of training and the context of the post-test interacted. This difference score represents the discrepancy in their abilities to count the moras in the two contexts. Group W (word training) showed a greater discrepancy between the scores of the two contexts than group S (sentence training). Given training only in one context, group W's ability to gen-

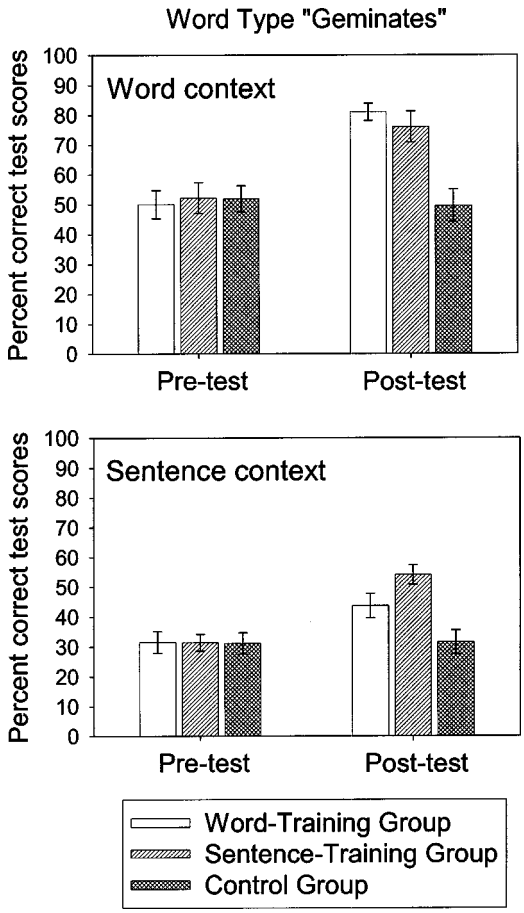


FIG. 3. Mean percent-correct scores on word type geminates at pretest and post-test for word training ($n=18$), sentence training ($n=17$), and control ($n=18$) groups. The error bars represent one standard error from the mean.

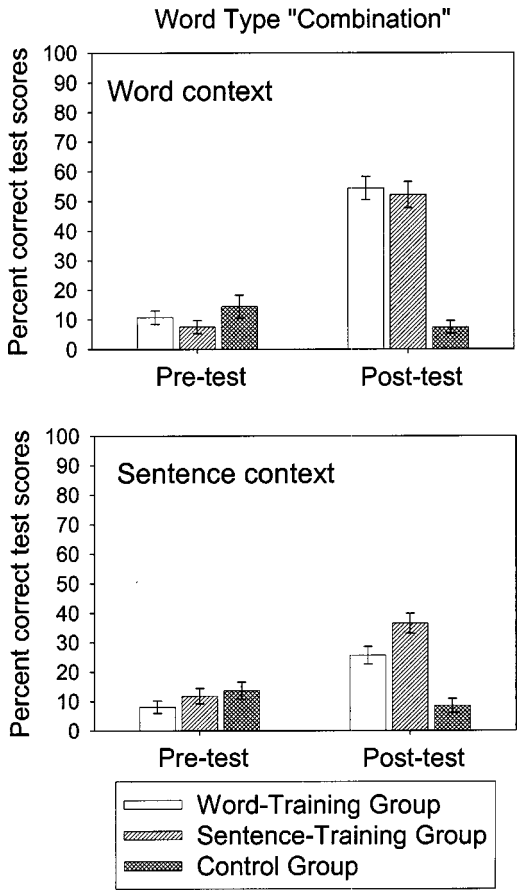


FIG. 4. Mean percent-correct scores on word type combination at pretest and post-test for word training ($n=18$), sentence training ($n=17$), and control ($n=18$) groups. The error bars represent one standard error from the mean.

eralize to the sentence context was smaller than group S's ability to generalize to the word context.

C. Results by word type

The above ANOVA showed a significant four-way interaction [$F(6,150)=2.78$, $p=0.014$]. Figures 2–5 show the three groups' test scores in the two contexts for the four word types. The results of short segments (Fig. 5) clearly differed from the rest of the word types: the three groups' scores differed from each other in neither the pretest nor the post-test ($p>0.05$). This indicates that there was no effect of training for this word type, in which the mora counting was the same as syllable counting. The results for long vowels, geminates, and combination (Figs. 2–4) were similar to those of all word types (Fig. 1) in that the post-test scores of groups W and S were higher than that of group C in the word and the sentence contexts ($p<0.01$) with one exception. In the geminates sentence context, group W's score was not significantly higher than group C's, but was not significantly lower than group S's ($p>0.05$).

The specific question of interest was whether there was a training type by context interaction in groups W's and S's post-test scores in each of these three word types, as found in the data of all word types combined. For long vowels, the post-test difference scores (word context minus sentence context) for group W (32.6%) and group S (25.1%) did not significantly differ [$tD(1,33)=1.28$, $p>0.1$], indicating that the two groups showed a similar amount of discrepancy between the two contexts. However, for geminates, the difference score of group W (37.4%) was significantly higher than that of group S (22.0%) [$tD(1,33)=2.64$, $p=0.01$], which means that group W showed a greater gap between the two contexts than group S did at the post-test. Similarly, for combination, the difference score of group W (28.9%) was significantly higher than that of group S (15.7%) [$tD(1,33)=2.84$, $p<0.01$]. Thus, word types geminates and combination were driving the training type by context interaction found in the data of all word types reported in the previous section.

IV. DISCUSSION AND CONCLUSIONS

The results of the pretest indicated that the NE speakers' accuracy in identifying phonemic vowel and consonant length differences in Japanese was initially low ($M=39\%$). The participants had a tendency to count the number of syllables when they were asked to count the number of moras. Because of this bias towards syllables, the participants performed well on word type short segments in which the number of moras and the number of syllables coincided. However, for the other three word types in which the number of moras and the number of syllables always differed, they initially performed worse than for short segments. The NE

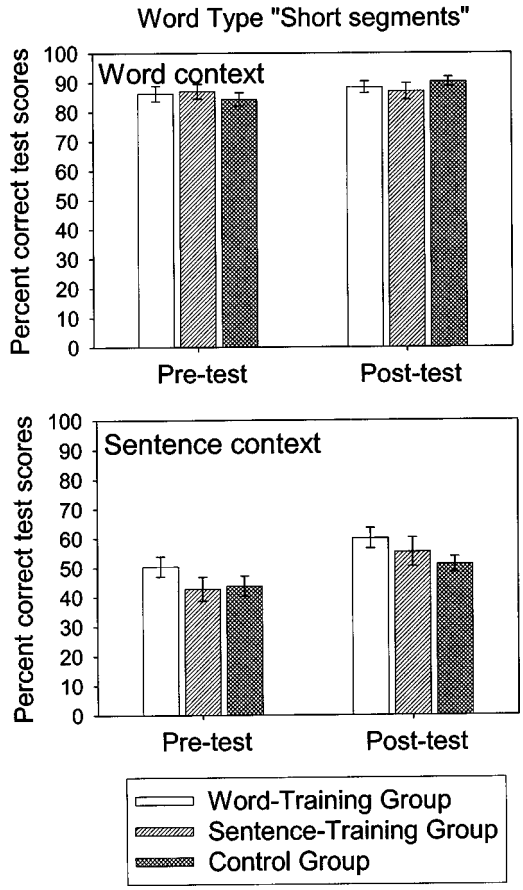


FIG. 5. Mean percent-correct scores on word type short segments at pretest and post-test for word training ($n=18$), sentence training ($n=17$), and control ($n=18$) groups. The error bars represent one standard error from the mean.

speakers lacked accuracy in detecting long vowels and geminate consonants that contribute one additional count to the number of syllables, and in distinguishing these long segments from the short counterparts.

The first question addressed by this study was whether the present training method enabled participants to perceive length contrasts accurately. Effects of training were found in the overall pre/post-test scores, and in three of the four word types examined: For long vowels, geminates, and combination, the trained groups showed the post-test scores significantly higher than that of the control group, while their pretest scores did not differ. In contrast, for word type short segments, neither the pretest nor the post-test scores differed between the control and the trained groups. As mentioned earlier, counting moras in short segments was the same as counting syllables (as in English) for 80% of the test items, and this was likely to be responsible for the consistent high scores of all three groups on the pretest and post-test. Thus, the trained groups increased their scores on long vowels, geminates, and combination while maintaining high scores on short segments. The training groups' improvement must not be attributed to rote memory, as none of the tested words appeared in their training, and they received no feedback on the tested items throughout this study. Taken together, these results suggest that the training increased the participants' sensitivity to duration of segments, and enabled them to identify more accurately the number of moras of Japanese words spoken both in isolation and in sentences.

The results that learners' accuracy in perceiving L2 length contrasts increased with training suggest that adults are capable of perceptual learning, and this is consistent with findings of the previous L2 training studies (Morosan and Jamieson, 1989; Logan *et al.*, 1991; Yamada *et al.*, 1995; Rochet, 1995; Bradlow *et al.*, 1997; Wang *et al.*, 1999; Tajima *et al.*, 2002). The present study specifically extended Yamada *et al.*'s (1995) finding in that adult NE speakers who did not know any Japanese could learn to perceive vowel and consonant length contrasts not only in a fixed position of disyllabic words (Yamada *et al.*, 1995) but also in various positions of one to six mora words. It is important to note, however, that the perceptual accuracy even of the trained participants at the post-test was far from perfect. Especially for identifying word type combination that had a combination of a geminate consonant and a long vowel, or two long vowels in a word, there is much room for further improvement. This may partially be due to the fact that training materials included only 24 words for this word type out of 489 words. It is of interest to explore whether longer training time with an increased number of this type of words would result in further improvement, or whether there is an upper limit for adults' perceptual learning capacity.

The second question addressed by this study was whether there are differential effects of word versus sentence training. The overall amount of improvement made by the word-training and the sentence-training groups did not differ from each other. That is, the two groups, regardless of the type of training, showed about the same improvement in their overall test scores, combining the two contexts. The word-training group generalized their learning to the sentence context, and the sentence-training group generalized their learning to the word context. However, the word-training group showed a greater discrepancy between the post-test scores of the two contexts than the sentence-training group, with the word test scores being higher than the sentence test scores (Fig. 1). This suggests that, given training in only one context, the degree of generalization to the untrained context was greater for the sentence training than the word training.

The effect of context was robust: the three groups' scores were consistently lower in the sentence than in the word context, and the overall improvement made from the pre- to the post-tests was smaller in the sentence context. This was the case even though the same set of words was tested in the word and the sentence tests, the order of which was counterbalanced across participants. These results suggest that an ability to perceive difficult L2 contrasts in isolated words might not automatically warrant an ability to do the same in sentences, and suggest that the two abilities involve different perceptual mechanisms. This point is consistent with Greenspan *et al.*'s (1988) findings mentioned in the Introduction. They concluded, "... the perceptual skills acquired in learning to recognize words in fluent sentences form a functional superset of the skills acquired during training with isolated words." (pp. 426–427). It is interesting that similar findings were obtained from NE speakers learning L1 synthetic speech and an L2.

What are the implications of these results for theories of L2 perception and learning? Although various factors have been identified that determine the difficulty learners face in the perception of L2 contrasts (e.g., Best, 1995; Flege, 2002), the current L2 models, which are based on findings from the isolated contexts, cannot explain the discrepancy between the word and the sentence contexts found in the present study. If the context of words-in-isolation versus words-in-sentences is a vital factor for L2 contrast perception, the current L2 perception and learning models must be amended to account for this discrepancy. An immediate question arises as to whether the discrepancy between the two contexts is applicable only to length contrasts, or whether it is applicable to other suprasegmental contrasts (e.g., Wang *et al.*, 1999) or segmental contrasts (Logan *et al.*, 1991; Morosan and Jamieson, 1989). Other important questions to be investigated in the future are what contributes to this context discrepancy, and how it diminishes or increases over a course of L2 acquisition.

One factor that might contribute to the context discrepancy is speaking rate or absolute duration of speech. As mentioned in the Introduction, the duration of words spoken in sentences tends to be shorter than duration of the same words spoken in isolation (Sagisaka and Tohkura, 1984). Pisoni (1993) and Sommers *et al.* (1992) argued that perceptual details associated with differences in speaking rate are not lost as a result of perceptual analysis, and become an integral part of the mental representation of speech. It is possible that when L2 learners are trained to detect small acoustic differences of duration (to distinguish between short and long members of the contrasts) as in the sentence training, the ability gained from this training is applicable to detecting larger differences in duration as in the word context. However, the opposite might not be automatically guaranteed: being trained to detect large differences in duration might not always result in detecting smaller differences in duration successfully.

Another possible factor that might explain the context discrepancy is that a different perceptual mechanism underlies perception of sentences, compared to the mechanism involved in perception of isolated words. For example, many participants in sentence training commented that extracting target words from carrier sentences was one of the most difficult tasks during the training. It is possible that the participants with sentence training, but not those with word training, were given an opportunity to gain a skill needed for segmenting a stream of speech into words and segments. Further, it is possible that the participants with sentence training, but not those with word training, had learned to use the temporal information of sentences for determining the length of segments in the words. This might be particularly important since a Japanese phonemic short vowel spoken at a slow rate can be longer in absolute duration than a phonemic long vowel spoken at a fast rate (Hirata, in press). These skills might be a superset of the skills (Greenspan *et al.*, 1988) that word training did not provide, and this might explain why the participants with word training generalized to the untrained context to a lesser degree.

In conclusion, the NE speakers who did not know any Japanese gained sensitivity to vowel and consonant length in Japanese words, and this suggests adults' malleability to perceive L2 contrasts that do not exist in their L1. This study also demonstrated that the context of words in isolation versus words in sentences is an important factor that affects the difficulty of L2 contrast perception. Further research is necessary to identify the mechanisms that underlie L2 perception of isolated words versus words in sentences.

ACKNOWLEDGMENTS

This study was supported by funding from the Research Council of Colgate University. I thank James E. Flege for assisting me in conducting this experiment at his laboratory. I also thank Jack Dovidio, Spencer Kelly, Alexander L. Francis, and Satomi Imai for their help in statistical analysis and in improving the manuscript. Thanks also go to Jon Bernard for creating the training program. I am also grateful to Randy L. Diehl, Robert F. Port, and an anonymous reviewer for their insightful reviews. A portion of this study appeared at the 15th International Congress of Phonetic Sciences in 2003.

APPENDIX A

Words used in training. The number of words in each word type was constrained by Japanese phonotactics, e.g., geminates do not occur word-initially or word-finally, whereas long vowels can occur in any position. The number of times the words appeared was constrained by carrier sentences, e.g., a noun (instead of an adjective or a verb) must be used in a sentence such as block 6 in training session 4 (see Appendix B).

Number of times used in training	Type of words				Total
	Short segments	Long vowels	Geminates	Combination	
Once	202	111	65	17	395
Twice	32	31	15	4	82
Three times	3	2	1	2	8
Four times	0	1	1	1	3
Five times	1	0	0	0	1
Total number of words	238	145	82	24	489
Total number of trials	280	183	102	35	600

APPENDIX B

Carrier sentences used in training session 4. (Each training session had six new sentences that did not appear in other sessions.)

Block 1 /_____ to itːe kudasai/ ''Please say ____.''
Block 2 /ija _____ dʒanai/ ''No, it is not ____.''
Block 3 /hai _____ to kikoemaʃita ga/ ''Yes, it sounded like ____.''
Block 4 /soredʒa _____ dʒa aɾimasen ne/ ''Then, it is not ____.''
Block 5 /ano: _____ desu ne/ ''Um, it is ____, isn't it?''
Block 6 /dewa _____ ni tsuite kangaemasu/ ''Then, I will think about ____.''

APPENDIX C

Carrier sentences used in tests. (None of these sentences appeared in training sessions.)

Block 1 /____ dʒa aɾimasen/ ''It is not ____.''
Block 2 /____ wa kaite aɾimasu ka/ ''Is ____ written?''
Block 3 /koɾe wa ____ to jomimasu/ ''This reads ____''
Block 4 /____ wa kaite aɾimasen/ ''____ is not written.''
Block 5 /____ dʒa nai desu jo/ ''It is not ____.''
Block 6 /soko wa ____ to jomimasu/ ''That reads ____.''

Beckman, M. (**1982**). ''Segment duration and the 'mora' in Japanese,'' Phonetica **39**, 113–135.

Beddor, P. S., and Gottfried, T. L. (**1995**). ''Methodological issues in cross-language speech perception research with adults,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 207–232.

Best, C. T. (**1995**). ''A direct realist view of cross-language speech perception,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., and Tohkura, Y. (**1997**). ''Training Japanese listeners to identify English /r/–/l/: IV. Some effects of perceptual learning on speech production,'' J. Acoust. Soc. Am. **101**, 2299–2310.

Flege, J. E. (**1995**). ''Second language speech learning: Theory, findings and problems,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 233–277.

Flege, J. E. (**2002**). ''Methods for assessing the perception of vowels in a second language,'' in *Issues in Clinical Linguistics*, edited by E. Fava and A. Mioni (UniPress, Padova), pp. 3–28.

Francis, A. L., and Nusbaum, H. C. (**2002**). ''Selective attention and the acquisition of new phonetic categories,'' J. Exp. Psychol. Hum. Percept. Perform. **28**, 349–366.

Greenspan, S. L., Nusbaum, H. C., and Pisoni, D. B. (**1988**). ''Perceptual learning of synthetic speech produced by rule,'' J. Exp. Psychol. Learn. Mem. Cogn. **14**, 421–433.

Han, M. (**1992**). ''The timing control of geminate and single stop consonants in Japanese: A challenge for non-native speakers,'' Phonetica **49**, 102–127.

Han, M. (**1994**). ''Acoustic manifestations of mora timing in Japanese,'' J. Acoust. Soc. Am. **96**, 73–82.

Hirata, Y. (**1999a**). ''Effects of sentence- vs word-level perceptual training on the acquisition of Japanese durational contrasts by English native speakers,'' Proceedings of the 14th International Congress of Phonetic Sciences **2**, 1413–1416.

Hirata, Y. (**1999b**). ''Acquisition of Japanese rhythm and pitch accent by English native speakers,'' doctoral dissertation (University of Chicago).

Hirata, Y. in press. ''Effects of speaking rate on the vowel length distinction in Japanese,'' J. Phonetics.

Kondo, Y. (**1995**). ''Production of schwa by Japanese speakers of English: A crosslinguistic study of coarticulatory strategies,'' doctoral dissertation (University of Edinburgh, UK).

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). ''Training Japanese listeners to identify English /r/ and /l/: A first report,'' J. Acoust. Soc. Am. **89**, 874–886.

Morosan, D., and Jamieson, D. G. (**1989**). ''Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word position or task,'' J. Speech Hear. Res. **32**, 501–511.

Noto, H. (**1992**). *Communicating in Japanese* (Sotakusha, Tokyo).

Oguma, R. (**2000**). ''Perception of Japanese long vowels and short vowels by English-speaking learners,'' Japanese-Language Education around the Globe **10**, 43–55. (in Japanese with English abstract).

Pisoni, D. B. (**1993**). ''Long-term memory in speech perception: Some new findings on talker variability, speaking rate, and perceptual learning,'' Speech Commun. **13**, 109–125.

Rochet, B. L. (**1995**). ''Perception and production of second-language speech sounds by adults,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 379–410.

Sagisaka, Y., and Tohkura, Y. (**1984**). ''Phoneme duration control for speech synthesis by rule,'' Trans. Inst. Electron. Commun. Eng. Jpn., Part A **J67-A**, 629–636.

Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (**1992**). ''The effects of speaking rate and amplitude variability on perceptual identification,'' J. Acoust. Soc. Am. **91**, 2340.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., and Jenkins, J. J. (**1998**). ''Perceptual assimilation of American English vowels by Japanese listeners,'' J. Phonetics **26**, 311–344.

Tajima, K., Rothwell, A., and Munhall, K. G. (**2002**). ''Native and non-native perception of phonemic length contrasts in Japanese: Effect of identification training and exposure,'' J. Acoust. Soc. Am. **112**, 2387.

Toda, T. (**1997**). ''Strategies for producing mora timing by non-native speakers of Japanese,'' in *Acquisition of Japanese as a Second Language*, edited by D. G. Shuutoku Kenkyuukai (Bonjinsha, Tokyo), pp. 157–197.

Tsukada, K. (**1999**). ''An acoustic phonetic analysis of Japanese-accented English,'' doctoral dissertation (Macquarie University).

Ueyama, M. (**2000**). ''Prosodic transfer: An acoustic study of L2 English vs L2 Japanese,'' doctoral dissertation (University of California at Los Angeles).

Wang, Y., Jongman, A., and Sereno, J. A. (**2002**). ''Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training,'' J. Acoust. Soc. Am. **113**, 1033–1043.

Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (**1999**). ''Training American listeners to perceive Mandarin tones,'' J. Acoust. Soc. Am. **106**, 3649–3658.

Yamada, T., Yamada, R. A., and Strange, W. (**1995**). ''Perceptual learning of Japanese mora syllables by native speakers of American English: Effects of training stimulus sets and initial states,'' Proceedings of the 13th International Congress of Phonetic Sciences **1**, 322–325.

Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners[a)]

Mead C. Killion, Patricia A. Niquette, and Gail I. Gudmundsen
Etymotic Research, Inc., 61 Martin Lane, Elk Grove Village, Illinois 60007

Lawrence J. Revit
Revitronix, Brownsville, Vermont 05037

Shilpi Banerjee[b)]
Northwestern University, Evanston, Illinois 60208

(Received 30 September 2003; revised 29 June 2004; accepted 30 June 2004)

This paper describes a shortened and improved version of the Speech in Noise (SIN™) Test (Etymotic Research, 1993). In the first two of four experiments, the level of a female talker relative to that of four-talker babble was adjusted sentence by sentence to produce 50% correct scores for normal-hearing subjects. In the second two experiments, those sentences-in-babble that produced either lack of equivalence or high across-subject variability in scores were discarded. These experiments produced 12 equivalent lists, each containing six sentences, with one sentence at each adjusted signal-to-noise ratio of 25, 20, 15, 10, 5, and 0 dB. Six additional lists were also made equivalent when the scores of particular pairs were averaged. The final lists comprise the ''QuickSIN'' test that measures the SNR a listener requires to understand 50% of key words in sentences in a background of babble. The standard deviation of single-list scores is 1.4 dB SNR for hearing-impaired subjects, based on test-retest data. A single QuickSIN list takes approximately one minute to administer and provides an estimate of SNR loss accurate to ± 2.7 dB at the 95% confidence level. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1784440]

PACS numbers: 43.71.Ky, 43.71.Gv, 43.72.Dv [KWG] Pages: 2395–2405

I. INTRODUCTION

Hearing aid wearers report that their biggest problem with their hearing aids is that of understanding speech in background noise, and consumer surveys polling approximately 80 000 households have consistently revealed consumer dissatisfaction with hearing aids in noisy environments (Kochkin, 1992, 1993, 1995, 1996, 2000, 2002). Kochkin (2002) reported that only 30% of hearing aid wearers were satisfied with their hearing aids in noisy situations.

As summarized below, recent evidence suggests that the wide range of satisfaction with hearing aids in noise reflects a wide range in the ability of hearing aid wearers to understand speech in a background of noise. By analogy with hearing loss, ''SNR loss'' (signal-to-noise ratio loss) refers to the increase in signal-to-noise ratio required by a listener to obtain 50% correct words, sentences, or words in sentences, compared to normal performance.[1]

Published reports indicate a wide range of SNR loss in persons with similar pure tone hearing losses (Lyregaard, 1982; Dirks *et al.*, 1982; Killion, 1997; Killion and Niquette, 2000; Taylor, 2003). The standard audiometric test battery does not measure or predict the ability to understand speech in noise (Killion and Niquette, 2000). Figure 1 shows data on 100 hearing-impaired listeners. Those with a 40–60 dB pure-tone-average (PTA) loss, for example, have SNR losses ranging from less than 2 dB (no more trouble hearing in noise than normal-hearing listeners) to greater than 20 dB (severe loss of ability to hear in noise). Without knowledge of a listener's SNR loss, it is virtually impossible to give realistic expectations for their potential improvement in noise with hearing aids. One person with a 50 dB PTA loss but without SNR loss may report little or no difficulty hearing in noise with hearing aids, while another with the same PTA loss, but a severe SNR loss, may require a remote FM microphone in order to understand speech in noise. Just as important, knowing the SNR loss makes it possible for the hearing professional to recommend the appropriate technology (e.g., omnidirectional microphones, directional microphones, array microphones, close-talking FM microphones) required for the listener to function in commonly encountered noisy situations.

SNR testing is only recently becoming common in clinical practice (Strom, 2003). Nilsson *et al.* (1994) described the hearing in noise test (HINT) that uses sentences in continuous speech-spectrum shaped noise and an adaptive procedure that gives the SNR for 50% correct for whole sentences. The use of whole-sentence scoring on the HINT has the advantage that whole sentences are tested, and the disadvantage that a greater number of sentences is required for a given statistical reliability than when key-word scoring is used. Similarly, the use of continuous noise has the advantage of reducing the variability in noise level, and the disad-

[a)]Portions of this work were published in the QuickSIN™ test manual, Etymotic Research, Inc., © 2001.
[b)]Present address: Starkey Laboratories, Eden Prairie, MN 55344.

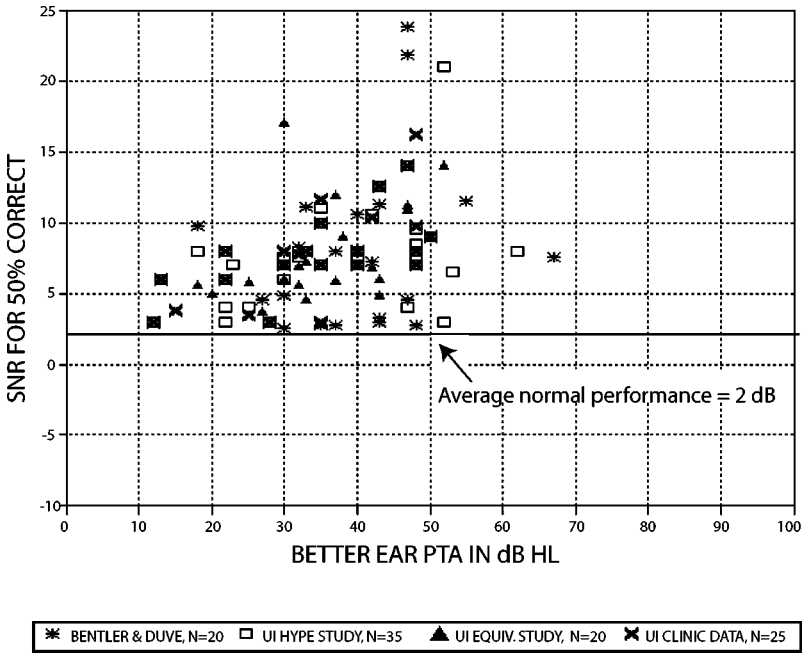


FIG. 1. Signal-to-noise ratio for 50% correct on the SIN test (70 dB HL presentation level) versus three-frequency average pure-tone hearing loss in the better ear (average of 0.5, 1, and 2 kHz). Four data sets obtained at the University of Iowa Speech and Hearing Clinic. From Killion and Niquette (2000), with permission.

vantage that it is less representative of everyday speech-in-noise situations than babble noise.

The purpose of the present experiments was to develop a speech-in-noise test which would (1) estimate SNR loss in one to two minutes, (2) be easy to administer, (3) have good face validity, (4) have simplified scoring, and (5) provide list equivalency for both normal and hearing-impaired subjects.

When designing a speech-in-noise task, the choice of speech and background noise materials is a compromise between realism and reproducibility. Monosyllabic words, recorded and played back at uniform intensity levels, are not representative of speech in the real world. Sentences spoken with natural dynamics have greater dynamic range than monosyllabic words, and are thus a more valid representation of real speech (Villchur, 1982). Furthermore, the effects of coarticulation are not well represented on monosyllabic word lists.

Likewise, a constant-level background noise, while easy to control and reproduce, is not typical of that encountered by most people in their everyday environments. Fikret-Pasa (1993) examined the intensity variations as a function of time of the background noise encountered in everyday situations (e.g., shopping malls and crowded restaurants) and found level variations having standard deviations of 2.8 to 8.4 dB, for maximum and minimum sound level meter readings, respectively. In contrast, Fikret-Pasa measured virtually no variation in level in available speech-spectrum-noise maskers, and only a 1-dB variation in level in two examples of multi-talker babble, both of which contained so many talkers that the result was a constant-level murmur. She found that the Auditec four-talker babble (Auditec of St. Louis, 1971) had more level variations than any of the other commercially available noises, presumably because the babble talkers were instructed to speak naturally (Carver, 1991). Use of a background noise with level variations is particularly important for a test used with compression hearing aids, so that the compression circuits are not clamped in a fixed-gain setting by the noise.

Fikret-Pasa (1993) also chose four-talker babble because it represents a realistic simulation of a social gathering, in which the listener may tune out the target talker and tune in one or more of the other nearby talkers using what Broadbent (1958) labeled ''selective listening.'' More subtly, the use of constant-level noise in speech-reception research eliminates the temporary gaps in the noise of real talkers, gaps which those with normal hearing appear able to make use of when listening in noise (Bacon *et al.*, 1998).

The SIN test (Killion and Villchur, 1993; Etymotic Research, 1993) was developed for estimating the signal-to-noise ratio required by hearing impaired subjects to obtain adequate intelligibility under different hearing aid processing conditions. In accordance with standard threshold procedures, the signal-to-noise ratio required for 50% correct words in sentences is abbreviated as ''SNR-50.''

The SIN test combined a Massachusetts Institute of Technology recording (female talker) of the IEEE sentences [Institute of Electrical and Electronics Engineers (IEEE), 1969] as signal, and a four-talker babble (Auditec of St. Louis, 1971) as competing noise. Each IEEE sentence has five key words, which are scored as correct or incorrect. Rabinowitz *et al.* (1992) reported experiments with cochlear implant subjects indicated that the IEEE sentences are comprised of words that are not highly predictable from the surrounding context, resulting in a performance-intensity function closer to that obtained with word scoring than with whole-sentence scoring.

The ten-sentence IEEE lists were reportedly phonetically balanced (IEEE, 1969). However, no attempt was made to maintain these list groupings on the SIN test. Egan (1948) observed that the distribution of sounds in speech depends on the topic being discussed and who is speaking. The phonetic balancing of a list of words or words-in-sentences is influ-

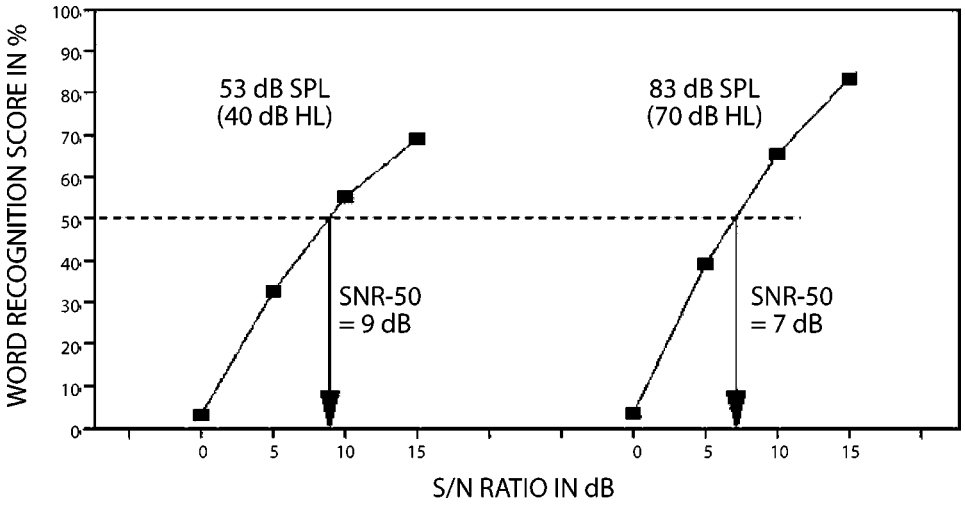


FIG. 2. Illustration of the graphical scoring method used in the SIN test. Each SNR level contains 25 words, so the number correct is multiplied by 4 to arrive at the percent-correct score. The percent correct is then plotted for each SNR. At the point where the 50% line is crossed, a vertical line is drawn down. The SNR-50 is then interpolated by the value given on the *X* axis.

enced by the talker; when the same words or sentences are recorded using a different talker, the original phonetic balancing is not maintained. Hood and Poole (1980) demonstrated this conclusively when they selected ''difficult'' and ''easy'' words, and found that ''words that were formerly difficult became easy and vice versa'' when recorded by a second speaker. Recently, Martin *et al.* (2000) found that words selected randomly from a dictionary could not be statistically distinguished from the phonetically balanced NU-6 words (Tillman and Carhart, 1966). Since the IEEE sentences had no standard talker, it seemed likely that two sets of phonetically balanced sentences might produce significantly different average intelligibility when spoken by different talkers, which is what was found in early SIN test experiments.

The SIN test uses the first 360 of the 720 IEEE sentences (lists 1–36), divided into nine blocks of 40 sentences each. Each SIN test block contains two sections, the first administered at a 70 dB HL presentation level, and the second at 40 dB HL. These levels were chosen to represent the range of typically loud and quiet speech levels encountered by most people in everyday life. Each section contains 20 sentences, five sentences at each signal-to-noise ratio of 15, 10, 5, and 0 dB. Each sentence contains five key words, which are scored as correct or incorrect, resulting in 25 key words at each SNR. The key words repeated correctly are summed for each SNR and multiplied by 4 to obtain a percent-correct score. The percentage scores are manually plotted for each SNR, using a graph with SNR on the abscissa and percent correct on the ordinate. The SNR-50 is interpolated by drawing a horizontal line at the 50% point, and dropping a vertical line at the SNR that intersects the 50% line. As shown in Fig. 2, scores are plotted separately for the 70 and 40 dB HL presentations.

Test time for a single SIN test block (using both of the recommended presentation levels of 70 and 40 dB HL) is approximately 6 min if the subject responds promptly. While the SIN test offered good face validity, some practitioners reported that it was too time-consuming for clinical use and that the graphical scoring for the test was difficult. *Posthoc* analysis of the SIN test (Killion *et al.*, 1996; Bentler, 2000) revealed that not all of the test blocks were equivalent, resulting in too few lists for some clinical comparisons and research purposes. Both floor and ceiling effects were also noted (Bentler, 2000).

Killion *et al.* (1996) reported that on blocks 3, 4, 5, 6, and 8 of the SIN test, normal-hearing listeners, as a group, produced similar SNR-50 estimates across blocks, and hearing-impaired listeners, as a group, also produced similar SNR-50 estimates across blocks. They reported that the standard deviation of a *single* block test score is 0.8 dB, based on test-retest data. Thus the critical difference for a two-way comparison at the 95% confidence level is 2.2 dB, using one half-block at a single presentation level (e.g., 70 dB HL) for each condition, and 1.6 dB using two half-blocks for each condition. To explain, for the average of two lists, the predicted 1.6 dB critical difference at the 95% confidence level is equal to 1.96*(sqrt2*0.8)/sqrt2, where the second sqrt2 refers to the use of two lists.

Bentler (2000) reported that SIN test blocks 1, 2, and 9 were equivalent to each other and that blocks 3, 4, and 5 were equivalent to each other. Test-retest correlations were high, with a 95% confidence-level critical difference between two single-block scores (single levels) of 2.4 dB (equivalent to a single-block test score standard deviation of 0.87 dB). Cox *et al.* (2001) developed a revised version of the SIN test, the RSIN, which reallocated the recorded SIN test materials into four ''modified dual blocks'' based on the data from Bentler (2000). Cox *et al.* reported a critical difference of 1.7 dB SNR for these RSIN dual blocks. The standard deviation inferred from their report for a *single-block* test score is thus also 0.87 dB.

From these three reports, it would appear that a conservative value for the standard deviation of one single-presentation-level SIN test block would be 0.85 dB. This value was used in the design of the four experiments conducted during the development of the QuickSIN test. Since the SIN test employed five sentences at each SNR and the QuickSIN was to employ only one, the predicted standard deviation for a single QuickSIN test score was 0.85*sqrt5 or 1.9 dB.

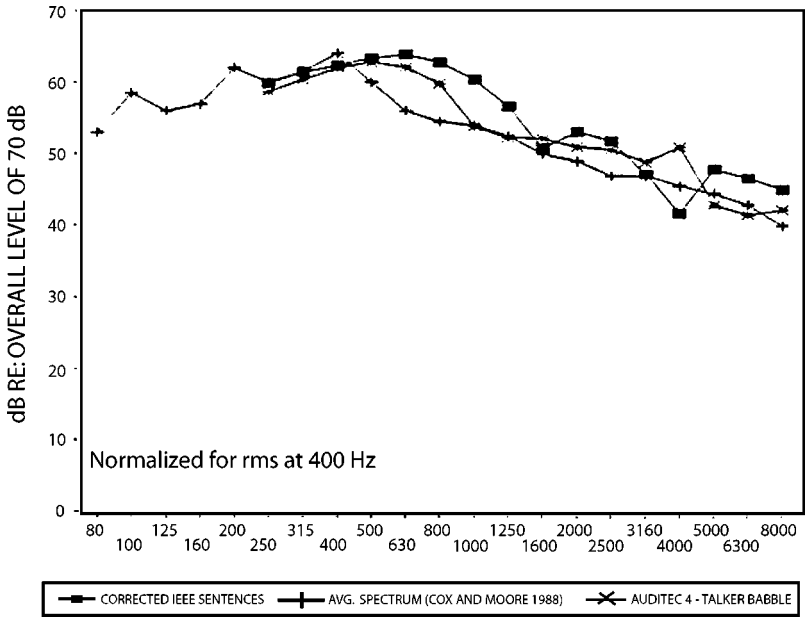


FIG. 3. Long-term average one-third-octave spectrum of IEEE sentences after filtering for microphone placement artifact, compared to measured long-term average, one-third-octave spectrum of Auditec four-talker babble and Cox and Moore (1988) average spectrum for continuous male speech. From Fikret-Pasa (1993), with permission.

II. EXPERIMENT 1

The purpose of experiment 1 was to normalize the SNR of the sentence-babble pairs, so that after normalization each pair had the same SNR-50 for normal-hearing subjects.

A. Method

1. Subjects

Sixteen adult subjects with normal hearing participated in experiment 1.[2] Normal hearing was defined as pure tone thresholds equal to or better than 20 dB HL for octave frequencies 250 to 4000 Hz. Subject ages ranged from 18 to 51 years, with an average age of 24 years and a median age of 20 years.

2. Stimuli

The same equalized MIT recording of a female talker and the same four-talker babble described above were used here, except the sentences were the second group of 360 IEEE sentences, lists 37–72 (recall that the first group of 360 sentences was used in the SIN test). A spectral analysis of the equalized female talker, the four-talker babble, and Cox and Moore's (1988) average spectrum for continuous male speech is shown in Fig. 3.

An ''Alpha 1'' master digital recording of the sentences was made, with the MIT female talker on channel 1 and the four-talker Auditec babble on channel 2. The instantaneous levels of both the MIT female talker and the babble talkers ebb and flow at any presentation level; as a result, the SNR required for 50% correct in a given sentence depends on the babble with which that sentence is paired. In the two-channel master recording of the sentences and babble, each pair was time locked, meaning that the time relationship between each sentence and its corresponding babble segment was fixed. The master recording was made so that all subsequent rerecordings of a given sentence had the same time-locked relationship between speaker and babble segments.

In previous studies using the SIN test, Killion *et al.* (1996) and Bentler (2000) found that normal-hearing subjects score 50% correct at approximately a 2-dB SNR. In experiment 1 the prerecorded calibration tones found on the MIT sentence recordings and on the Auditec babble recordings were used to set a nominal 2-dB signal-to-noise ratio for each sentence-babble combination.[3] Additional recordings of all 360 sentences were made at -1 dB SNR and $+5$ dB SNR in the same manner. The result was three Alpha 1 CDs, one each at nominal average signal-to-noise ratio of -1, $+2$, and $+5$ dB, which covered the range of likely SNR-50 values for each sentence-babble pair.

3. Stimulus presentation

In all experiments, speech materials were routed through the speech channel of standard clinical audiometers, and testing took place in sound-treated test booths. Each CD contained a calibration track, and VU meters were set for ''0'' using the calibration tone prior to testing.

The Alpha 1 sentences were presented at 70 dB HL via ER-3A insert earphones. Twelve subjects were tested monaurally, and four subjects were tested binaurally. Prior to the test session, three prototype lists were administered at 70 dB HL to familiarize the subjects with the task. Each prototype list consisted of six IEEE sentences (IEEE, 1969), female talker, in four-talker babble (Auditec of St. Louis, 1971) with one sentence at each SNR of 25, 20, 15, 10, 5, and 0 dB. The 360 Alpha 1 sentences were then presented at -1, $+2$, and $+5$ dB SNR, in that order.

4. Scoring

One point was given for each of five key words repeated correctly in each sentence. Half credit was given for words close to the target word, e.g., ''cat'' for ''cats.'' The SNR-50 was calculated for each sentence using a formula based on the Tillman–Olsen (1973) recommended method for obtaining spondee thresholds.[4] These scoring methods were used in all four experiments.

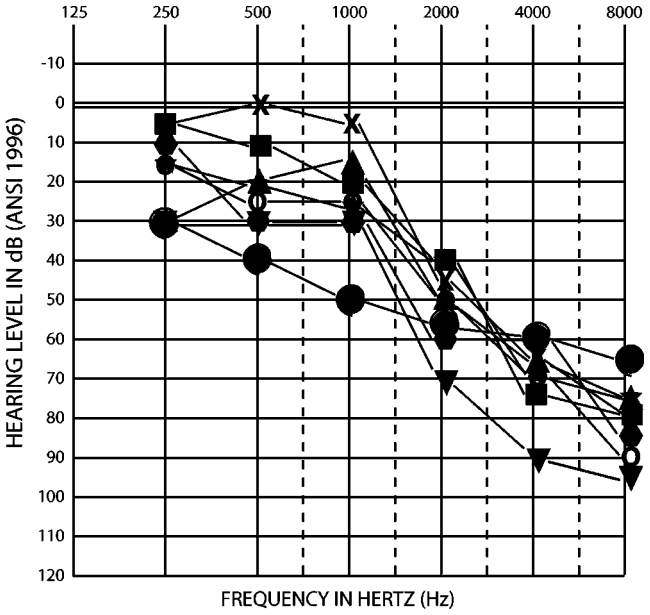


FIG. 4. Composite audiogram for eight hearing-impaired adult subjects from experiment 2.

B. Results and discussion

Experiment 1 provided, for each sentence-babble combination, SNR-50 values for each subject. The across-subject average SNR-50 for the 360 Alpha1 IEEE sentences was +2.5 dB, and the standard error of the mean for the 16 subjects was 0.3 dB. The experimentally derived SNR-50 value for each of the time-locked sentence-babble combinations ranged from −1.3 to +5.9 dB across the 360 sentences. The subject-data-based SNR-50 values thus often differed substantially from the nominal values based on the prerecorded calibration tones. In addition to the ebb and flow in the talker and babble, some sentences were simply easier than others. The individual-sentence subject-data-based SNR-50 values formed the basis for the recordings in experiment 2.

III. EXPERIMENT 2

The purpose of experiment 2 was to assess the equivalence of the 360 time-locked sentence-babble combinations after they had been readjusted for SNR equivalence based on the data from experiment 1.

A. Method

1. Subjects

Six normal-hearing adult subjects and eight hearing-impaired adult subjects with symmetrical sloping, severe high-frequency sensorineural hearing loss participated in experiment 2. Normal hearing was defined as pure tone thresholds equal to or better than 20 dB HL for octave frequencies 250 to 8000 Hz. The normal-hearing subjects ranged in age from 20 to 23 years, with an average age of 22 years and a median age of 22 years. Figure 4 shows the audiograms of the eight hearing-impaired subjects; thresholds represent the test ear for the seven subjects who were tested monaurally, and the better ear for the subject who was tested binaurally. The hearing-impaired subjects ranged in age from 60 to 78 years, with an average age of 69 years and a median age of 66 years.

2. Stimuli

A new set of recordings was made based on the SNR-50 values from experiment 1. For these recordings, the values from experiment 1 were used to readjust the recorded SNR of each sentence-babble combination to bring it to an expected value of 2 dB [the average SNR-50 for normal-hearing adults at a 70 dB HL presentation level on the SIN test (Bentler, 2000)]. For example, sentence 1 on list 37 had an SNR-50 of 3.5 dB in experiment 1, so the level of the babble associated with this sentence was reduced by 1.5 dB to produce an expected SNR-50 of 2 dB. A new master recording was made using the readjusted babble levels, preserving the previous time-locked relationship between sentences and babble. From this master, a set of ''Alpha 2'' CD recordings was made of all sentences, with the sentence-to-babble ratio readjusted to nominal SNR values of 0, +5, +10, and +15 dB.

3. Stimulus presentation

The Alpha 2 recordings were presented monaurally at 70 dB HL as described in experiment 1 for subjects with normal hearing, and at a presentation level judged to be ''Loud, but O.K.'' (just below discomfort) for subjects with hearing loss (Valente and Van Vliet, 1997). These presentation levels were chosen to ensure that nearly all speech cues were audible. Frequency response shaping was not used for hearing-impaired subjects, since previous research (Skinner, 1976; Dirks, 1982; van Buuren *et al.*, 1995) showed negligible differences in SNR at these high presentation levels as the frequency response was changed. Monaural presentation was used for seven of the eight hearing-impaired subjects.

The normal-hearing subjects listened to the lists at the 0 and +5 dB SNRs, since they were expected to obtain nearly 100% correct at +10 dB SNR and above. To determine the appropriate SNRs to use for testing the hearing-impaired subjects, two half-blocks of the SIN test were administered to each subject. The resulting SNR-50 value was used to determine which two Alpha 2 CDs to use. For example, the +5 and +10 dB SNR recordings were used for a subject with an SNR-50 of 7 dB on the SIN test. In all cases, the lists with the lower (most difficult) SNR were administered first.

B. Results and discussion

The across-subject average SNR-50 for the six normal-hearing subjects was 2.4 dB, and the across-sentence average of the across-subject standard deviations was 1.6 dB, giving a standard error of the mean for the six normal-hearing subjects of 0.65 dB. The six-subject average value of 2.4 dB was not significantly different than the expected value of 2 dB (for this small sample, the ±95% confidence interval around the average value of 2.4 dB was 1.1 to 3.7 dB). The across-subject standard deviation was small, indicating good equivalence for the readjusted sentence-babble combinations.

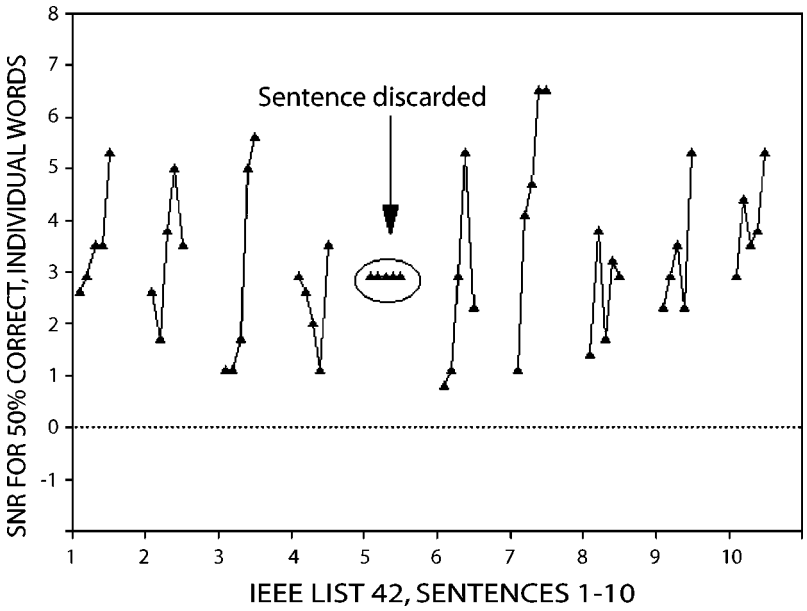


FIG. 5. SNR-50 data for individual words in IEEE list 42, sentences 1–10, derived from an analysis of data from five normal-hearing subjects from experiment 1. Sentences in which the SNR-50 for each of the five key words varied by less than or equal to 2 dB were discarded. Sentence 5 illustrated an unusual case in which all words had identical SNR-50 values.

The across-subject SNR-50 for the eight sloping-hearing-loss subjects was 7.4 dB, and the across-sentence average of the across-subject standard deviations was 2.9 dB. The higher across-subject standard deviation for the hearing impaired subjects resulted from the wide range of apparent SNR losses they exhibited. These data were used in the sentence selection process of experiment 3.

IV. EXPERIMENT 3

The purpose of experiment 3 was to assess the equivalence of selected time-locked sentence-babble combinations, using subjects with normal hearing and subjects with hearing loss. The selection of these sentences was made based on data from experiments 1 and 2.

A. Method

1. Subjects

a) Group 1. Twenty-six normal hearing adult subjects, 15 subjects from 12 Beta sites and 11 subjects from the University of Iowa Speech and Hearing Clinic, participated in experiment 3. Normal hearing was defined as pure tone thresholds equal to or better than 20 dB HL for octave frequencies 250 to 4000 Hz. All subjects met this criterion except two; one subject had a 25 dB HL threshold at 250 Hz in one ear, and another subject had a threshold of 30 dB HL at 4000 Hz in one ear. Subjects ranged in age from 20 to 58 years, with an average age of 32 years and a median age of 27 years.

b) Group 2. Eighteen hearing-impaired adult subjects from ten Beta sites participated in experiment 3. Hearing impaired subjects had symmetrical sensorineural hearing losses. Audiometric criteria for the hearing-impaired subjects were (*i*) mild to moderate sloping loss, with a minimum 20-dB drop from 500 to 4000 Hz; (*ii*) mild to moderate flat loss, with less than a 15-dB drop from 500 to 4000 Hz; or (*iii*) severe loss, with thresholds from 60 to 90 dB for frequencies 500 to 4000 Hz. Subjects ranged in age from 20 to 77 years, with an average age of 57 years and a median age of 58 years.

2. Stimuli

Data from experiments 1 and 2 were used to select sentences that were equivalent for normal-hearing and hearing-impaired subjects. Selection criteria were the following.

(i) Better than average across-subject variability in sentence scores: Only sentences with an across-subject standard deviation of less than 1.5 dB were retained, based on experiment 2 data from six normal-hearing subjects. Since the criterion was 0.1 dB below the *average* of 1.6 dB, slightly more than half of the sentences were discarded as having greater than average across-subject standard deviation.[5]

(ii) The across-subject average SNR-50 value for the normal-hearing subjects was within 1.5 dB of the grand average value for each retained sentence, based on experiment 2 data.

(iii) The across-subject average SNR-50 value for the hearing-impaired subjects fell within 2 dB of the grand average value for each selected sentence, based on experiment 2 data from eight subjects with sloping, high-frequency sensorineural hearing loss.

(iv) The SNR-50 value was calculated for each of the 1800 key words found in the 360 sentences, using data from experiment 1 for five randomly selected subjects from the 16 normal-hearing subject data pool. Data from five subjects were considered adequate, since it resulted in 15 data points for each word, or 27 000 total data points. For a sentence to be retained, the range of SNR-50 values across the five key words in the sentence had to be greater than 2 dB (see Fig. 5).

(v) Sentences containing language that is no longer contemporary were eliminated. For example, ''The smell

of burned rags itches my nose'' and ''The vamp of the shoe had an old buckle'' were eliminated.

The purpose of the above steps was to select time-locked sentence-babble combinations that had good reliability across subjects and sentences. Some variation in SNR of words within sentences was desirable so that the sentences were more representative of everyday speech than sentences with nearly the same SNR for each word. These combined procedures eliminated 75% of the original 360 sentences, leaving 89 sentences meeting the criteria.

Beta recordings were made using 84 of the 89 selected sentence-babble combinations. This yielded 14 lists of six sentences each, with one sentence at each of the following signal-to-noise ratios: 25, 20, 15, 10, 5, and 0 dB. These were labeled Beta lists 1–14. Since additional lists were desired, the standard deviation limit in step *ii* (above) was increased from 1.5 to 2.0 dB. This change yielded enough additional sentences for seven more lists of six sentences each, labeled Beta lists 15–21.

3. Stimulus presentation

Beta lists were presented binaurally at 70 dB HL to normal-hearing subjects. Most subjects (19 of 26) were tested with ER-3A insert earphones; four subjects were tested in sound field, and three subjects with TDH headphones. At a presentation level of 70 dB HL, the quietest speech cues should have been at 40 dB HL. Even taking into account the field-referenced high-frequency rolloff of the TDH-39 and ER-3A earphones, no speech cues should have fallen below 30 dB HL. van Buuren *et al.* (1995) found no difference in intelligibility in noise for 18 of 25 frequency responses, covering most of the dynamic range of their hearing-impaired listeners, providing further evidence suggesting that the effect of the earphones should have been minimal. Hearing-impaired subjects were tested at 70 dB HL except when their three-frequency pure tone average (PTA) exceeded 50 dB HL, in which case they were tested at a level they judged to be ''Loud, but OK.'' The Beta list presentations were counterbalanced to control for potential order effects.

B. Results and discussion

The across-subject, across-list SNR-50 average for the 26 normal-hearing subjects in group 1 was 1.9 dB, nearly identical to the original SIN test average of 2 dB. As in experiment 2, the average fell well within the 95% confidence interval of the expected value: The single-list test-score standard deviation of SNR-50 was 1.25 dB for the normal-hearing subjects, giving a standard error of the mean of 0.25 dB.

The performance of the hearing-impaired subjects in group 2 was so diverse that it precluded extraction of the desired normative data. Our motivation for creating an efficient SNR-loss test was the finding that SNR performance is not predictable from hearing thresholds (Killion and Niquette, 2000). Only a few of our subjects had SNR-50 scores at the same SNRs, and none had SNR-50 scores at the higher SNRs (15, 20, and 25 dB). Thus, it was not possible to assess the SNR equivalence of the lists for the higher SNRs. The single-list test-score standard deviation (derived from individual test-retest scores) was 1.4 dB, however, which was an encouraging result since we had predicted 1.9 dB from SIN test data. Nonetheless, the results from the 18 hearing-impaired subjects in experiment 3 made it clear that an exceedingly large number of hearing-impaired subjects would need to be screened in order to obtain enough data for statistically useful tests of list equivalence at the higher signal-to-noise ratios. We therefore chose a different approach in experiment 4.

V. EXPERIMENT 4

The purpose of experiment 4 was to assess list equivalence for five of the six signal-to-noise ratios used in the test (25, 20, 15, 10, and 5 dB).

A. Method

1. Subjects

Twenty-five normal-hearing adult subjects (22 from the University of Iowa Speech and Hearing Clinic and 3 from the Northwestern University Speech and Hearing Clinic) participated in experiment 4. None of these subjects participated in experiments 1, 2, or 3. Normal hearing was defined as pure tone thresholds equal to or better than 20 dB HL for octave frequencies 250 to 4000 Hz. One subject failed this criterion in one ear, with thresholds of 30 and 25 dB HL at 2000 and 3000 Hz, respectively. Subjects ranged in age from 18 to 49 years, with an average age of 26 years and a median age of 24 years.

2. Stimuli

To simulate a wide range of SNR losses, the Beta-list recordings were low-pass filtered. Low-pass cutoff frequencies of 2000, 1400, 1100, 850, and 750 Hz were chosen to provide expected SNR-50 scores of 5, 10, 15, 20, and 25 dB. The choice of cutoff frequencies was made using the count-the-dot articulation index method (Mueller and Killion, 1990) and the relationship between AI and SNR-50 (Killion and Christensen, 1998, Fig. 7). The procedure was as follows: (1) a desired SNR-50 value was chosen; (2) the Killion and Christensen (1998) data were used to find the corresponding AI value; and (3) a low-pass filter cutoff frequency was selected using a trial and error method to filter out the required number of ''dots.'' For example, filtering out all speech sounds above 1100 Hz with a fairly steep filter slope will remove 63% of the speech cues (leaving 37% audible, or an articulation index of 0.37). Figure 7 of Killion and Christensen (1998) indicates that 63 ''missing dots'' should correspond to a 15 dB SNR-50, whether the cues are inaudible as a result of filtering or hearing loss. As shown in their Fig. 7, the average measured SNR-50 of our subjects under the various filtering conditions was closely correlated with the predicted value. By using these filtered sentences with normal-hearing subjects, only the lack of audible speech cues could be expected to affect the measured SNR-50 values.

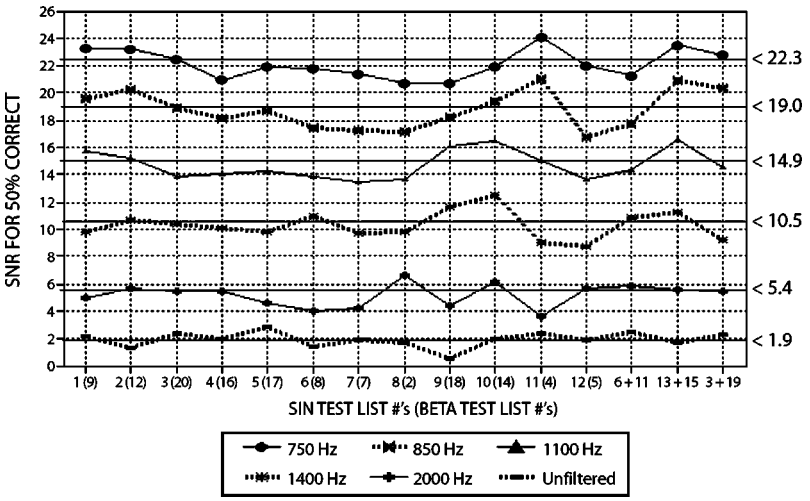


FIG. 6. Across-subject average SNR-50 data for lists included in the final QuickSIN CD. The original list numbers are given in parentheses next to the new list numbers shown on the abscissa.

3. Stimulus presentation

The filtered recordings were presented binaurally at 70 dB HL via ER-3A insert earphones. Subjects were tested in three sessions over several days, and list presentation order was randomized to minimize potential order effects. The most difficult condition (750 Hz low pass) was presented first, followed in order by the less difficult conditions (850, 1100, 1400, and 2000 Hz). Learning effects were not expected to significantly affect the subjects' performance because (a) the testing took place over several days and (b) few words were understood in the more difficult filtering conditions. When similar lists (unfiltered) were retested on the *same* day, Chung (2001) reported learning effects of 1.6 to 3.6 dB, which suggests that learning effects should have little effect on the present results, where each successive list presentation was made at a 5-dB-greater SNR.

B. Results and discussion

1. Selection of lists

The final selection of equivalent lists was made using normal-hearing data from experiment 3 and the filtered data from experiment 4. Twelve lists had experimental SNR values that fell within ± 2.2 dB of the average in each condition (unfiltered and all filtered conditions) and became lists 1–12 on the final QuickSIN CD. Three pairs of lists were also found whose *pair averages* met the above criteria. Typically one list score in a pair would be high and the other would be low under similar conditions. These became list pairs 13/14, 15/16, and 17/18 on the final QuickSIN CD. Three lists (labeled A, B, and C) failed the equivalency criteria and were included on the final QuickSIN CD for practice only. These lists can be used to familiarize the listener with the task prior to testing.

Figure 6 shows the across-subject average SNR-50 data for the lists included on the final QuickSIN CD. The original list numbers are given in parentheses next to the new list numbers shown on the abscissa.

2. Application of test

SNR loss is defined as the difference between the test subject's threshold, and average-normal threshold: More precisely SNR loss is equal to the test subject's SNR-50 in dB minus the average-normal SNR-50 in dB. Since the average normal SNR-50 on the QuickSIN test is 2 dB, the simple formula for SNR loss is SNR loss $= 25.5 -$ Total Correct.

The time interval between sentences in the final recording is 5 s, providing adequate response time for most subjects. Each QuickSIN list takes approximately 1 min to administer if the subject responds promptly. If the subject requires more time to respond, the CD can be paused between sentences.

3. AI predictions versus subject performance

The correlation between articulation-index calculated results and 25-subject average results was good. Recall that the expected values of SNR-50 for the filtered lists were 25, 20, 15, 10, and 5 dB, based on the AI predictions. The results from experiment 4 were similar to the predicted values: 22.5, 19.0 14.9, 10.5, and 5.4 dB. Figure 7 shows the correlation between articulation-index calculated results and 25-subject average results.

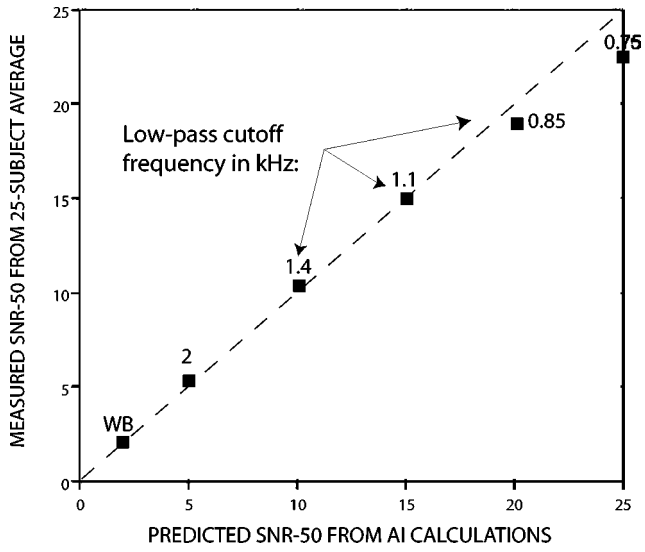


FIG. 7. Comparison between average SNR-50 for normal-hearing subjects listening to filtered speech in noise and predicted values based on the count-the-dots version of the articulation index.

TABLE I. Reliability: number of QuickSIN lists required for a given accuracy. An 80% confidence interval is normally adequate for clinical testing, where the results of any one test are used in context with other factors. A single QuickSIN list is accurate to ±2.2 dB at the 80% confidence interval. A 95% confidence interval is common for research reporting, where a reduced risk of error is normally required. In this case, a single QuickSIN list is accurate to ±2.7 dB at the 95% confidence interval.

	No. of lists								
	1	2	3	4	5	6	7	8	9
95% C.I.+(dB)	2.7	1.9	1.6	1.4	1.2	1.1	1.0	1.0	0.9
80% C.I.+(dB)	2.2	1.6	1.3	1.1	1.0	0.9	0.8	0.8	0.7

VI. GENERAL DISCUSSION

A pure tone audiometric threshold, measured using the standard Hughson–Westlake technique, is accurate to about 5 dB at the 80% confidence level. (1.28×SDEV; Witting and Hughson, 1940). An 80% confidence level is normally adequate for clinical audiometric testing, where the results of any one test are used in context with other factors. A 95% confidence level is common for research reporting, where a reduced risk of error is normally required. A QuickSIN score obtained in 1 min from a single list is accurate to ±2.2 dB at the 80% confidence level and ±2.7 dB at the 95% confidence level.

Table I shows the number of lists required for a given accuracy (such as ±2.2 dB) for confidence levels of 80% and 95%. The numbers in Table I are based on the across-subject root-mean-square standard deviation of 1.4 dB SNR for single-list test scores found for the hearing-impaired subjects in experiment 2. This value comes from two numbers: (a) the 1.3 dB standard deviation derived from the combined individual *test-retest* scores (441 comparisons), and (b) the *across-list* standard deviation of 0.6 dB. If only normal-hearing subjects are used, the single list test score standard deviation drops from 1.4 to 1.25 dB.

A standard deviation of 1.4 dB is slightly better than the standard deviation that would have been expected based on the SIN test, which employs five sentences at each SNR level. As discussed above, the single-block SIN test score standard deviation of 0.85 dB multiplied by the square root of 5 would predict a standard deviation of 1.9 dB for the QuickSIN test, because the QuickSIN test uses only one sentence at each SNR level. The more careful preselection of sentences used in the QuickSIN test may have contributed to the slightly better result.

Averaging the results of several QuickSIN lists improves reliability compared to a single list. For example, one list with the assumed standard deviation of 1.4 dB gives an 80% confidence level of 1.6×1.4=±2.2 dB. A 50% increase in testing is required to improve from a confidence level of 80% to 95% at a given criterion, e.g., two lists give an 80% confidence interval of ±1.6 dB; three lists provide ±1.6 dB at the 95% confidence level.

The use of multiple lists is particularly important when QuickSIN lists are used to compare two conditions (often two hearing aids or hearing aid adjustments). In this case, the real differences may not be large. A difference of 3.9 dB can be expected 5% of the time when two successive lists are administered, for example, *even if the lists are perfectly equivalent*.

TABLE II. QuickSIN critical differences for comparing two conditions (e.g., two hearing aids or hearing aid adjustments). If one list per condition is used, results must be greater than 3.2 dB different to be considered statistically significant at the 80% confidence interval, and 3.9 dB at the 95% confidence interval.

	Lists per condition								
	1	2	3	4	5	6	7	8	9
95% C.D.+(dB)	3.9	2.7	2.2	1.9	1.7	1.6	1.5	1.4	1.3
80% C.D.+(dB)	3.2	2.2	1.8	1.6	1.4	1.3	1.2	1.1	1.1

The preceding result illustrates the difficulty in validating the equivalence of lists to better than 2 dB. With a standard deviation of 1.4 dB for a single list, even if all 12 lists are administered to ten subjects and their SNR results are averaged, the lists with the highest and lowest scores across the 12 lists will differ by 1.8 dB 5% of the time. If all 12 lists are administered to a single individual, then by the Bonferoni inequality for multiple comparisons (Miller, 1966), a difference of 6.1 dB between the two lists with the highest and lowest measured SNR can be expected 5% of the time.

Table II shows the number of lists required for the comparison between two conditions at an 80% or 95% confidence level. For a critical difference of 1.9 dB, for example, four lists are required for each condition at the 95% confidence level. For a critical difference of 1.4 dB, eight lists are required for each condition at the 95% confidence level.

The use of half-word credit and phoneme scoring were debated during QuickSIN test development. Whole-word scoring was preferred in order to simplify the scoring procedure and to reduce the variability in scoring among clinicians. Achieving consistency in scoring without compromising test accuracy was the goal. Instructions for scoring the SIN test promoted awarding half-credit for partially correct answers because an analysis of the relative independence of IEEE words indicated that 25 words in five sentences, using half-word scoring, gives the equivalent of 27 independent words with whole-word scoring (Fikret-Pasa, 1993). Bentler (2000), however, observed poor interobserver reliability using partial-word scoring. By reexamining the test-retest data from the 26 normal-hearing subjects from experiment 3, whole-word and half-word scoring were compared. Results indicated only a slight improvement (from a standard deviation of 1.25 to a standard deviation of 1.21) when half-word scoring was used. These findings suggest that whole-word scoring on the QuickSIN test is adequate for clinical testing.

VII. CONCLUSIONS

The QuickSIN test provides 12 equivalent lists for testing normal-hearing and hearing-impaired subjects. The test is time efficient; the administration of a single list takes approximately 1 min. The standard deviation for an SNR estimate using a single list is 1.4 dB. Averaging multiple lists results in a lower standard deviation. In addition to the 12 equivalent lists, another six lists are equivalent to the 12 when used in designated list pairs.

ACKNOWLEDGMENTS

The four-talker babble recording was incorporated under license from Auditec of St. Louis. The IEEE recordings were incorporated under license from Massachusetts Institute of Technology. Ruth Bentler and Donna Devine provided much of the subject data for experiments 3 and 4. The following individuals contributed Beta-site data: Nancy Aarts, Harvey Abrams, Rose Allen, Pauline Bailey, Dawn-Marie Bass, Pamela Buethe, Jodi Cook, Paul Efros, David Hawkins, Erica Johnson, Krista Johnson, Melissa Kluck, Barbara Kruger, Lisa Lamson, Stacey Matson, Kristi Mohr, Susan Phillips, Tricia Roh, Juliette Sterkens, Becky Warner, and audiologists at the Eye and Ear Institute, Pittsburgh, PA.

[1]A 50% correct score on words in sentences typically corresponds to 90% correct on complete sentences (e.g., Killion and Christensen, 1998).

[2]After these experiments were completed, we discovered that English was not the first language of one of the subjects, who nonetheless spoke English fluently. After examining the data, we found that this subject's ability to understand speech in noise was indistinguishable from that of the other subjects and, more importantly, it would have made no difference in the results if that subject had been excluded. We thus retained her data.

[3]The calibration tones on the Auditec four-talker babble recording and on the MIT female-talker recording had been adjusted in accordance with Section 6.2.11 of ANSI S3.6-1996 (specification for audiometers): ''...the level of the rms sound pressure of a 1000 Hz signal [is to be] adjusted so that the deflection of the volume level indicator produced by the 1000 Hz signal is equal to the average peak deflection produced by the speech signal'' (ANSI, 1996). The same calibration method was employed in the NU-4 and NU-6 speech tests (Tillman *et al.*, 1963; Tillman and Carhart, 1966) and the SIN test (Etymotic Research, 1993).

[4]The Tillman–Olsen (1973) recommended method for obtaining spondee thresholds provides a simple method for estimating SNR-50 using the total number of words repeated correctly. In the Tillman–Olsen method, two spondees are presented at each level, starting at a level where all spondees are repeated correctly, then decreasing in 2-dB steps until no responses are obtained for several words. The starting level plus 1 dB, minus the total number of spondees repeated correctly, is the spondee threshold. The simple arithmetic comes from the use of 2-dB steps and two words per step. If the audiometer has 5-dB steps, the corresponding method would use five words per step and take the starting level plus 2.5 dB (half the step size), minus the total number of spondees repeated correctly.

[5]Although this procedure worked well on the average, some sentences with low-variability in sentence score were discarded and some sentences with high-variability were included by this procedure because the statistical variability in standard deviation estimates is moderately large.

ANSI (**1996**). ANSI S3.6-1996 ''American National Standard Specification for Audiometers'' (American National Standards Institute, New York).

Auditec of St. Louis (**1971**). ''Four-talker babble,'' 2515 S. Big Bend Boulevard, St. Louis, Missouri 63143-2105.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). ''The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds,'' J. Speech Lang. Hear. Res. **41**, 549–563.

Bentler, R. A. (**2000**). ''List equivalency and test-retest reliability of the Speech in Noise Test,'' Am. J. Audiol. **9**, 84–100.

Broadbent, D. E. (**1958**). *Perception and Communication* (Pergamon, New York).

Carver, W. (**1991**). Auditec of St. Louis, St. Louis, MO. Personal communication.

Chung, K. (**2001**). ''Effects of input-output functions on speech recognition and preference ratings (for subjects with normal hearing and with selected hearing losses),'' doctoral dissertation, Northwestern University, University Microfilms, Ann Arbor, MI.

Cox, R. M., and Moore, J. N. (**1988**). ''Composite speech spectrum for hearing aid gain prescriptions,'' J. Speech Hear. Res. **31**, 102–107.

Cox, R. M., Gray, G. A., and Alexander, G. C. (**2001**). ''Evaluation of a revised Speech in Noise (RSIN) test,'' J. Am. Acad. Audiol **12**, 423–432.

Dirks, D. D. (**1982**). ''Comments Regarding 'Speech Discrimination Ability in the Hearing-Impaired,' '' in *The Vanderbilt Hearing Aid Report*, edited by G. Studebaker and F. Bess (Monographs in Contemporary Audiology, Upper Darby, PA), pp. 44–50.

Dirks, D. D., Morgan, D. E., and Dubno, J. R. (**1982**). ''A procedure for quantifying the effects of noise on speech recognition,'' J. Speech Hear Disord. **47**, 114–123.

Egan, J. (**1948**). ''Articulation testing methods,'' Laryngoscope **61**, 891–909.

Etymotic Research (**1993**). ''The SIN Test,'' (Compact Disk) 61 Martin Lane, Elk Grove Village, IL 60007.

Fikret-Pasa, S. (**1993**). ''The effects of compression ratio on speech intelligibility and quality,'' doctoral dissertation, Northwestern University, University Microfilms, Ann Arbor, MI.

Hood, J. D., and Poole, J. P. (**1980**). ''Influence of the Speaker and Other Factors Affecting Speech Intelligibility,'' Audiology **19**, 434–455.

Institute of Electrical and Electronics Engineers (**1969**). ''IEEE recommended practice for speech quality measurements,'' Appendix C. Global Engineering Documents, Boulder, CO.

Killion, M. C. (**1997**). ''The SIN report: Circuits haven't solved the hearing-in-noise problem,'' Hear. J. **50**(10), 28–32.

Killion, M. C., and Christensen, L. A. (**1998**). ''The case of the missing dots: AI and SNR loss,'' Hear. J. **51**, 32–47.

Killion, M. C., and Niquette, P. A. (**2000**). ''What can the pure-tone audiogram tell us about a patient's SNR loss?'' Hear. J. **53**, 46–53.

Killion, M. C., and Villchur, E. (**1993**). ''Kessler Was Right—Partly: But SIN Test Shows Some Aids Improve Hearing in Noise,'' Hear. J. **46**(9), 31–35.

Killion, M. C., Olsen, W. O., Clifford, C. L., VanVliet, D. D., Rose, D. E., Bensen, D. E., Marion, M. W., Tillman, P. A., Hawkins, D. B., Dalzell, S. M., and Fabry, D. A. (**1996**). ''Preliminary data on the SIN Test,'' presented at the annual convention of the American Academy of Audiology, Salt Lake City, UT.

Kochkin, S. (**1992**). ''MarkeTrak III identifies key factors in determining consumer satisfaction,'' Hear. J. **45**(8), 39–44.

Kochkin, S. (**1993**). ''MarkeTrak III: Why 20 million in US don't use hearing aids for their hearing loss,'' Hear. J. **46**(1), 20–27.

Kochkin, S. (**1995**). ''Customer satisfaction and benefit with CIC hearing instruments,'' Hear. Rev. **2**(4), 16–24.

Kochkin, S. (**1996**). ''Customer satisfaction and subjective benefit with high performance hearing aids,'' Hear. Rev. **3**(12), 16–26.

Kochkiǹ, S. (**2000**). ''MarkeTrak V: 'Why my hearing aids are in the drawer: The consumers' perspective,'' Hear. J. **53**(2), 34–42.

Kochkin, S. (**2002**). ''10-Year customer satisfaction trends in the US hearing instrument market,'' Hear. Rev. **9**(10), 14–46.

Lyregaard, P. E. (**1982**). ''Frequency selectivity and speech intelligibility in noise,'' Scand. Audiol. Suppl. **15**, 113–122.

Martin, F. N., Champlin, C. A., and Perez, D. D. (**2000**). ''The Question of Phonetic Balance in Word Recognition Testing,'' J. Am. Acad. Audiol. **11**, 489–493.

Miller, R. (**1966**). *Simultaneous Statistical Inference* (McGraw–Hill, New York).

Mueller, H. G., and Killion, M. C. (**1990**). ''An easy method for calculating the articulation index,'' Hear. J. **43**(9), 14–17.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). ''Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise,'' J. Acoust. Soc. Am. **95**, 1085–1099.

Rabinowitz, W. M., Eddington, D. K., Delhorne, L. A., and Cuneo, P. A. (**1992**). ''Relations among different measures of speech reception in subjects using a cochlear implant,'' J. Acoust. Soc. Am. **92**, 1869–1881.

Skinner, M. W. (**1976**). ''Speech intelligibility in noise-induced hearing loss: Effects of high frequency compensation,'' doctoral dissertation, Washington University, University Microfilms, Ann Arbor, MI.

Strom, K. E. (**2003**). ''The HR 2003 dispenser survey,'' Hear. Rev. **10**(6), 522–538.

Taylor, B. J. (**2003**). ''Speech-in-noise tests: How and why to include them in your basic test battery,'' Hear. J. **56**(1), 40–46.

Tillman, T. W., and Carhart, R. (**1966**). ''An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University auditory test No. 6,'' SAM-TR-66-55.

Tillman, T. W., and Olsen, W. O. (**1973**). ''Speech audiometry,'' in *Modern Developments in Audiology (Second Edition)*, edited by J. Jerger (Academic, New York), pp. 37–74.

Tillman, T. W., Carhart, R., and Wilber, L. (**1963**). ''A test for speech discrimination composed of CNC monosyllabic words (N.U. auditory test No. 4),'' SAM-TDR-62.
Valente, M., and Van Vliet, D. D. (**1997**). ''The independent hearing aid fitting forum (IHAFF) protocol,'' Trends Amplif. **2**(1), 6–35.
van Buuren, R. A., Festen, J. M., and Plomp, R. (**1995**). ''Evaluation of a wide range of amplitude-frequency responses for the hearing impaired,'' J. Speech Hear. Res. **38**, 211–221.
Villchur, E. (**1982**). ''The evaluation of amplitude-compression processing for hearing aids,'' in *The Vanderbilt Hearing Aid Report*, edited by G. Studebaker and F. Bess (Monographs in Contemporary Audiology, Upper Darby, PA), pp. 141–143.
Witting, E. G., and Hughson, W. (**1940**). ''Inherent accuracy of a series of repeated clinical audiograms,'' Laryngoscope **50**, 259–269.

Speech activity detection and enhancement of a moving speaker based on the wideband generalized likelihood ratio and microphone arrays

Ilyas Potamitis[a)]
Wire Communications Laboratory, Electrical and Computer Engineering Department, University of Patras, Sofocleous-Adiparou 1, 265 00 Rion, Patras, Greece

Eran Fishler[b)]
Electrical Engineering Department, Princeton University, Princeton, New Jersey

(Received 15 August 2003; accepted for publication 11 June 2004)

The subject of this work is a unifying treatment of estimating the Direction of Arrival (DOA), detecting speech activity and suppressing noise in the case of a moving speaker by using a linear microphone array. The approach is based on the generalized likelihood ratio test applied to the framework of far-field, wideband moving sources (W-GLRT). It is shown that under certain distributional assumptions the W-GLRT provides a framework for the evaluation of DOA measurements against spurious DOAs, probabilistic speech activity detection as well as speech enhancement. As regards speech enhancement, we demonstrate the direct connection of W-GLRT with enhancement based on subspace methods. In addition, through the concept of *directive a priori SNR* we demonstrate its indirect connection with Minimum Mean Square Error spectral (MMSE_SA) and log-spectral gain modification (MMSE_LSA). The efficiency of the approach is illustrated on a moving speaker when either additive white Gaussian or babble noise is present in the acoustical field at very low SNRs. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1781622]

PACS numbers: 43.72.Ew [DOS] Pages: 2406–2415

I. INTRODUCTION

Microphone arrays have received considerable attention during the last decade as a potential replacement for headset and directional microphones when used in speech communication applications. This is due to the fact that the spatial selectivity of their receptive field (beamforming process) has the advantage over one-channel based techniques of suppressing interfering talkers and localized noise sources, and, to certain extent, reducing reverberation. Therefore microphone arrays are good candidates for a wide field of applications that includes videoconferencing,[1] multimedia conferencing,[2,3] speech recognition with regard to distant talkers,[4] speech enhancement for coding applications (e.g., hands-free mobile services).[5] These kinds of applications are greatly simplified if speech communication is based on hands-free acoustic interfaces that can focus on desired talkers while simultaneously suppressing noise and interfering talkers.

The output of the beamformer is a filtered estimate of the time-aligned microphone outputs where coherence of signals is achieved by using the estimated directions of arrival (DOAs) of the speakers. The DOAs can be estimated by high resolution spectral estimation techniques like MUSIC and Minimum Variance, or by finding the maximum power of a steered response over a range of angles or by first estimating the time delay of arrival between sets of microphones.[6]

Beamforming on a moving speaker imposes some restrictions on the way microphone outputs are processed. In order to avoid unrealistic stationarity assumptions one has to allow the derivation of the spatial correlation matrix from small numbers of frames (snapshots). However, speech pauses, reverberation, background noise, and the lack of long data segments to perform some kind of ensemble averaging leads to estimation of unreliable angle fixes. Although the beamforming process as such provides some degree of enhancement against directive interference, it is beneficial to incorporate noise suppression through an adaptive noise cancellation stage or postprocessing after the focusing procedure (see Ref. 6 and the references therein).

In this work, we present a unifying framework for DOA evaluation, voice activity detection, and speech enhancement based on setting a ''speech presence/absence hypothesis'' of the speech signal reaching the array and by forming the generalized log-likelihood ratio over all frequency bands (W-GLRT) while exploiting the spatial information provided by the array. The W-GLRT detector assumes that in each spectral bin both the source (if present) and the additive noise are zero-mean, i.i.d. complex Gaussian probability distributions with unknown variance. Assuming independence between spectral components the case of a single Gaussian point source embedded in additive Gaussian noise[7] is generalized to the case of wideband signals.

The broadband scanning scheme that estimates the DOA of a far-field moving source is based on finding the angle of arrival over the whole range of angles that minimizes a parameter of the W-GLRT.

By using the W-GLRT a decision between the null hy-

a)Electronic mail: potamitis@wcl.ee.upatras.gr
b)Electronic mail: efishler@ee.princeton.edu

pothesis ''noise only'' and the hypothesis ''speech is present'' is made. In this way, the concept of W-GLRT is also used to either associate a DOA measurement to an active speaker and allow the beamforming procedure to carry on, or to reject the DOA measurement and switch off the microphones in a software sense [i.e., act as a voice activity detector (VAD)]. One should note that the majority of VAD methods are based on single channel techniques (see Refs. 8–10, among many others) that focus on the spectral characteristics of the speech signal, while the proposed approach exploits both spectral and spatial characteristics of the sound source. There are few techniques reported in the literature for voice activity detection in the context of microphone arrays (see Refs. 11–13). Compared to the latter methods that are mainly based on a heuristic methodology our approach has the clear advantage of a rigorous probabilistic basis. Moreover we show that the probability of speech absence/presence which serves as statistical voice activity detector (VAD) and as a gain modification rule can be directly related to the W-GLRT.

Finally, it is demonstrated that under the distributional assumptions of W-GLRT the maximum likelihood estimation of the speech variance has a signal subspace interpretation allowing the direct reconstruction of the enhanced signal. Moreover, the introduction of the *directive a priori SNR* concept allows the association of powerful MMSE spectral[14] and log-spectral amplitude estimators[15] with the theory of microphone arrays. Therefore, the proposed technique combines the advantage of providing the spatially selective speech acquisition of a microphone array along with the nontonal residual noise (tonal noise consists of narrowband peaks across the spectrum, which is perceived as short tones varying in frequency from frame to frame producing a warbling sound that is very annoying to the human ear) of the MMSE Spectral and log-Spectral Amplitude estimators. Finally, implementation code and sample recordings can be found at http://slt.wcl.ee.upatras.gr/potamitis/GLv2.zip.

II. DESCRIPTION OF THE ALGORITHM

Due to the nonstationary nature of speech signals the microphone outputs are segmented into fixed blocks. We assume that the signal is stationary in each block; in addition we assume that after performing FFT on each block, each bin can be considered as a narrowband signal. The output from a linear microphone array of M sensors receiving a far-field wideband signal sampled at time instances $t=1,...,T$ is

$$\mathbf{x}(f_k,t)=[X_1(f_k,t),\ldots,X_M(f_k,t)]^T\in\mathbf{C}^{M\times 1},$$

where $X_m(f_k,t)$ $m=1,...,M$ is the DFT coefficient of the received signal at frequency f_k received at the mth microphone. Let D be half the size of the FFT, then $f_k=f_s(k-1)/2D$ $k=1,...,D$, where f_k is the frequency corresponding to the kth harmonic line of the spectrum and f_s is the sampling frequency. The array observation vector is assumed to satisfy the following model:

$$\mathbf{x}(f_k,t)=\mathbf{a}(f_k,\theta_0)s(f_k,t)+\mathbf{n}(f_k,t), \tag{1}$$

where $\mathbf{a}(f_k,\theta_0)$ is the steering vector of the source in the direction of θ_0 at frequency f_k. For a linear microphone array this vector is given by

$$\mathbf{a}(f_k,\theta)=[1,e^{-j\kappa d\cos(\theta)},\ldots,e^{-j\kappa d(M-1)\cos(\theta)}]^T. \tag{2}$$

Let $N_m(f_k,t)$ $m=1,...,M$ be the DFT coefficients of the noise process at frequency f_k received at the mth microphone. As such, $\mathbf{n}(f_k,t)$ is set as

$$\mathbf{n}(f_k,t)=[N_1(f_k,t),\ldots,N_M(f_k,t)]^T\in\mathbf{C}^{M\times 1}. \tag{3}$$

where $\kappa=2\pi f/c$, d is the spacing between adjacent microphones, and c is the sound velocity ($c\approx 343$ m/s). Let $\mathbf{S}(t)\in\mathbf{C}^{D\times D}$ and $\mathbf{N}(t)$, $\mathbf{A}(\theta_0)\in\mathbf{C}^{M\times D}$ be the matrices holding the DFT of the source, noise, and steering vectors, respectively, in all frequencies, then

$$\mathbf{X}(t)=[\mathbf{x}(f_1,t),\ldots,\mathbf{x}(f_D,t)],$$

$$\mathbf{S}(t)=\mathrm{diag}[s(f_1,t),\ldots,s(f_D,t)],$$

$$\mathbf{N}(t)=[\mathbf{n}(f_1,t),\ldots,\mathbf{n}(f_D,t)],$$

$$\mathbf{A}(\theta)=[\mathbf{a}(f_1,\theta),\ldots,\mathbf{a}(f_D,\theta)].$$

Under the null hypothesis H_0 speech is absent, i.e., $\mathbf{X}(t)=\mathbf{N}(t)$, $t=1,...,T$, while under the alternative, H_1, speech is present. That is,

$$\begin{aligned} &H_0:\mathbf{X}(t)=\mathbf{N}(t),\\ &H_1:\mathbf{X}(t)=\mathbf{A}(\theta_0)\mathbf{S}(t)+\mathbf{N}(t). \end{aligned} \tag{4}$$

The W-GLRT detector is based on the following assumptions:

(a) The spectral components are jointly independent.
(b) $\mathbf{x}(f_k,t)$ and $\mathbf{n}(f_k,t)$ are zero mean, temporally white, uncorrelated complex Gaussian random vector processes.
(c) θ_0 is unknown.

By using the above-noted assumptions, the joint probability density function of the received vectors $\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T)$ under hypothesis H_i $(i=0,1)$ are given by

$$p_0(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T)|\boldsymbol{\psi}_0)=\prod_{k=1}^{D}p_{0,f}(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T)|\boldsymbol{\psi}_{0,f_k}),$$

$$p_1(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T)|\boldsymbol{\psi}_1)=\prod_{k=1}^{D}p_{1,f}(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T)|\boldsymbol{\psi}_{1,f_k},\theta_0),$$

where $\boldsymbol{\psi}_{i,fk}\in\boldsymbol{\Psi}_{i,fk}$ stand for the unknown parameters under hypothesis $i\in\boldsymbol{\Psi}_{i,fk}$ denotes the parameter space of $\boldsymbol{\psi}_{i,fk}$ all defined for each f_k; $\psi_0=[\psi_{0,f_1},\ldots,\psi_{0,f_D}]$; $\psi_1=[\theta_0,\psi_{1,f_1},\ldots,\psi_{1,f_D}]$; $\psi_{0,\phi}=[\sigma^2_{n0,f_k}]$; and $\psi_{1,\phi}=[\sigma^2_{n1,f_k},\sigma^2_{s,f_k}]$. The W-GLRT based on T array snapshots in the frequency domain is used for deciding whether to accept H_0 or H_1 and is given by

$$L(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T))=\log\frac{\max_{\psi_0\in\Psi_0}p_0(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T)|\psi_0)}{\max_{\psi_1\in\Psi_1}p_1(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T)|\psi_1)}$$

$$=\log\frac{p_0(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T)|\hat{\psi}_0)}{p_1(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T)|\hat{\psi}_1)}, \tag{5}$$

where $\hat{\psi}_i$ stands for the maximum likelihood estimate of the unknown parameters under hypothesis *i*. Under the null hypothesis, the unknown parameter vector ψ_0 contains the unknown noise level in the various spectral bins f_k, whereas under the alternative hypothesis, the unknown parameter vector contains the unknown direction, the source powers, and the noise levels in the various frequencies. In the W-GLRT detector we compare $L(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T))$ to a threshold, denoted by γ (the detector's threshold). If $L(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T))>\gamma$ then H_0 is accepted, otherwise H_1 is accepted. In operational conditions, the detector's threshold is derived from the initial frames denoted by *Tini*, which are assumed to have only noise and is set to $\gamma=\min L(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_{Tini}))$.

One should note that due to possible movement of the speaker the number of snapshots on which W-GLRT is based is restricted. A small number of snapshots lead to more spurious DOAs due to sparse data while a large number of snapshots restricts the resolution of the likelihood to discern speech from noise (because of the nonstationary nature of the speech signal). For 8 kHz sampling, the number of overlapping snapshots should be between 1 and 5 (i.e., 64–224 ms). Every complex spectral coefficient in each snapshot forms a circular, zero mean, white, complex Gaussian vector with covariance matrix $\mathbf{R}_{\mathbf{x},f_k}=\sigma^2_{n0,f_k}\mathbf{I}$ under the null hypothesis, and $\mathbf{R}_{\mathbf{x},f_k}=\sigma^2_{s,f_k}\mathbf{a}(f_k,\theta_0)\mathbf{a}^H(f_k,\theta_0)+\sigma^2_{n1,f_k}\mathbf{I}$ under the alternative. The detection problem is now set as

$$H_0:\mathbf{X}(\mathrm{t})\sim\prod_{k=1}^{D}G(\mathbf{0},\sigma^2_{n0,f_k}\mathbf{I}),$$
$$H_1:\mathbf{X}(\mathrm{t})\sim\prod_{k=1}^{D}G(\mathbf{0},\sigma^2_{s,f_k}\mathbf{a}(f_k,\theta_0)\mathbf{a}^H(f_k,\theta_0)+\sigma^2_{n1,f_k}\mathbf{I}), \tag{6}$$

where G in Eq. (6) denotes the Gaussian density. Under the Gaussian assumption and independence of spectral components, the W-GLRT as set in Eq. (5) takes the form of

$$\begin{aligned}L(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T))&=-\max_{\psi_0\in\Psi_0}\sum_{k=1}^{D}T\{\log|\mathbf{R}_{\mathbf{X},f_k}(\psi_0)|+\mathrm{Tr}\{\mathbf{R}_{\mathbf{X},f_k}(\psi_0)^{-1}\hat{\mathbf{R}}(f_k)\}\}\\&\quad+\max_{\psi_1\in\Psi_1}T\sum_{k=1}^{D}\{\log|\mathbf{R}_{\mathbf{X},f_k}(\psi_1)|+\mathrm{Tr}\{\mathbf{R}_{\mathbf{X},f_k}(\psi_1)^{-1}\hat{\mathbf{R}}(f_k)\}\}\\&=-T\sum_{k=1}^{D}\{\log|\mathbf{R}_{\mathbf{X},f_k}(\hat{\psi}_0)|+\mathrm{Tr}\{\mathbf{R}_{\mathbf{X},f_k}(\hat{\psi}_0)^{-1}\hat{\mathbf{R}}(f_k)\}\}\\&\quad+T\sum_{k=1}^{D}\{\log|\mathbf{R}_{\mathbf{X},f_k}(\hat{\psi}_1)|+\mathrm{Tr}\{\mathbf{R}_{\mathbf{X},f_k}(\hat{\psi}_1)^{-1}\hat{\mathbf{R}}(f_k)\}\},\end{aligned} \tag{7}$$

where

$$\hat{\mathbf{R}}(f_k)=\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(f_k,\mathrm{t})\mathbf{x}^H(f_k,\mathrm{t})$$

is the empirical correlation matrix at each frequency and Tr( ) stands for the trace of a matrix.

Substituting the exact distributions of Eq. (6) in Eq. (7) results in

$$\begin{aligned}L(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T))&=-T\left(M\sum_{k=1}^{D}\log(\hat{\sigma}^2_{n0,f_k})+\sum_{k=1}^{D}\frac{\mathrm{Tr}(\hat{\mathbf{R}}(f_k))}{\hat{\sigma}^2_{n0,f_k}}\right)+\max_{\theta_0}T\left\{\sum_{k=1}^{D}\log|\hat{\sigma}^2_{s,f_k}\mathbf{a}(\theta_0,f_k)\mathbf{a}^H(\theta_0,f_k)+\hat{\sigma}^2_{n1,f_k}\mathbf{I}|\right.\\&\quad\left.+\sum_{k=1}^{D}\mathrm{Tr}\{(\hat{\sigma}^2_{s,f_k}\mathbf{a}(\theta_0,f_k)\mathbf{a}^H(\theta_0,f_k)+\hat{\sigma}^2_{n1,f_k}\mathbf{I})^{-1}\hat{\mathbf{R}}(f_k)\}\right\}.\end{aligned} \tag{8}$$

In order to derive $L(\mathbf{X}(t_1),\ldots,\mathbf{X}(t_T))$, one needs to calculate the second part of Eq. (8) which is related to θ_0 from where the maximum likelihood estimate (ML) of the unknown source bearing is estimated by finding the peak with respect to the angle of arrival. Under the hypothesis H_0 speech is absent while the noise variance is

$$\hat{\sigma}^2_{n0,f_k}=\mathrm{Tr}\frac{\hat{\mathbf{R}}(f_k)}{M}.$$

The variance of the spectral coefficient of speech under the hypothesis H_1 is σ^2_{s,f_k} in each frequency bin found by projecting the empirical correlation matrix on the subspace

spanned by the steering vector in the direction of $\hat{\theta}_0$ while the noise estimate is based on its orthogonal counterpart. Since the noise exists in the steering vector subspace as well as in speech, one should subtract the variance of the amount of noise projected on this subspace from the spectral variance of speech (for the narrowband case see, e.g., Ref. 16):

$$\hat{\sigma}^2_{n1,f_k} = \frac{1}{M-1}\mathrm{Tr}\left\{ \mathbf{I} - \frac{\mathbf{a}(f_k,\hat{\theta}_0)\mathbf{a}^H(f_k,\hat{\theta}_0)}{\mathbf{a}^H(f_k,\hat{\theta}_0)\mathbf{a}(f_k,\hat{\theta}_0)}\hat{\mathbf{R}}(f_k)\right\}, \quad (9)$$

$$\hat{\sigma}^2_{s,f_k} = \left\{ \frac{\mathbf{a}^H(f_k,\hat{\theta}_0)\hat{\mathbf{R}}(f_k)\mathbf{a}(f_k,\hat{\theta}_0)}{(\mathbf{a}^H(f_k,\hat{\theta}_0)\mathbf{a}(f_k,\hat{\theta}_0))^2} - \frac{\hat{\sigma}^2_{n1,f_k}}{\mathbf{a}^H(f_k,\hat{\theta}_0)\mathbf{a}(f_k,\hat{\theta}_0)}\right\}. \quad (10)$$

The enhanced speech signal in each spectral bin can be directly reconstructed from Eq. (10). After the subspace subtraction has been applied to all frequencies, when steered in direction $\hat{\theta}_0$ the beamformer returns a filtered output for each $\hat{\theta}_0$. The beamformer's output is given by

$$\hat{\mathbf{x}}(f_k,\mathrm{t}) = \mathbf{w}^H(f_k,\hat{\theta}_0)\mathbf{x}(f_k,\mathrm{t}). \quad (11)$$

The proposed technique is independent of the beamforming method applied to reconstruct the speech signal. However, in this work we applied minimum variance beamforming[6] and, therefore,

$$\mathbf{w}(f_k,\hat{\theta}_0) = \frac{\hat{\mathbf{R}}^{-1}(f_k)\mathbf{a}(f_k,\hat{\theta}_0)}{\mathbf{a}^H(f_k,\hat{\theta}_0)\hat{\mathbf{R}}^{-1}(f_k)\mathbf{a}(f_k,\hat{\theta}_0)}.$$

Based on the assumption that the human ear is relatively insensitive to phase distortion,[17] we focus on the short-time amplitude of the speech signal, leaving the noisy phase of the beamformer's output unprocessed. The noisy phase of the beamformer's output is added to the enhanced output. That is,

$$s(f_k,t) = \hat{\sigma}_{s,f_k}(\hat{\theta}_0)\exp(j\measuredangle\hat{\mathbf{x}}(f_k,t)) \quad (12)$$

and subsequently the underlying frame is reconstructed by using an inverse FFT and the weighted overlap-and-add method.

III. CALCULATION OF SPEECH ABSENCE/PRESENCE PROBABILITY

We show that conditioned on the current input to the array we can derive the speech absence probability, which is based on the likelihood of Eq. (5), calculated for each snapshot (i.e., $T=1$). By direct application of the Bayes rule and assuming independence between frequency bins we derive:

$$p(H_0|\mathbf{X}(t)) = \frac{p(H_0)\Pi_{k=1}^{D} p_{0,f_k}(\mathbf{x}(f_k,t)|H_0)}{p(H_0)\Pi_{k=1}^{D} p_{0,f_k}(\mathbf{x}(f_k,t)|H_0) + p(H_1)\Pi_{k=1}^{D} p_{1,f_k}(\mathbf{x}(f_k,t)|H_1)}$$

and finally:

$$p(H_0|\mathbf{X}(t)) = \frac{1}{1+q\left(\frac{\Pi_{k=1}^{D} p_{0,f_k}(\mathbf{x}(f_k,t)|H_0)}{\Pi_{k=1}^{D} p_{1,f_k}(\mathbf{x}(f_k,t)|H_1)}\right)^{-1}}, \quad (13)$$

where $q=p(H_1)/p(H_0)$ is the *a priori* probability ratio of the two hypotheses and is set the same value for every f_k. The value of q is set experimentally to provide a smooth speech presence/absence probability as in Refs. 10, 14 for the one channel case while a recursive calculation independently for each frequency is applicable as in Refs. 8 and 9.

Combining Eq. (5) and Eq. (13) yields:

$$p(H_0|\mathbf{X}(t)) = \frac{1}{1+q\exp(-L(\mathbf{X}(t)))}. \quad (14)$$

The assumption of spectral independence is also used in Ref. 10 to form a global soft decision about speech presence. However, in our case the log-likelihood is calculated from signal subspace techniques for the multichannel case. Equation (14) allows the calculation of a speech presence probability (SPP); $p(H_1|\mathbf{X}(t))=1-p(H_0|\mathbf{X}(t))$, the aim of which is threefold:

(a) It serves as a voice activity detector (VAD).
(b) It allows the estimated noise power to be updated regardless whether speech is present or not (see Refs. 9 and 10).
(c) It can be used as an additional gain function for speech enhancement [see Eq. (17) in Sec. IV].

IV. MMSE SPECTRAL AMPLITUDE AND LOG-SPECTRAL AMPLITUDE ESTIMATION FOR MICROPHONE ARRAYS

The W-GLRT is based on the assumption that the DFT coefficients of speech and noise are independently distributed as complex Gaussian pdfs with a time-varying unknown variance as summarized in the following:

$$p(\hat{\mathbf{x}}(f_k,t)|\mathrm{H}_0) = \frac{1}{\pi\sigma^2_{n,f_k}(\hat{\theta}_0)}\exp\left(\frac{-|\hat{\mathbf{x}}(f_k,t)|^2}{\sigma^2_{n,f_k}(\hat{\theta}_0)}\right), \quad (15)$$

$$p(\hat{\mathbf{x}}(f_k,t)|\mathrm{H}_1)=\frac{1}{\pi(\sigma^2_{s,f_k}(\hat{\theta}_0)+\sigma^2_{n,f_k}(\hat{\theta}_0))}\times\exp\left(\frac{-|\hat{\mathbf{x}}(f_k,t)|^2}{\sigma^2_{s,f_k}(\hat{\theta}_0)+\sigma^2_{n,f_k}(\hat{\theta}_0)}\right). \quad (16)$$

The suppression gain functions proposed by Ephraim and Malah (1984, 1985) for the problem of one channel speech enhancement are based on the same distributional assumptions. The latter gain functions are constructed to minimize the mean-square error estimates of spectrum[14] and of log-spectrum[15] (the latter known to be better associated with speech perception). The reader is referred to Refs. 14 and 15 for a detailed derivation. Here we briefly present both enhancement rules, primarily with a view to associate them with microphone arrays through the concept of *directive a priori* SNR which is DOA dependent. As first stated in Ref. 18 and later in Refs. 8–10 the noise suppression rule can be modified based on the knowledge of global speech presence,

$$\begin{aligned}s(f_k,t)&=E\{s(f_k,t)|\hat{\mathbf{x}}(f_k,t)\}\\&=E\{s(f_k,t)|\hat{\mathbf{x}}(f_k,t),H_0\}p(H_0|\mathbf{X}(\mathrm{t}))\\&\quad+E\{s(f_k,\mathrm{t})|\hat{\mathbf{x}}(f_k,t),H_1\}p(H_1|\mathbf{X}(\mathrm{t}))\\&=E\{\mathrm{s}(f_k,t)|\hat{x}(f_k,t),H_1\}p(H_1|\mathbf{X}(\mathrm{t})). \quad (17)\end{aligned}$$

The rules are summarized in their corresponding gain functions $G^{\mathrm{SA}}_{f_k}$ Eq. (18) and $\mathrm{G}^{\mathrm{LSA}}_{f_k}$ Eq. (19) applied to $\hat{\mathbf{x}}(f_k,t)$,

$$E\{s(f_k,t)|\hat{\mathbf{x}}(f_k,t),H_1\}=G_{f_k}(\xi_{f_k}(\hat{\theta}_0,t),\gamma_{f_k}(\hat{\theta}_0,t))\hat{\mathbf{x}}(f_k,t),$$

$$G^{\mathrm{SA}}_{f_k}(\xi_{f_k}(\hat{\theta}_0,t),\gamma_{f_k}(\hat{\theta}_0,t))=\frac{\sqrt{\pi}}{2}\sqrt{\frac{\xi_{f_k}(\hat{\theta}_0,t)}{\gamma_{f_k}(\hat{\theta}_0,t)(\xi_{f_k}(\hat{\theta}_0,t)+1)}}\times M\left(\frac{\gamma_{f_k}(\hat{\theta}_0,t)\xi_{f_k}(\hat{\theta}_0,t)}{1+\xi_{f_k}(\hat{\theta}_0,t)}\right), \quad (18)$$

$$M(z)=\exp\left(-\frac{z}{2}\right)\left\{(1+z)I_0\left(\frac{z}{2}\right)+zI_1\left(\frac{z}{2}\right)\right\},$$

where I_0 and I_1 are the modified Bessel functions of zero and first order, and

$$G^{\mathrm{LSA}}_{f_k}(\xi_{f_k}(\hat{\theta}_0,t),\gamma_{f_k}(\hat{\theta}_0,t))=\frac{\xi_{f_k}(\hat{\theta}_0,t)}{\xi_{f_k}(\hat{\theta}_0,t)+1}\exp\left(\frac{1}{2}\int_{u_{f_k}(\hat{\theta}_0,t)}^{\infty}\frac{e^{-t}}{t}dt\right), \quad (19)$$

where $\xi_{fk}(\hat{\theta}_0,t)$ is the *a priori* SNR ratio. In our multichannel framework the *a priori* SNR ratio can be defined for every possible angle. Equation (19) requires the calculation of the *a priori* SNR in the direction of $\hat{\theta}_0$, which is time varying due to movement of the speaker. In that sense the *a priori* SNR is DOA dependent (hereafter *directive a priori* SNR ratio) and is calculated by substituting Eq. (20) in Eqs. (18) and (19). One should note that $\xi_{fk}(\hat{\theta}_0,t)$ is derived from signal subspace projections [i.e., directly from Eqs. (9) and (10)] already derived for the calculation of W-GLRT and not by using the derivation of the one-channel method (see Refs. 8–10). Following a similar reasoning $\gamma_{fk}(\hat{\theta}_0,t)$ is the *directive a posteriori* SNR ratio,

$$\xi_{f_k}(\hat{\theta}_0,t)=\frac{\sigma^2_{s,f_k}(\hat{\theta}_0)}{\sigma^2_{n,f_k}(\hat{\theta}_0)},$$

$$\gamma_{f_k}(\hat{\theta}_0,t)=\frac{|\hat{\mathbf{x}}(f_k,t)|^2}{\sigma^2_{n,f_k}(\hat{\theta}_0)},$$

$$u_{f_k}(\hat{\theta}_0,t)=\frac{\gamma_{f_k}(\hat{\theta}_0,t)\xi_{f_k}(\hat{\theta}_0,t)}{\xi_{f_k}(\hat{\theta}_0,t)+1}. \quad (20)$$

Finally a time-smoothed version of the $\xi_{fk}(\hat{\theta}_0,t)$ is employed, conducive to the elimination of musical noise: [Tonal noise consists of narrowband peaks across the spectrum, which is preceived as short tones varying in frequency from frame to frame producing a warbling sound that is very annoying to the human ear.]

$$\hat{\xi}_{f_k}(\hat{\theta}_0,\mathrm{t})=a\frac{s^2(f_k,t)}{\sigma^2_{n,f_k}(\hat{\theta}_0)}+(1-a)\max(0,\gamma_{f_k}-1),\qquad a=0.98, \quad (21)$$

where α is a weighting term acting as a low pass filter.

V. SIMULATION EXPERIMENTS

The room acoustics simulation takes place in a typical virtual 6.8 m×4.55 m×3 m room and is carried out by using specialized software for spatial sound rendering.[20] The full impulse response is calculated by summing line-of-sight and additional ray-traced reflection contributions from the moving virtual sound source, allowing for the sound emission profile and the wall's reflection profiles as direction-dependent attenuation; the impulse response is computed frame-wise and averaging takes place to smooth boundaries of frames. The initial and final times of each speech frame are rounded to the nearest integer sample time and all the samples in between are interpolated using third-order polynomial interpolation. The four walls possess a 0.7 reflection coefficient while the ceiling and the floor possess 0.9 and 0.4, respectively. We made use of an array of $M=8$ omnidirectional microphones with $d=0.1$ m spacing between microphones with its center located at (3.4,0.25,1.6) m. During his movement the speaker is continuously active uttering randomly selected recordings from the TIMIT database.[21] The speaker starts from (6.4,0.25,1.6) m at $t=0$ as shown in Fig. 1 and performs a circular movement with angular velocity 0.1257 rad/s. He walks for 6.25 s and stops at 135°. At $t=8.25$ s he moves on and at 14.5 s stops moving at 90° (at broadside position). At $t=17.5$ s he continues the circular motion until $t=30$ s. The DOAs were calculated by using

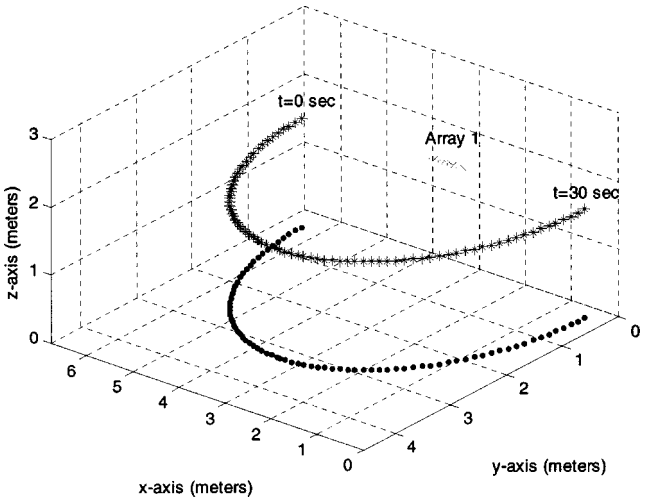


FIG. 1. Trajectory of a moving speaker inside a reverberant room. The reflection coefficient of the walls is 0.7 while the ceiling and the floor have 0.9 and 0.4, respectively.

hamming-windowed frame (512 samples) with 50% overlap at 8 kHz sampling, by using a 512 points FFT.

VI. SPEECH ACTIVITY DETECTION EVALUATION

Since a speaker can move at random inside a room and DOA estimation is dependent, among other factors, on the reverberation and the distance of the speaker from the microphones, a quantitative analysis of speech activity detection is very difficult. In order to have an indication of the performance we performed ten simulations with the same experimental settings as described in the last paragraph using the recordings of five male and five female speakers of the TIMIT database (30 s of concatenated utterances). An example of the detection results using the W-GLRT, the signal presence probability (SPP), and the final VAD decision is depicted in Fig. 2(a) and its detail in Fig. 2(b). The recordings of each set were corrupted in turn by additive white Gaussian noise (AWGN), and babble noise ranging from -10 to 20 dB taken from the NOISEX-92 database.[22] Babble noise was composed of a real recording of many people discussing—a situation commonly encountered in large cafeterias. Speech corpora (a total number of almost 10 000 frames) were manually tagged into ''inactive'' and ''active speech'' categories. The ''active speech'' category was further partitioned into onset, offset, and strongly present speech in order to assess the capability of the detection algorithm for different parts of typical speech segments. The onset speech refers to the low energy part of speech after a stop and before the vocal cord vibration or consonant reaches a steady state (strongly present speech) begins. Offset speech is also the low energy part starting from the end of the steady state of the strongly present speech and fading to silence.

When a speaker pauses, there is no speech signal available to carry out a correct estimate of the DOA. Detection of speech presence can reduce the misclassification of a speech/speaker recognition engine by rejecting noise segments and saves channel bandwidth in the case where the microphone array is involved in a wireless application (e.g., distributed speech recognition).

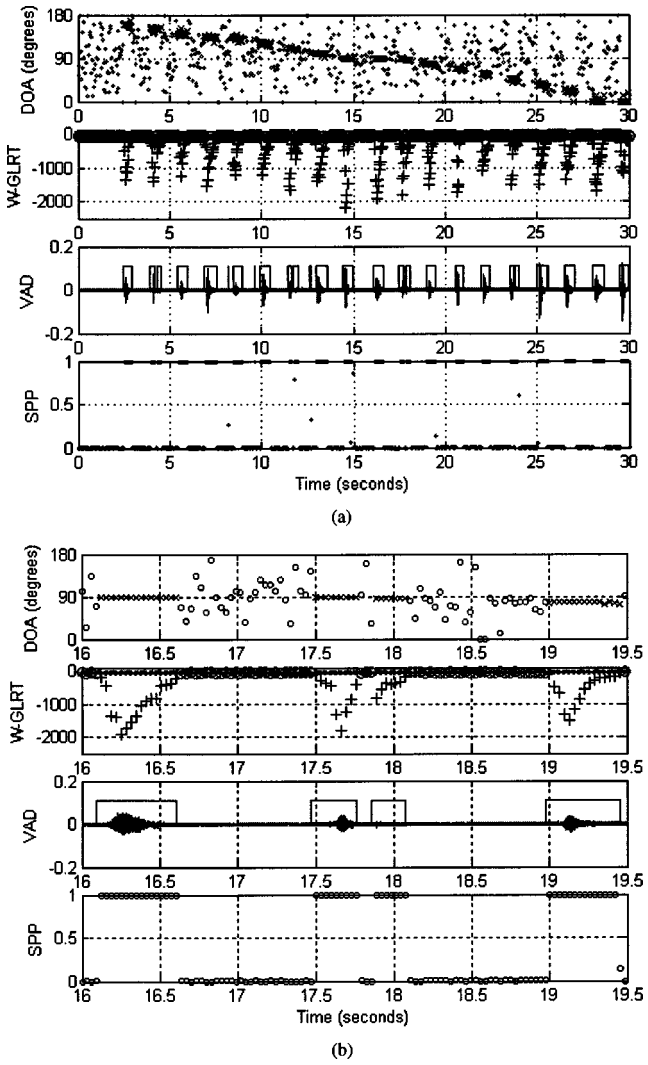


FIG. 2. (a) DOA measurements of speech corrupted by AWGN at 10 dB. Circles are measurements attributed to noise while crosses indicate speech-originating DOAs. (b) Log-likelihood scores (circles are rejected measurements). (c) Noisy speech signal and VAD decision. (d) Signal presence probability (SPP). (e) Detail of (a) (from 16 to 19.5 s).

The performance of the detector can by analyzed by calculating the receiver operating characteristic (ROC) curve which is the parametric plot of probability of detection (P_D) versus the probability of false alarm (P_F), based on the two hypotheses [see also Eq. (16)]:

$$H_0\colon \textit{speaker not active (speech pauses)},$$

$$H_1\colon \textit{speaker active}.$$

$P_D = Pr\{H_1/H_1\}$ and $P_F = Pr\{H_1/H_0\}$. The proposed method is based on a probabilistic model which allows directly inferring speech presence/absence either by Eq. (14) or by Eq. (5). One can observe that Eq. (14) is a scaled version of W-GLRT to produce valid probability scores and, therefore, must demonstrate the same performance. The validation over all SNRs is based on the manually tagged speech corpora comprising segments labeled H_0 and H_1.

The parts of the speech signal labeled as strong speech presence demonstrated gradual degradation with decreasing SNR and are adequately detected even at extremely low

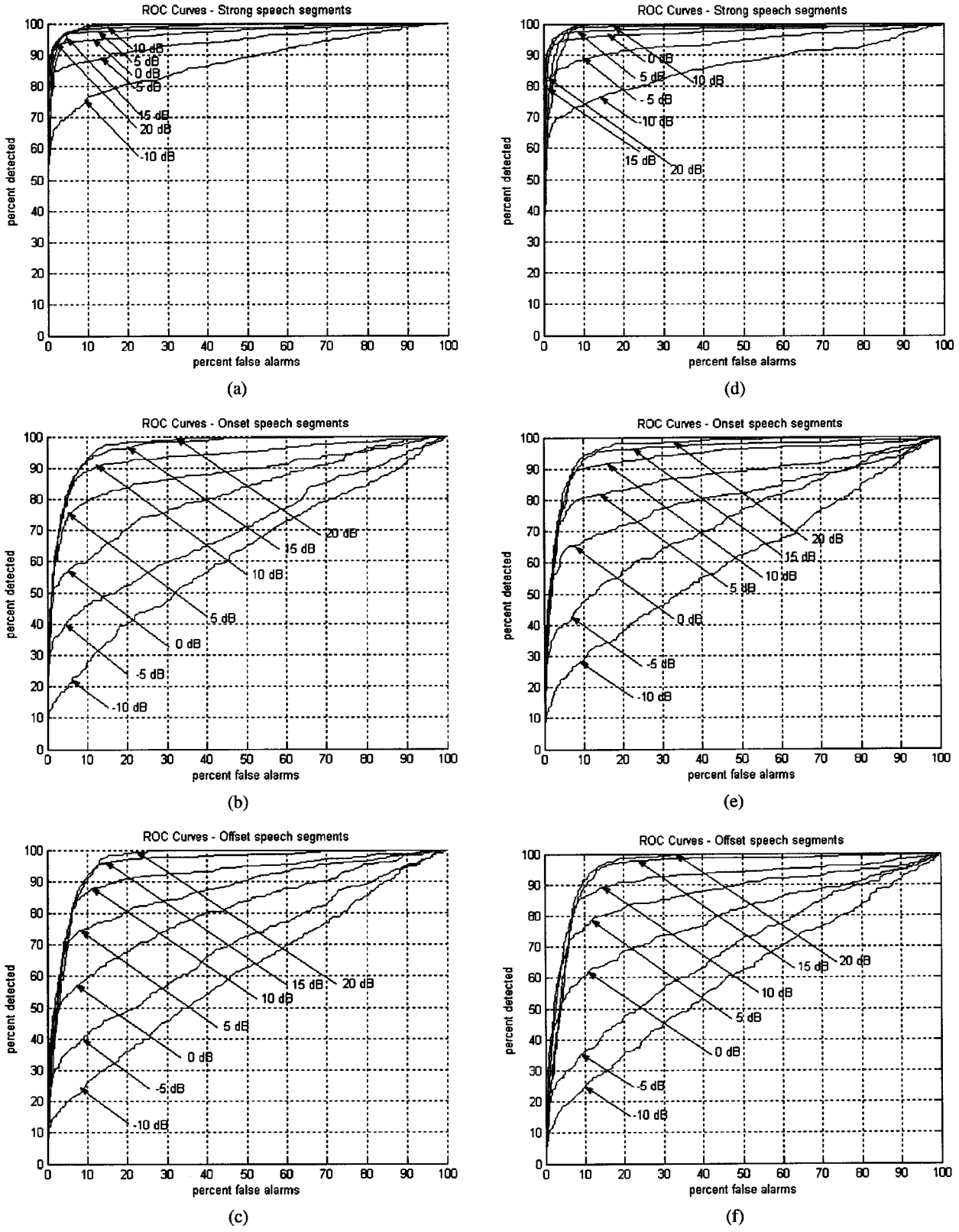


FIG. 3. Detection accuracy validation. ROC curves for SNRs ranging from −10 to 20 dB and using one snapshot. (a)–(c) White Gaussian corruption. (d)–(f) Babble noise corruption. (a),(d) Strong speech presence. (b),(e) Onset speech. (c),(f) Offset speech.

SNRs. For SNRs from 0 to 20 dB even for values of P_F close to zero the P_D is very high for AWGN and babble noise corruption as depicted in Figs. 3(a) and (d).

The speech segments that are mostly vulnerable to noise corruption are the onset and offset parts that possess low energy [see Figs. 3(b)–(c) and Figs. 3(e)–(f)]. For these parts, P_D rises much faster than P_F for SNRs as low as 5 dB SNR. For lower SNRs, the slow improvement of P_D with

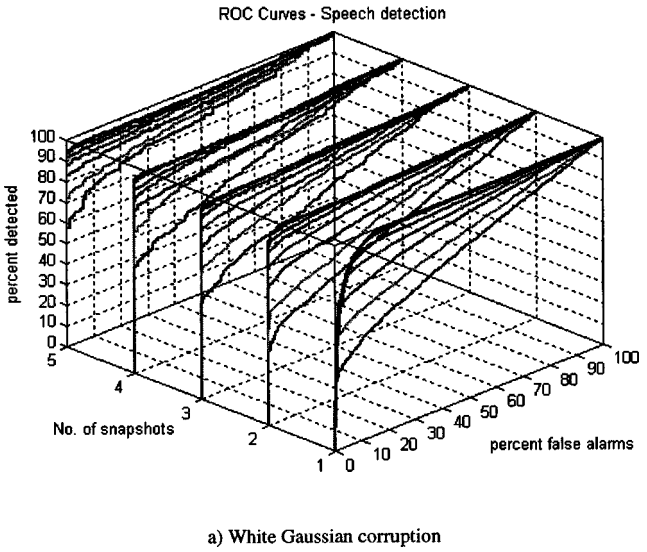


FIG. 4. Detection accuracy validation. ROC curves for SNRs ranging from −10 dB (top down curve) to 20 dB (upper curve) with a step of 5 dB and by employing a varying number of snapshots ranging from $T=1,...,5$. White Gaussian noise.

increasing P_F indicates that the predominance of noise can lead to the misclassification of low energy speech for low P_F.

For most applications operating in an environment with SNR as low as 5 dB a small value for P_F is tolerable. For noisier environments (down to −5 dB) the parts of speech that suffer most are the beginning and end of isolated speech segments that can be clipped.

At this point one should note that the VAD decision was carried out on a per frame basis (every 64 ms at 8 kHz sampling). In order to assess the extent the speech detection could be assisted if one increased the number of snapshots the VAD decision is based on we performed detection experiments on the same data by varying the number of snapshots from $T=1,...,5$. In order for a DOA and its corresponding W-GLRT score to be valid, the direction of sound propagation must remain constant through the averaging process of the spatial correlation matrix. Therefore, the number of snapshots employed by the algorithm cannot be increased arbitrarily since this would implicitly imply that the speaker is static whereas we want to retain the generality of the proposed technique to deal with possibly moving speakers. Figure 4 illustrates the relation between P_D and P_F as the SNR and number of snapshots employed in the derivation of the VAD decision varies. For a block of five snapshots, which is a reasonable number for speakers performing normal movement, the P_D rises abruptly over 80% for SNRs down to 0 dB for P_F close to zero. The P_D degrades gradually with a decreasing number of available snapshots and decreasing SNR. One should also note that the results derived in Fig. 4 assume two states, that is, active speech and nonactivity and, therefore the onset, offset, and strong speech presence states are merged into the category of speech activity. The depicted ROC results are derived under the assumption of AWGN corruption. The ROC curves for the babble noise case are quite similar [compare also Figs. 2(a)–(c) to Figs. 2(d)–(f).

TABLE I. MOS results of the proposed array-VAD combined with noise suppression for Gaussian noise corruption (circularly moving speakers). *Subspace subtraction* refers to Eq. (10), MMSE_SA refers to Eq. (18), MMSE_LSA to Eq. (19).

SNR (dB)	Subspace subtraction	MMSE SA	MMSE LSA	Unprocessed signal
−10	2.25	1.65	**2.32**	1.15
−5	2.72	2.90	**2.95**	1.81
0	3.05	3.22	**3.55**	2.25
5	3.75	3.93	**3.95**	2.82
10	4.21	4.11	**4.29**	3.25
15	**4.62**	4.13	4.34	3.83
20	**4.82**	4.32	4.45	4.45

TABLE II. MOS Scale for assessing the quality of the perceived speech and the degradation due to the enhancement procedure.

Points	Quality	Degradation assessment
5	Excellent	Imperceptible
4	Good	Just Perceptible, not Annoying
3	Fair	Perceptible, slightly Annoying
2	Poor	Annoying, not Objectionable
1	Bad	Annoying and Objectionable

VII. SPEECH ENHANCEMENT EVALUATION

We performed subjective mean opinion score (MOS) tests[17] in order to assess the quality of perception of the enhanced signals. Comparative evaluation of the proposed algorithm was conducted and the quality and impairment of the enhancement results were scored by ten listeners. The results were averaged over ten utterances from five randomly selected male and five female speakers of the TIMIT database. All speakers followed the scenario of Sec. V. Tables I–III show the MOS results provided by the enhancement techniques when the array VAD is applied (noise only frames are nulled and noisy speech is enhanced). As regards the quality of perception, in general, the noise reduction of the proposed array-VAD combined with the enhancement rule of Eq. (20) affected smaller percentage of nontonal noise than the subspace technique (although it inflicted an echo-like artifact) and at low SNRs it consistently outperformed the latter. However, at higher SNRs the subspace subtraction had a small advantage.

TABLE III. MOS results of the proposed array-VAD combined with noise suppression for babble noise corruption (circularly moving speakers). The notation is the same as in Table I.

SNR (dB)	Subspace subtraction	MMSE SA	MMSE LSA	Unprocessed signal
−10	1.05	1.1	**1.2**	1
−5	1.55	1.3	**1.95**	1.6
0	**2.85**	2.05	2.35	1.85
5	**4.05**	3.15	3.45	2.55
10	**4.9**	3.65	3.85	3.25
15	**5**	4.4	4.55	4.1
20	**5**	4.7	4.85	4.65

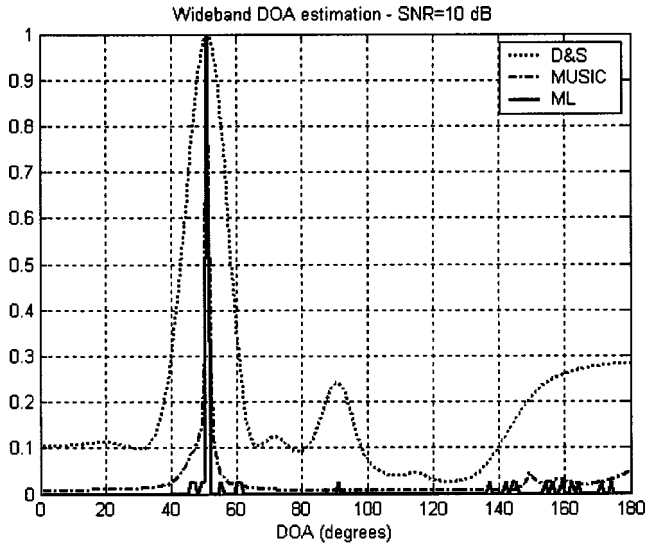


FIG. 5. Normalized DOA measurements of a voiced speech segment. Speech signal corrupted by AWGN at 10 dB. Comparative results for delay and sum (D&S), MUSIC and maximum likelihood (ML). True DOA is at 50°. All methods use five snapshots in the frequency domain.

VIII. ML WIDEBAND DOA ESTIMATION

In general, the wideband DOA-estimation results of MUSIC, ESPRIT, minimum variance, and ML are very accurate when dealing with a small number of sound sources and in the case of our application with a single speaker. The applicability of all aforementioned techniques to general multispeaker scenarios is greatly complicated as none can directly associate DOAs with particular sources raising data association problems which are beyond the scope of this article (see also Ref. 6, p. 424). In our comparative tests, the wideband version of the DOA estimators is derived following the incoherent approach; therefore the wideband array output of each microphone is decomposed into narrowband components by using discrete Fourier transform (DFT). Subsequently, the frequency bands of the DFT transformed block corresponding to the narrowband components with the largest power values are selected and each narrowband DOA estimation technique is applied independently to each harmonic line. The DOAs are derived from the histogram and clustering of the angle of arrivals of all high power spectral bins.

For the case of a single speaker, the performance of the ML—DOA estimation as described in Sec. II compares equally to MUSIC and minimum variance. Although the ML technique seems to return estimates that are more concentrated around the true angle values as illustrated in Fig. 5, the peaks corresponding to different methods are rarely erroneous for the single active-speaker case. However, for multiple *narrowband* sources the superiority of the ML solution over MUSIC has been established.[23] In the framework of the semicircular movement of the speaker as shown in Fig. 1, the root mean square error averaged over ten simulations of all methods is about 2° for the case of 20 SNR dB (AWGN corruption) using five snapshots to about 5° for the case of −10 dB and one snapshot excluding outlier values. The distinct advantage provided by the proposed approach is that once the W-GLRT and the parameters related to it are calculated one can simultaneously attack the DOA estimation, voice activity detection, and speech enhancement problem whereas a unifying statistical framework for these functions is lacking for the other techniques.

IX. CONCLUSIONS

We demonstrated that W-GLRT provides a means to introduce the spatial selectivity of microphone arrays to voice activity detection. W-GLRT is derived on a per block basis and, therefore, can serve as an online VAD detector of moving speakers. Moreover the statistical formulation of W-GLRT serves as a unifying treatment of DOA estimation, statistical voice detection, and speech enhancement. The detection and enhancement capabilities of W-GLRT proved robust at very low SNRs for Gaussian noise corruption and the colored non-Gaussian case of babble noise corruption.

[1] Y. Huang, J. Benesty, G. Elko, and R. Mersereau, "Real-time passive source localization: An unbiased linear-correction least-squares approach," IEEE Trans. Speech Audio Process. **9**, 943–956 (2001).
[2] P. Aarabi and S. Zaky, "Robust sound localization using multi-source audiovisual information fusion," Information Fusion **2**, 209–223 (2001).
[3] D. Rabinkin, R. J. Renomeron, A. Dahl, J. French, J. Flanagan, and M. Bianchi, "A DSP implementation of source location using microphone arrays," Proceedings of the SPIE, Vol. **2846**, pp. 88–99, Denver, Colorado, August 1996.
[4] T. Yamada, S. Nakamura, and K. Shikano, "Distant-talking speech recognition based on a 3-D Viterbi search using a microphone array," IEEE Trans. Speech Audio Process. **10**, 48–56 (2002).
[5] M. Hoffman, Z. Li, and D. Khataniar, "GSC-based spatial voice activity detection for enhanced speech coding in the presence of competing speech," IEEE Trans. Speech Audio Process. **9**, 175–179 (2001).
[6] D. Johnson and D. Dudgeon, *Array Signal Processing: Concepts and Techniques* (Prentice Hall, New York, 1993).
[7] J. Friedmann, E. Fishler, and H. Messer, "General asymptotic analysis of the generalized likelihood test for a Gaussian point source under statistical or spatial mismodeling," IEEE Trans. Signal Process. **50**, 2617–2631 (2002).
[8] D. Malah, R. Cox, and A. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments," Proc. IEEE ICASSP **1**, 789–792 (1999).
[9] I. Soon, S. Koh, and C. Yeo, "Improved noise suppression filter using self-adaptive estimator of probability of speech absence," Signal Process. **75**, 151–159 (1999).
[10] N. Kim and J. Chang, "Spectral enhancement based on global soft decision," IEEE Signal Process. Lett. **7**, 108–110 (2000).
[11] Y. Hioka and N. Hamada, "Voice activity detection with array signal processing in the Wavelet domain," IEICE Trans. Fundamentals **E86-A**, 2802–2811 (2003).
[12] F. Beritelli, S. Casale, and A. Cavallaro, "A multi-channel speech/silence detector based on time delay estimation and fuzzy classification," Proc. IEEE ICASSP **1**, 93–96 (1999).
[13] R. Bouquin and G. Le Faucon, "Proposal of a voice activity detector for noise reduction," Electron. Lett. **30**, 930–932 (1994).
[14] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **32**, 1109–1121 (1984).
[15] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **33**, 443–445 (1985).
[16] P. Stoica and A. Nehorai, "On the concentrated stochastic likelihood function in array signal processing," Circuits Syst. Signal Process. **14**, 669–674 (1995).
[17] J. Deller, J. Proakis, and J. Hansen, *Discrete-time Processing of Speech Signals* (Prentice–Hall, Englewood Cliffs, NJ, 1993), p. 296.
[18] R. McAulay and M. Malpass, "Speech enhancement using a soft decision noise suppression filter," IEEE Trans. Speech Audio Process. **28**, 137–145 (1980).

[19] O. Cappé, ''Elimination of the musical noise phenomenon using a minimum mean-square error short-time spectral amplitude estimator,'' IEEE Trans. Speech Audio Process. **32**, 345–349 (1994).
[20] http://www.sonicspot.com/3daudio/3daudio.html
[21] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren, ''DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM, NIST Speech Disc 1-1.1'' (1993).
[22] A. Varga and H. Steeneken, ''Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems,'' Speech Commun. **12**, 247–251 (1993).
[23] A. Jaffer, ''Maximum likelihood direction finding of stochastic sources: A separable solution,'' Proc. IEEE of the IEEE International Conference on Acoustics, Speech, and Signal Processing., 1998, **vol. 5**, pp 2893–2896.

Perfect harmony: A mathematical analysis of four historical tunings

Michael F. Page[a)]
The Pingry School, Martinsville Road, Martinsville, New Jersey 08836

(Received 30 September 2003; revised 13 July 2004; accepted 14 July 2004)

In Western music, a musical interval defined by the frequency ratio of two notes is generally considered consonant when the ratio is composed of small integers. Perfect harmony or an ''ideal just scale,'' which has no exact solution, would require the division of an octave into 12 notes, each of which would be used to create six other consonant intervals. The purpose of this study is to analyze four well-known historical tunings to evaluate how well each one approximates perfect harmony. The analysis consists of a general evaluation in which all consonant intervals are given equal weighting and a specific evaluation for three preludes from Bach's ''Well-Tempered Clavier,'' for which intervals are weighted in proportion to the duration of their occurrence. The four tunings, 5-limit just intonation, quarter-comma meantone temperament, well temperament (Werckmeister III), and equal temperament, are evaluated by measures of centrality, dispersion, distance, and dissonance. When all keys and consonant intervals are equally weighted, equal temperament demonstrates the strongest performance across a variety of measures, although it is not always the best tuning. Given C as the starting note for each tuning, equal temperament and well temperament perform strongly for the three ''Well-Tempered Clavier'' preludes examined. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1788732]

PACS numbers: 43.75.Bc [SEM] Pages: 2416–2426

I. INTRODUCTION

In Western music, a musical interval defined by the frequency ratio of two notes is generally considered pleasing or consonant when the ratio is composed of small integers. For example, an octave has a ratio of 2:1, and two notes an octave apart are often perceived as being essentially the same note. In addition to unison (1:1 ratio) and the octave (2:1), six other musical intervals are generally considered consonant: perfect fifth, perfect fourth, major and minor thirds, and major and minor sixths. As discussed by Tenney,[1] the meaning of consonance has evolved over time, but despite different theories about the causes of consonance, there is strong agreement that intervals defined by small integer ratios are the most consonant. Perfect harmony would require the division of an octave into 12 tones, each of which would be used as the root note for the other consonant intervals that have the following frequency ratios: 3:2 (perfect fifth), 4:3 (perfect fourth), 5:4 (major third), 6:5 (minor third), 5:3 (major sixth), and 8:5 (minor sixth). As described by Hall,[2] perfect harmony or an ''ideal just scale'' has no exact solution, that is, there is no set of 12 tones that can exactly meet all the constraints imposed by the small-integer ratios of the consonant intervals.

The purpose of this study is to analyze four well-known historical tunings, each of which defines a 12-tone scale, to evaluate how well each one approximates perfect harmony. Because each tuning includes unison and an octave by definition, the evaluation focuses on the six other consonant intervals. The four tunings in this study are particular cases of just intonation (JI), quarter-comma meantone temperament (MT), well temperament (Werckmeister III) (WT), and equal temperament[3] (ET). These were selected because each has held a prominent place in the evolution of tunings in Western music. The analysis consists of two phases: a general evaluation in which all consonant intervals are given equal weighting and a specific evaluation for three preludes from Bach's ''Well-Tempered Clavier,''[4] for which intervals are weighted in proportion to the duration of their occurrence in the music. Recent work on tempered intervals has focused on the perception of consonance (see, for example, Vos and van Vianen[5]). The analysis reported here uses functions of physical variables, specifically frequencies and spectra, which are correlated with the perceptual variables of pitch and timbre, to compare tunings. Such quantitative measures can be used to complement perceptual studies about the consonance of a tuning. In several areas, comparisons to the results of perceptual studies have been included here. Previous work by Hall[6] examined the performance of a wide range of tunings with respect to specific pieces of music using a single measurement statistic, while Rasch[7] examined a variety of regular[8] tunings using three criteria, but without examination of any specific music. This study extends their work by applying a comprehensive set of evaluation criteria, including the ones used by Hall and Rasch, and by evaluating regular and irregular tunings in the context of both equal and composition-specific weightings. This analysis also extends the work by Barnes[9] on the ''Well-Tempered Clavier'' to the extent that all consonant intervals, not just the major thirds, are examined; however, this study is not intended to justify any specific temperament as the one Bach intended for the ''Well-Tempered Clavier.'' Finally, the analysis includes a quantitative measure of dissonance introduced by Sethares[10] which is based on the Plomp–Levelt curves[11] for perception

a)Electronic mail: michaelfpage@att.net

TABLE I. Definitions of the intervals in a 12-tone scale relative to the tonic for four historical tunings. Columns (a) and (b) provide the common names and symbols for the intervals. Columns (c)–(f) provide the frequency ratios that define these intervals for one version of 5-limit just intonation, quarter-comma meantone temperament, well temperament (Werckmeister III), and equal temperament, respectively.

Interval name (a)	Interval symbol (b)	5-Limit just intonation (JI) (c)	Quarter-comma meantone temperament (MT) (d)	Well temperament (Werckmeister III) (WT) (e)	Equal temperament (ET) (f)
Unison	p1	1/1	1/1	1/1	1/1
Minor second	m2	16/15	$2187/2048\times(80/81)^{7/4}$	256/243	$2^{1/12}/1$
Major second	M2	9/8	$9/8\times(80/81)^{2/4}$	$9/8\times 1/P^{2/4}$	$2^{2/12}/1$
Minor third	m3	6/5	$32/27\times(81/80)^{3/4}$	32/27	$2^{3/12}/1$
Major third	M3	5/4	$81/64\times(80/81)^{4/4}=5/4$	$81/64\times 1/P^{3/4}$	$2^{4/12}/1$
Perfect fourth	p4	4/3	$4/3\times(81/80)^{1/4}$	4/3	$2^{5/12}/1$
Tritone	tt	45/32	$729/512\times(80/81)^{6/4}$	1024/729	$2^{6/12}/1$
Perfect fifth	p5	3/2	$3/2\times(80/81)^{1/4}$	$3/2\times 1/P^{1/4}$	$2^{7/12}/1$
Minor sixth	m6	8/5	$6561/4096\times(80/81)^{8/4}=25/16$	128/81	$2^{8/12}/1$
Major sixth	M6	5/3	$27/16\times(80/81)^{3/4}$	$27/16\times 1/P^{3/4}$	$2^{9/12}/1$
Minor seventh	m7	9/5	$16/9\times(81/80)^{2/4}$	16/9	$2^{10/12}/1$
Major seventh	M7	15/8	$243/128\times(80/81)^{5/4}$	$243/128\times 1/P^{3/4}$	$2^{11/12}/1$
Octave	p8	2/1	2/1	2/1 where P= 531441/ 524288	2/1

of pure tones, but which he generalized to account for the harmonic partials of complex tones.

Equal temperament, which divides an octave into 12 equally spaced half-tones, has been the standard tuning since the 19th Century. However, during the 20th Century, it has been criticized by proponents of just intonation as a ''matter of expediency'' since it simplifies instrument design and playing techniques (see, for example, Doty[12]). This study demonstrates that equal temperament performs strongly in approximating perfect harmony when all keys and all consonant intervals are given equal weighting compared to the particular cases of the other tunings examined, as well as for the performance of the three ''Well-Tempered Clavier'' preludes examined. The findings here agree with the results of Rasch[13] in demonstrating that equal temperament is an appropriate tuning for the ''Well-Tempered Clavier.'' Additionally, well temperament also performs strongly, especially for the specific ''Well-Tempered Clavier'' preludes. The results for the specific preludes are subject to the selection of C as the starting note for each tuning,

II. FOUR HISTORICAL TUNINGS INCLUDED IN THIS STUDY

A. 5-Limit just intonation

As defined by Barbour,[14] just intonation refers to any 12-note scale that contains some pure fifths and major thirds with the ratios given earlier. Probably the earliest example of just intonation is the tuning credited to the Greek mathematician and philosopher Pythagoras (6th Century B.C.). Pythagorean tuning is based on multiples of the prime numbers two and three, but because it does not utilize the number five, the thirds and sixths do not sound consonant by most standards. This tuning was popular, nonetheless, up through the Middle Ages.

From the 14th to the 16th Centuries, several music theorists were involved in introducing the use of consonant thirds and, as such, were forced to abandon Pythagorean tuning. Many variations of just intonation with some pure fifths and major thirds have been defined, with one common definition in terms of frequency ratios shown in Table I, column (c).[15] This particular version of just intonation agrees with what Doty[12] referred to as the ''harmonic duodene of C.'' The modifier ''5-limit'' indicates that the number five is the largest prime number used as a base number in the frequency ratios.

Although just intonation was widely accepted by theorists during the Renaissance, the growth of fixed-pitch instruments, such as pipe organs and keyboard instruments, meant that just intonation had limited use in practice. Modulation to other keys produces some very unacceptable intervals. Although the frequency ratios in Table I, column (c), for the consonant intervals are based on small integers, modulation to other keys creates ratios as complex as 1024/675.

Efforts were made during the 16th and 17th Centuries to create keyboard instruments with significantly more than 12 notes per octave in an attempt to achieve the ratios in Table I, column (c), for at least several keys. But these instruments

did not become popular, and the concept of temperament, or altering the tuning of certain notes, took hold. For this study, an initial tone of C was set at 261.63 Hz and the ratios in Table I, column (c), were applied to obtain the remaining notes in a scale. These tones were modulated through all keys and analyzed to determine how close the 72 ratios formed by the six consonant intervals of interest are to the pure ratios for perfect harmony.[16]

B. Quarter-comma meantone temperament

Despite the interest of theorists in just intonation, meantone temperament, first introduced in the early 1400s, became the predominant method of tuning during the 16th and 17th Centuries. In meantone, the major thirds of Pythagorean tuning are made more consonant by a slight tempering of the perfect fifths. In addition, the tone located between the root note (say C) and the higher note in the major third (E) is positioned at the mean of the interval. This results in two tones (C–D and D–E) of equal size, whereas just intonation has tones of different sizes.

There are many ways to accomplish meantone, but the predominant method, quarter-comma meantone, was devised by Pietro Aaron in 1523. Each perfect fifth is reduced by one-quarter of a syntonic comma.[17] Then other tones are constructed by adding fifths together (multiplying their respective ratios), which after four fifths yields a major third. This process is continued until all but one of the fifths is tempered. The result is eight pure major thirds, but also one wide fifth, known as a ''wolf'' interval because of its sharp sound. The frequency ratios used to construct a 12-tone scale in quarter-comma meantone are shown in Table I, column (d). The complete set of frequency ratios for quarter-comma meantone temperament with each note used in turn as the root note was also calculated for comparison to the frequency ratios for perfect harmony.

C. Well temperament (Werckmeister III)

As composers started to make greater use of modulation in a single composition, the existence of wolf intervals in quarter-comma meantone made this tuning less attractive. This led to the development of well temperaments, including the popular Werckmeister III (c. 1691) named after its inventor, Andreas Werckmeister. In this temperament, the goal is to keep the fifths and thirds as close to just intonation as possible, while spreading out the wolf intervals. This is accomplished by spreading the Pythagorean comma,[18] the difference between twelve perfect fifths and seven octaves, across four intervals. As a result, the wolf intervals are eliminated, the major thirds are sharper, four of the perfect fifths have been flattened, and there are once again different sized tones.

In well temperament, all the keys are usable, but they sound different due to the different size of the tones. Different keys are thought to correspond to different ''colors'' or moods. The frequency ratios used to construct a 12-tone scale in well temperament (Werckmeister III) are shown in Table I, column (e). The complete set of frequency ratios with each note used in turn as the root note was also found.

Barnes[9] analyzed the suitability of Werckmeister III as the temperament intended by Bach for the ''Well-Tempered Clavier,'' concluding that the piece was written for a temperament similar to that tuning and that such temperaments are suitable for its performance. However, Rasch[13] pointed out deficiencies in Barnes' methodology and concluded from an analysis of late 17th and early 18th century music theory that equal temperament is a ''particularly appropriate tuning'' for the ''Well-Tempered Clavier.''

D. Equal temperament

Equal temperament, defined by equally spaced half tones, gradually took over as the standard, replacing the well temperaments during the 19th Century. The frequency ratio between any two adjacent notes is $2^{1/12}$:1. The frequency ratios used to construct a 12-tone scale in equal temperament are shown in Table I, column (f). These values remain the same when the notes are modulated to other keys.

Although equal temperament appeared in references as far back as the 16th Century, it was not adopted sooner in part because it was difficult to tune keyboard instruments to equal temperament. There continues to be a school of thought that equal temperament is a poor compromise (see Doty[12]), since it does not use any ratio with small integers except for the octave. However, this study found that equal temperament performs strongly in approximating perfect harmony when all keys and all consonant intervals are given equal weighting compared to the particular cases of the other tunings examined and also performs well for the three ''Well-Tempered Clavier'' preludes examined.

III. GENERAL EVALUATION OF FOUR TUNINGS

A. Data

The frequency ratios given for each of the tunings in Table I and for perfect harmony in the introduction were converted into cents. One cent is defined as 1/1200 of an octave and, therefore, a ratio, r, is equivalent to x cents, where $x = 1200 \log_2 r$. The deviation between perfect harmony and each tuning is shown in Table II, with the absolute deviation being the absolute value of the data given in that table. Because each scale consists of 12 fixed pitch classes, no distinction is made between enharmonically equivalent intervals, and keys are referred to by the labels given in Table II.

Just noticeable differences in frequency are explained by Hall[19] based on the work by Wier, Jesteadt, and Green[20] and Harris[21] that characterized the sensitivity to changes in frequency for pure tones as a function of loudness and a reference frequency. For much of the frequency range of a piano, the just noticeable difference for pure tones is about 1 Hz, and for complex waveforms, it may be as little as 0.1 Hz at low frequencies. Discrimination between two tones depends on multiple factors, such as the choice of interval, absolute pitch level, tone duration, and the musical training of the listener. A rough rule of thumb states that differences of less than five cents are generally considered imperceptible. However, Hall reports that under certain conditions, such as in the case of a well-trained musician listening deliberately for mis-

TABLE II. Deviations between perfect harmony and four tunings, measured in cents. Each value represents the difference in the size of the interval created by a specific tuning (e.g., just intonation) and the interval defined as perfect harmony. For each tuning, the deviations are found for 72 intervals (six consonant intervals of interest times 12 keys).

	Key											
	C	C#	D	Eb	E	F	F#	G	G#	A	Bb	B
Just intonation (JI)												
m3	0.00	−41.06	−21.51	−41.06	0.00	0.00	−21.51	0.00	−41.06	0.00	−21.51	0.00
M3	0.00	0.00	0.00	0.00	41.06	0.00	41.06	0.00	0.00	41.06	0.00	41.06
p4	0.00	−19.55	0.00	0.00	0.00	21.51	0.00	0.00	0.00	21.51	0.00	0.00
p5	0.00	0.00	−21.51	0.00	0.00	0.00	19.55	0.00	0.00	0.00	−21.51	0.00
m6	0.00	−41.06	0.00	−41.06	0.00	0.00	0.00	0.00	−41.06	0.00	−41.06	0.00
M6	0.00	21.51	0.00	0.00	41.06	21.51	41.06	0.00	0.00	21.51	0.00	41.06
Meantone temperament (MT)												
m3	−5.38	−5.38	−5.38	−46.44	−5.38	−46.44	−5.38	−5.38	−5.38	−5.38	−46.44	−5.38
M3	0.00	41.06	0.00	0.00	0.00	0.00	41.06	0.00	41.06	0.00	0.00	41.06
p4	5.38	5.38	5.38	−35.68	5.38	5.38	5.38	5.38	5.38	5.38	5.38	5.38
p5	−5.38	−5.38	−5.38	−5.38	−5.38	−5.38	−5.38	−5.38	35.68	−5.38	−5.38	−5.38
m6	−41.06	0.00	0.00	−41.06	0.00	−41.06	0.00	0.00	0.00	0.00	−41.06	0.00
M6	5.38	46.44	5.38	5.38	5.38	5.38	46.44	5.38	46.44	5.38	5.38	5.38
Well temperament (WT)												
m3	−21.51	−15.64	−9.78	−21.51	−9.78	−21.51	−15.64	−15.64	−15.64	−3.91	−21.51	−15.64
M3	3.91	21.51	9.78	15.64	15.64	3.91	21.51	9.78	21.51	15.64	9.78	15.64
p4	0.00	0.00	5.87	0.00	0.00	0.00	5.87	5.87	0.00	5.87	0.00	0.00
p5	−5.87	0.00	−5.87	0.00	0.00	0.00	0.00	−5.87	0.00	0.00	0.00	−5.87
m6	−21.51	−15.64	−9.78	−15.64	−3.91	−21.51	−9.78	−15.64	−15.64	−3.91	−21.51	−9.78
M6	3.91	21.51	15.64	21.51	15.64	9.78	21.51	9.78	21.51	15.64	15.64	15.64
Equal temperament (ET)												
m3	−15.64	−15.64	−15.64	−15.64	−15.64	−15.64	−15.64	−15.64	−15.64	−15.64	−15.64	−15.64
M3	13.69	13.69	13.69	13.69	13.69	13.69	13.69	13.69	13.69	13.69	13.69	13.69
p4	1.96	1.96	1.96	1.96	1.96	1.96	1.96	1.96	1.96	1.96	1.96	1.96
p5	−1.96	−1.96	−1.96	−1.96	−1.96	−1.96	−1.96	−1.96	−1.96	−1.96	−1.96	−1.96
m6	−13.69	−13.69	−13.69	−13.69	−13.69	−13.69	−13.69	−13.69	−13.69	−13.69	−13.69	−13.69
M6	15.64	15.64	15.64	15.64	15.64	15.64	15.64	15.64	15.64	15.64	15.64	15.64

tuning, the ear may be much more discriminating. Perceptual studies such as performed by Vos[22] provide validation for this observation. In Vos' study, musically trained subjects were asked to rate tempered fifths and major thirds in terms of ''purity,'' where differences in purity ratings imply discrimination between notes. For comparisons between tones that each contained 20 harmonics, the mean subjective purity ratings for temperings of ± 2 cents with a tone duration of 0.5 s were statistically significantly different from the mean ratings for zero tempering.

Beyond the issue of whether a difference is noticeable is whether it is tolerable in terms of pleasant sounding. The amount of mistuning that is tolerated depends on the function of the mistuned tone within the local tonal hierarchy and the direction of the interval in the case of melody. Deviations of 10–20 cents are common in actual performances, since wind and string instruments, for example, easily vary by such amounts. Such deviations generally do not correlate with a perception of roughness or dissonance, whereas deviations greater than 25–30 cents tend to be unpleasant sounding and difficult to use in practice. Vos[22] reported that the strong difference in subjective purity ratings between pure and tempered intervals could be mainly attributed to the occurrence of beats or roughness.

Although just intonation and meantone have a large number of zero entries in Table II, these two tunings also are susceptible to large deviations, which limit the intervals and keys that can be played. Well temperament is successful in eliminating the wolf tones for which meantone was criticized, and similarly, equal temperament is characterized by moderate, rather than extreme, deviations from pure consonance.

B. Analysis and results

1. Measures of central tendency and dispersion

A statistical summary of the absolute values of the data in Table II shows how well each tuning performs. The smaller the absolute deviations, the closer a tuning is to perfect harmony. Two measures of central tendency and three measures of variability for the absolute deviations from perfect harmony are given in Table III.

Well temperament and equal temperament have the smallest means, that is, are closest to perfect harmony on average. Just intonation has a median of zero, which is the smallest of all the tunings. In this application, the mean seems a more useful measure than the median, since it explicitly incorporates very large deviations which would tend to be unpleasant sounding.

The results in Table III for the mean absolute deviations for meantone and equal temperament agree with those presented by Rasch[7] in his Table VI as the ''mean tempering of a tuning.'' The methodology used in this study allows as well

TABLE III. Measures of central tendency and dispersion for absolute deviations from perfect harmony. Two measures of central tendency, the mean and median, and three measures of dispersion, the variance, range, and interquartile range, are shown. Equal temperament performs as well as or better than the other tunings for three of the five parameters.

	Absolute deviations			
Measures	JI	MT	WT	ET
Mean	11.51	12.41	10.43	10.43
Median	0	5.38	9.78	13.69
Variance	270.09	264.42	60.74	36.53
Range	41.06	46.44	21.51	13.69
Interquartile range	21.51	0	11.73	13.69

for the calculation of this parameter for irregular tunings, including just intonation and well temperament.

Of greater interest are the measures of variability, which show how far each tuning deviates from perfect harmony. The variance and range of the absolute deviations show a trend of decreasing variability across the centuries, from meantone to well temperament to equal temperament, corresponding to the oldest to newest standard.[23] The interquartile range, which measures the difference between the 25th and 75th percentiles, shows the opposite trend, reflecting that the middle of the data for well temperament and equal temperament is more disperse. The zero interquartile range for meantone reflects that over half the data values were equal to 5.38, the value of the median. Just intonation has the largest interquartile range.

A graphical representation in a box-whisker plot, shown in Fig. 1, illustrates the wider variability for just intonation and meantone temperament. The ends of the ''box'' correspond to the 25th and 75th percentiles, with a line drawn at the 50th percentile, i.e., the median. The ''whiskers'' or lines at the ends of the figure correspond to the 2.5th and 97.5th percentiles, thereby capturing 95% of the data between those values. The graphs show how compact well temperament and equal temperament are overall, even though they have

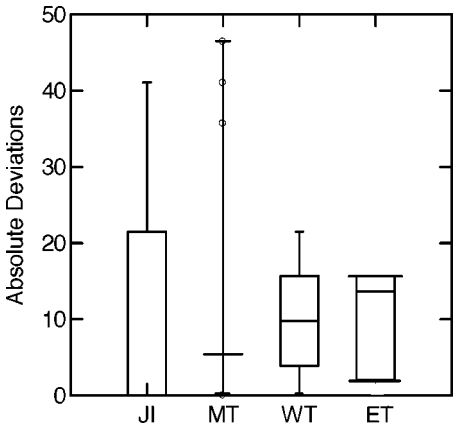


FIG. 1. Box-whisker plots of the absolute deviations of the tunings from perfect harmony. Each plot depicts the 2.5th, 25th, 50th, 75th, and 97.5th percentiles for one of the tunings, just intonation (JI), meantone temperament (MT), well temperament (WT), and equal temperament (ET). Only meantone has extreme points, designated by circles. Well temperament and equal temperament have less overall spread than just intonation and meantone temperament.

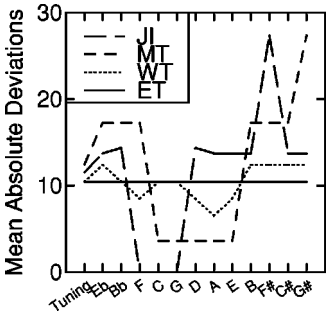


FIG. 2. Mean absolute deviations by key. The mean absolute deviation between each tuning and perfect harmony by key, as well as for the overall tuning, is displayed for each tuning. Just intonation and meantone perform the best for C and nearby keys, whereas equal temperament and well temperament exhibit less overall variability and better performance for distant keys.

large interquartile ranges. The plot also shows extreme points, values that are more than 3.0 times the interquartile range from the 25th and 75th percentiles. Only meantone temperament has such extreme points.

Another view of the variability of each tuning is provided by an examination of the mean absolute deviations from perfect harmony by key. The results are subject to the selection of C as the starting note for each tuning. As shown in Fig. 2, just intonation and meantone are close to perfect harmony for C and nearby keys, arranged in order based on fifths, but have the most extreme deviations for distant keys. On the other hand, equal temperament does not vary by key, and well temperament exhibits modest variation. These patterns of variability between keys underlie the results displayed in the box-whisker plots of Fig. 1.

Of the five measures considered in this section, two of them—the mean and variance—are explicit functions of all deviations. These seem the best suited of the measures considered here to summarize the performance of the tunings since they account for large deviations, which tend to be unpleasant sounding. For both these measures, equal temperament is as good as or better than the other tunings, with well temperament also performing strongly.

2. Distance measures

The second stage of the analysis is based on calculating the distance between each of the four tunings and perfect harmony. Six different distance matrices, Euclidean, Chebyshev, City Block, Canberra, Minkowski, and Correlation distances, are used to determine whether the results will be invariant to the choice of metric. Each of the distance metrics is an explicit function of all deviations, including the largest ones. The definitions of the distance metrics and the results are provided in Table IV.

The distance results are consistent across the six measures. Equal temperament is closer to perfect harmony than any other tuning for five of the six distance measures and tied with well temperament on the sixth measure. Well temperament also performs strongly across all six measures.

Two of the distance measures in Table IV are comparable to two other measures reported by Rasch.[7] The quadratic mean tempering of a tuning used by Rasch is a con-

TABLE IV. Distances between four tunings and perfect harmony. The names and general definitions of the distance metrics are given in the first two columns; x_{ri} is the value in cents of the ith interval of the rth tuning and x_{si} is the associated value for perfect harmony. The remaining columns show the actual distances between perfect harmony and just intonation (JI), quarter-comma meantone temperament (MT), well temperament (Werckmeister III) (WT), and equal temperament (ET), respectively. The Canberra and Correlation distances are unitless, while the other metrics are in cents.

Measures	Definition	JI	MT	WT	ET
Euclidean distance	$\delta_{rs}=\sqrt{\sum_i (x_{ri}-x_{si})^2}$	170.27	173.57	110.46	102.27
Chebyshev distance	$\delta_{rs}=\max_i \lvert x_{ri}-x_{si}\rvert$	41.06	46.44	21.51	15.64
City block distance	$\delta_{rs}=\sum_i \lvert x_{ri}-x_{si}\rvert$	828.99	893.51	750.78	750.78
Canberra distance	$\delta_{rs}=\sum_i \frac{\lvert x_{ri}-x_{si}\rvert}{x_{ri}+x_{si}}$	0.8311	0.8882	0.7612	0.7610
Minkowski distance (with $w_i=1$ and $\lambda=3$)	$\delta_{rs}=\left(\sum_i w_i \lvert x_{ri}-x_{si}\rvert^{\lambda}\right)^{1/\lambda}$	102.71	107.77	60.32	53.55
Correlation distance	$\delta_{rs}=1-\frac{\Sigma_i (x_{ri}-\overline{x_r})(x_{si}-\overline{x_s})}{\sqrt{\Sigma_i (x_{ri}-\overline{x_r})^2 (x_{si}-\overline{x_s})^2}}$	0.0021	0.0022	0.0009	0.0008

stant c times the Euclidean distance, where the constant is the number of intervals in the calculation raised to the negative one-half power [$c=(72)^{-1/2}$]. Also, the maximum tempering of a tuning equals the Chebyshev distance between each tuning and perfect harmony. The generalized distance definitions provided here simplify the calculations for irregular tunings.

Figure 3 provides a graphical representation of the data in Table IV as star plots to highlight the differences between the tunings. Each star plot consists of six spokes, each corresponding to one of the distance metrics. For a given spoke, the maximum value in Table IV for a particular distance metric is plotted at the extreme end of a spoke and all other values for that metric are plotted proportionately. The star plot for equal temperament has the smallest area, showing that it is the closest to perfect harmony on these distance measures. The star plot for well temperament also has a relatively small area compared to just intonation and meantone temperament.

The distance measures were also examined by key, with the overall patterns for each tuning closely resembling that shown in Fig. 2 for the mean absolute deviations.

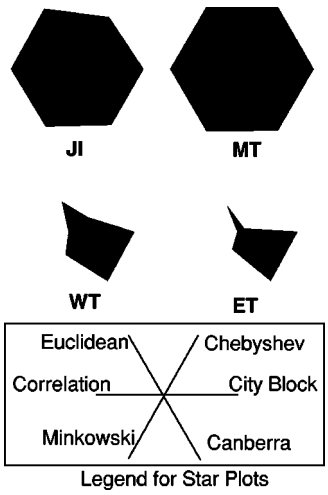


FIG. 3. Star plots of six distance measures. Six metrics, Euclidean, Chebyshev, City Block, Canberra, Minkowski, and Correlation, are used to measure the distance between a tuning and perfect harmony. The largest value for a metric is used to scale the plot on the corresponding spoke of the star plot. The star plot for equal temperament has the smallest area, which means it is the closest to perfect harmony on these measures. The plot for well temperament also has a relatively small area compared to just intonation and meantone.

3. Measure of dissonance

As explained by Sethares,[24] the concept of sensory consonance implies that consonance and dissonance depend not just on the interval between tones, but also on the spectra of the tones, which correlate with timber. The roughness or tonal dissonance experienced for two simultaneously sounded complex tones can be accounted for by interfering harmonics. The relationship between a perception of ''impurity'' and interfering harmonics is consistent with the primary findings of the perceptual experiments conducted by Vos.[22]

Based on the work of Plomp and Levelt[11] for pure tones, Sethares[10] developed a model for measuring the dissonance of two complex tones with base frequencies f and αf, as follows:

$$D=D_F+D_{\alpha F}+\sum_{i=1}^{n}\sum_{j=1}^{n} d(f_i,\alpha f_j,\nu_i,\nu_j), \tag{1}$$

where f_i and αf_i are the partials of the two tones, ν_i are the amplitudes of the partials, and D_F and $D_{\alpha F}$ are the dissonances for each tone calculated from its own partials, with

$$D_F=1/2\sum_{i=1}^{n}\sum_{j=1}^{n} d(f_i,f_j,\nu_i,\nu_j), \tag{2}$$

$$d(f_i,f_j,\nu_i,\nu_j)=\nu_i\nu_j(e^{-3.5s(f_i-f_j)}-e^{5.75s(f_i-f_j)}), \tag{3}$$

TABLE V. Dissonance ratios based on a model of sensory consonance. Two different sets of amplitudes were used for the harmonic partials of the two complex tones in an interval. These were incorporated into Eq. (1) to obtain an absolute dissonance measure for each tuning, summed across the 72 intervals of Table II. Each dissonance measure was then normalized by forming the ratio of the absolute dissonance to the corresponding dissonance level for perfect harmony.

	Dissonance ratios			
Amplitudes	JI	MT	WT	ET
Based on 0.88	1.143	1.179	1.164	1.158
Based on $1/n$	1.086	1.108	1.087	1.081

$$s = \frac{0.24}{0.021 f_i + 19}, \tag{4}$$

and $f_i < f_j$. Sethares created dissonance curves by letting α vary continuously over an interval such as $1 \leqslant \alpha \leqslant 2$. In this analysis, α assumes the values associated with the consonant intervals in Table I. Then a measure of dissonance for the 72 consonant intervals in Table II is found by summation and can be compared to the mean absolute deviation for the same 72 intervals. Since the measure of dissonance, *D*, depends on actual frequencies, not just the ratio of frequencies, its values will change for notes within different octaves. The calculations are performed by setting C equal to a frequency of 261.63 Hz. Two sets of amplitudes are used in the analysis: (1) following the analysis of Sethares,[10] amplitudes are set to fall at a rate of 0.88, and (2) as suggested by Doty[12] as typical of many common types of musical tones, amplitudes are decreased in direct proportion to their ordinal number (for example, the second harmonic has one-half the amplitude of the first). In both cases, $n = 6$ harmonic partials are used, similar to the analyses of Sethares.

A comparable measure of dissonance is also found for the perfect harmony intervals, and then the results for each tuning are calculated as a ratio of the dissonance for the tuning compared to the dissonance for perfect harmony. With this dissonance ratio measure, 1.0 equals perfect harmony and values closer to 1.0 indicate a tuning is closer to perfect harmony.

The results for the four tunings, presented in Table V, indicate the strongest performances are associated with just intonation and equal temperament and the weakest with meantone temperament. Compared to the earlier analysis based strictly on frequencies, rather than the spectra, just intonation is viewed as more consonant and well temperament as less so. However, overall there is little difference between the four tunings due to the added complexity introduced by the interfering harmonics.

To examine the differences by key, Fig. 4 displays the dissonance ratio by key with the amplitudes based on $1/n$. As before, the results for the measure of dissonance are subject to the selection of C as the starting note for each tuning. The general pattern by key is similar for the other set of amplitudes considered. Many similarities between Figs. 2 and 4 are apparent. For example, just intonation and mean temperament again perform best for the key of C and nearby keys and reach their most extreme values for distant keys.

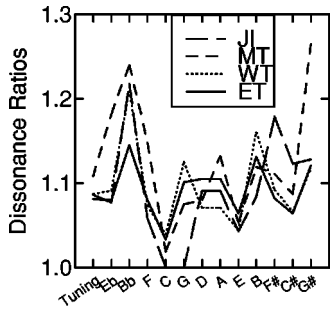


FIG. 4. Dissonance ratios by key. The dissonance ratio measure for each tuning compared to perfect harmony is displayed by key, as well as for the overall tuning. Just intonation and meantone perform the best for C and nearby keys, whereas equal temperament and well temperament exhibit less overall variability and better performance for distant keys. The performance pattern is similar to that of Fig. 2, indicating that the mean absolute deviation based on frequencies is in reasonable agreement with the dissonance ratio based on the spectra analyzed.

Similarly, equal temperament and meantone exhibit less variability than the other two tunings. However, they show considerably more variability than in Fig. 2 because of the dependence of the dissonance measure on the absolute frequencies. Equal temperament, in particular, has a variable dissonance ratio, unlike its mean absolute deviation, since the dissonance ratio is a function of the frequencies of the tones in an interval and the beats measured by their difference. Also, although in general the measure of dissonance for a given tuning decreases as the frequency increases, this is not the trend for the ratio of the dissonance of a tuning to that of perfect harmony. Nonetheless, the similarity between Figs. 2 and 4 indicates that the simple measure of mean absolute deviation provides a reasonable approximation to the dissonance ratio based on a perceptual model. Several of the distance measures discussed earlier also provide a close approximation, but the strongest relationship based on correlation measures exists between the dissonance ratio and the mean absolute deviation.

In this section, the four historical tunings have been compared to perfect harmony by means of a dissonance model derived from the results of perceptual experiments conducted by Plomp and Levelt. In this way, the dissonance measure provides an assessment of the tunings that more closely reflects human perception of consonance and dissonance and in particular accounts for the roughness associated with interfering harmonics. The later perceptual experiments conducted by Vos[22] found that the primary factor that differentiated the ratings of pure and tempered intervals was the presence of beats or roughness and that deletion of harmonic interference resulted in higher ratings. Although his research asked subjects to judge purity of tempered intervals, Vos noted that it is reasonable to assume that purity and consonance are comparable concepts. Additionally, an examination of the intervals in this study that were also examined by Vos (namely the fifths and major thirds corresponding to the regular tunings of equal temperament and meantone, as well as perfect harmony) indicates a similarity of results. The correlation coefficient between the measures of dissonance and the subjective impurity ratings for these intervals, as reported in Vos,[25] is 0.83. As tempering increases, the intervals have

TABLE VI. Weights for the consonant intervals appearing in three Bach preludes. For each piece of music, the number of beats for each of the consonant intervals was counted and classified by root note. The counts of beats were converted into weights by scaling each count by the total number of beats for a piece of music. The cells with a value of zero correspond to intervals that do not occur in the music.

	Key												
Prelude 1, C Major													
	C	C#	D	Eb	E	F	F#	G	G#	A	Bb	B	Total
m3	0.004	0	0.048	0	0.060	0.015	0.004	0.019	0	0.023	0	0.047	0.220
M3	0.130	0	0.008	0	0	0.035	0	0.035	0.004	0	0	0	0.214
p4	0.017	0	0.033	0	0.029	0	0	0.033	0	0.013	0	0.005	0.130
p5	0.079	0	0.027	0	0	0.021	0	0.081	0	0.004	0	0	0.213
m6	0	0	0	0	0.050	0	0	0.015	0	0	0	0.023	0.088
M6	0.029	0	0.013	0.004	0	0.021	0	0.042	0.019	0.004	0.004	0	0.136
Total	0.260	0	0.129	0.004	0.140	0.092	0.004	0.225	0.023	0.044	0.004	0.075	1.000
Prelude 3, C# Major													
	C	C#	D	Eb	E	F	F#	G	G#	A	Bb	B	Total
m3	0.023	0.002	0.007	0.058	0	0.106	0.005	0	0.005	0	0.016	0	0.222
M3	0	0.062	0	0.005	0.005	0	0.032	0	0.039	0.005	0	0.005	0.152
p4	0.005	0.018	0	0.044	0	0.005	0.005	0	0.079	0	0.014	0	0.169
p5	0.002	0.088	0	0.012	0	0	0.009	0	0.025	0	0.009	0.005	0.150
m6	0.035	0	0.005	0.007	0	0.072	0	0	0	0	0.023	0	0.141
M6	0.002	0.044	0	0.030	0.002	0	0.030	0	0.039	0	0.002	0.016	0.166
Total	0.067	0.215	0.012	0.155	0.007	0.182	0.081	0	0.187	0.005	0.065	0.025	1.000
Prelude 5, D Major													
	C	C#	D	Eb	E	F	F#	G	G#	A	Bb	B	Total
m3	0	0.018	0.035	0.003	0.027	0	0.083	0.012	0.038	0.005	0	0.029	0.249
M3	0.002	0	0.065	0	0	0.002	0.002	0.011	0	0.051	0.002	0.002	0.134
p4	0	0	0.002	0	0.027	0	0.002	0	0	0.119	0	0.002	0.151
p5	0	0	0.066	0	0.006	0	0.006	0.011	0	0.080	0.002	0.003	0.174
m6	0	0.008	0.002	0.002	0	0	0.060	0	0.009	0.041	0.002	0.002	0.124
M6	0.002	0	0.042	0	0.018	0.026	0	0.023	0	0.044	0.003	0.011	0.168
Total	0.003	0.026	0.211	0.005	0.079	0.027	0.153	0.056	0.047	0.340	0.008	0.047	1.000

larger dissonance measures, with major thirds having larger dissonance measures than fifths. However, the major thirds are less sensitive to large temperings than are the fifths, which is consistent with how equal temperament treats these two intervals. The dissonance measures for the fifths tend to follow an exponential curve, as modeled by Vos, both when examined across a tuning and for each key separately. Future work with perceptual experiments that incorporate more than regular tunings and additional intervals would be useful for comparison with modeled results.

IV. SPECIFIC EVALUATION OF TUNINGS BASED ON THREE PRELUDES

The analysis presented in Sec. III provides a general characterization of the performance of the four tunings using a variety of quantitative measures. Hall[2] argued that objective measurement of how well a tuning performs should be calculated with respect to specific pieces of music. Since a given piece will not likely include all consonant intervals in equal proportion, the performance of a tuning will vary from that described in the general case. This section presents an analysis for three preludes from Bach's ''Well-Tempered Clavier'' for which intervals are weighted in proportion to the duration of their occurrence in the music.

A. Data

Data were collected for three pieces from the ''Well-Tempered Clavier, Book I:'' Prelude 1, C Major; Prelude 3, C# Major; and Prelude 5, D Major. When the notes in a chromatic scale are arranged in order based on fifths, they stand as: Eb Bb F C G D A E B F# C# G#. Therefore, even though C, C#, and D are adjacent tones in a single scale, C and D are central keys, while C# is a peripheral key.

For each piece of music, the number of beats for each of the consonant intervals (m3, M3, p4, p5, m6, and M6) was counted and classified by root note. For example, in Prelude 1, perfect fifths occur for a duration of 51 equivalent beats, 19 of which have C as the root note. Chords were separated into their component two-note intervals. The counts of beats were converted into weights by scaling each count by the total number of beats for a piece of music. The resulting weights for the three preludes are given in Table VI. The cells with a value of zero correspond to intervals that do not occur in the music. This method of weights differs from that of Hall[2] in that the duration, not just the occurrence, of an interval is used to determine the weight.

A review of the pattern of zeros, as well as the rows and columns labeled ''Total'' in Table VI, indicates that all intervals do not appear in approximately equal proportion. For example, in Prelude 3, only 1.2% of all consonant intervals have D as a root note, compared to 21.5% for the tonic C#. Although Barnes[9] argues that intervals should occur more or less often according to their tuning quality, this seems too restrictive, as a piece written for equal temperament should not be expected to have intervals based on all notes appear in equal proportion. Charts of duration versus the absolute deviations from perfect harmony do show that intervals with large absolute deviations tend to be correlated with small

TABLE VII. Weighted measures of central tendency, dispersion, and distance for the absolute deviations from perfect harmony for three Bach preludes and the overall tuning. The weights in Table VI are applied to the absolute deviations to obtain weighted measures for the preludes. Equal weights of 1/72 are used for the overall tuning. Overall, just intonation and meantone temperament exhibit the most variability between different pieces of music, whereas well temperament and equal temperament are more stable and more moderate in size of deviations.

		Absolute deviations			
Measures	Tuning/prelude	JI	MT	WT	ET
Weighted mean	C Major prelude	2.51	5.30	7.57	10.36
	C# Major prelude	10.16	29.23	14.21	10.71
	D Major prelude	14.81	4.36	9.73	10.69
	Tuning	11.51	12.41	10.43	10.43
Weighted variance	C Major prelude	128.75	120.73	33.81	37.49
	C# Major prelude	268.09	622.53	110.15	36.54
	D Major prelude	239.56	83.85	30.50	37.39
	Tuning	270.09	264.42	60.74	36.53
Weighted Euclidean distance	C Major prelude	7.35	9.91	9.11	12.03
	C# Major prelude	19.22	34.55	17.25	12.29
	D Major prelude	21.17	6.17	11.17	12.31
	Tuning	20.07	20.46	13.02	12.05
Weighted Chebyshev distance	C Major prelude	21.51	46.44	21.51	15.64
	C# Major prelude	41.06	46.44	21.51	15.64
	D Major prelude	41.06	46.44	21.51	15.64
	Tuning	41.06	46.44	21.51	15.64

durations; however, some intervals with small or zero absolute deviations are also associated with small durations.

Finally, the weights in Table VI were combined with the absolute deviations derived from Table II to create performance measurements discussed in the next section. One limitation of this analysis is the need to assume a particular key as the starting point for each tuning, in this case the key of C. This choice of disposition necessarily impacts the results. Therefore, the results reported here are illustrative and would not necessarily be the same if another disposition were selected.

B. Analysis and results

1. Measures of central tendency and dispersion

Two measures of central tendency and dispersion for the weighted absolute deviations between each tuning and perfect harmony for the three Bach preludes are included in Table VII. Because of the similarity of results between measures, only a subset of the measures discussed in Sec. III appears in this and the following section. The ones selected are explicit functions of all deviations. Table VII also includes the values for the overall tuning for comparison, with each interval given equal weight in the calculations.

The weighted mean of the absolute deviations demonstrates the variability in meantone temperament between pieces. Although meantone performs well for the C major and D major preludes, the frequent occurrence in the C# prelude of minor and major thirds and sixths with large absolute deviations from perfect harmony contribute to the large mean for that piece. Meantone temperament has not been proposed as a likely tuning for the ''Well-Tempered Clavier,'' and the weighted mean absolute deviation supports that position. Just intonation, as defined in Table I, performs well for the C major prelude, but is the worst of the four tunings for the D major prelude. Well temperament and equal temperament demonstrate strong consistency from piece to piece. Although they are not always the best tuning, and equal temperament is sometimes the worst, both perform well across the different preludes.

Although Rasch[7] does not analyze specific pieces of music, he introduces a weighting scheme based on the circle of fifths to determine the mean tempering of keys. Mean tempering is equivalent to the mean absolute deviations reported here. Based on a simplified view of the relative importance of intervals within a key, Rasch concludes that for meantone, the key of C major has the smallest mean tempering and that for the central keys of F, C, G, and D major, meantone tuning is the best tuning of the regular tunings considered. The analysis of the three preludes from the ''Well-Tempered Clavier,'' which is based on weights derived from the music, shows that meantone performed slightly better for D major than C major, but considerably worse for C#, a peripheral key. Since Rasch's results are based on a simplified assumption on weights, small differences in findings such as reported here should be expected based on individual music. Also, just intonation outperformed meantone in the key of C, indicating how an irregular tuning may be closer to perfect harmony for some pieces.

Table VII also presents a measure of dispersion, the weighted variance for the absolute deviations from perfect harmony for the three preludes and overall tuning. Meantone temperament performs poorly for the C# major prelude, and just intonation demonstrates considerably more variability than equal temperament. Both well temperament and equal temperament perform the best on this measure across the preludes. The results in this section depend on the initial selection of C as the starting note for each tuning and are applicable to the particular version of just intonation defined in Table I.

2. Distance measures

Two distance measures for the three preludes and the overall tuning are included in Table VII. The values for the Euclidean distance for the overall tuning differ from the results in Sec. III by the constant $c=(72)^{-1/2}$ and equal the root mean square reported by Rasch,[7] while the Chebyshev distance equals his maximum tempering of a tuning. Additionally, the Euclidean distance metric used for the preludes equals the square root of the goodness-of-fit measure used by Hall.[2]

The weighted Euclidean distance between each tuning and perfect harmony shows similar patterns to those found for the weighted mean and variance. All four tunings perform well for the C major prelude, while well temperament and equal temperament continue to perform best across the pieces. The pattern for the Chebyshev distance differs from that of the other statistics. The maximum absolute deviation differs from that of the overall tuning only if intervals associated with the maximum value never appear in the piece of music. This occurs only once—for the C major prelude. In

this case, the minor and major thirds and sixths with an absolute deviation from perfect harmony of 41.06 for just intonation do not appear.

The analysis of the three preludes has shown that there can be considerable variability in the measures of central tendency, dispersion, and distance for a given tuning across different pieces, as well as between tunings for a single-piece. Perceptual experiments have also found this to be the case when musically trained subjects rate performances of different musical pieces, such as in the experiments conducted by Vos.[25] In that study, subjects were presented with musical fragments according to one of seven regular tunings (including equal temperament, meantone temperament, and perfect harmony) and asked to rate their acceptability. Vos determined through multiple regression that acceptability is highly correlated with the subjective purity ratings of isolated fifths and major thirds, suggesting that the acceptability ratings reflected a judgment of consonance. An analysis of variance of the experimental data revealed that mean acceptability ratings differ as a function of the tuning, the musical fragment, and the individual subjects, as well as interactions between them. When averaged across fragments and subjects, the mean acceptability ratings were about the same for tunings with between 0 and -5.4 cents deviations from perfect harmony for fifths (which includes equal temperament, meantone, and perfect harmony). However, the main effect of musical fragments was related to the mean absolute tempering, and a strong correlation was found between mean absolute tempering and overall acceptability. These results suggest that the mean absolute deviations (which are equivalent to mean absolute tempering) provide a reasonable approximation to perceptual ratings of consonance.

V. CONCLUSION

The general evaluation of the four historical tunings using measures of central tendency, measures of dispersion, distance measures, and a dissonance ratio support the conclusion that equal temperament performs strongly in approximating perfect harmony when all keys and all consonant intervals are given equal weighting compared to the particular cases of 5-limit just intonation, quarter-comma meantone temperament, and well temperament (Werckmeister III) analyzed.

Additionally, both equal temperament and well temperament provide the most consistently strong performances as tunings for three preludes from Bach's ''Well-Tempered Clavier.'' The findings here agree with the results of Rasch[13] who demonstrated that equal temperament is an appropriate tuning for the ''Well-Tempered Clavier'' based on evidence from German music theory. The analysis here is not intended to justify any specific temperament as the one Bach intended for this piece, but rather to use quantitative measurements of performance compared to perfect harmony to judge the suitability of these tunings. The results for the specific preludes are subject to the selection of C as the starting note for each tuning.

The measures of central tendency, dispersion, distance, and dissonance used in both the general and composition-specific analyses can be readily applied to other tunings and other pieces of music to evaluate both relative and absolute performance. These measures include ones that are equivalent to those defined by Rasch[7] for regular tunings and by Hall[2] for specific pieces of music. As presented here, these measures can be readily interpreted as measuring centrality and variability of a tuning compared to the standard of perfect harmony, as well as overall distance from pure consonance. The mean absolute deviation appears to be the measure best correlated with the dissonance ratio that reflects sensory dissonance, as well as with the results of perceptual studies of subjective acceptability. Comparing the mean absolute deviation with the results of perceptual experiments for both regular and irregular tunings should be considered for further research.

ACKNOWLEDGMENTS

The author would like to acknowledge Judith Page, Stephen McAdams, and two anonymous reviewers for their insightful comments and suggestions.

[1] J. Tenney, *A History of "Consonance" and "Dissonance"* (Excelsior Music, New York, 1988).
[2] D. E. Hall, ''Quantitative evaluation of musical scale tunings,'' Am. J. Phys. **42**, 543–552 (1974).
[3] The term ''temperament'' is used to describe a scale for which the frequency ratios deviate from simple small integer ratios. Intervals associated with such a scale are referred to as tempered intervals, in contrast to pure intervals. The general term ''tunings'' is used here to include temperaments.
[4] J. S. Bach, *The Well-Tempered Clavier*, Reprint with corrections and explanation of ornaments by Saul Novack (Dover, Toronto, 1983).
[5] J. Vos and B. G. van Vianen, ''Thresholds for discrimination between pure and tempered intervals: the relevance of nearly coinciding harmonics,'' J. Acoust. Soc. Am. **77**, 176–187 (1985).
[6] D. E. Hall, ''The objective measurement of goodness-of-fit for tunings and temperaments,'' J. Music Theory **17**, 274–290 (1973).
[7] R. A. Rasch, ''Description of regular twelve-tone musical tunings,'' J. Acoust. Soc. Am. **73**, 1023–1035 (1983).
[8] Rasch (1983) defines regular tunings as ones that have 11 fifths with the same amount of tempering (which may be no tempering).
[9] J. Barnes, ''Bach's keyboard temperament,'' Early Music **7**, 236–249 (1979).
[10] W. A. Sethares, ''Local consonance and the relationship between timbre and scale,'' J. Acoust. Soc. Am. **94**, 1218–1228 (1993).
[11] R. Plomp and W. J. M. Levelt, ''Tonal consonance and critical bandwidth,'' J. Acoust. Soc. Am. **38**, 548–560 (1965).
[12] D. B. Doty, *The Just Intonation Primer*, 3rd ed. (The Just Intonation Network, San Francisco, 2002).
[13] R. Rasch, ''Does 'well-tempered' mean 'equal-tempered'?'' in *Bach, Handel, Scarlatti Tercentenary Essays*, edited by P. Williams (Cambridge University Press, Cambridge, 1985), pp. 293–310.
[14] J. M. Barbour, *Tuning and Temperament: A Historical Survey* (Michigan State College Press, East Lansing, 1951).
[15] The ratios, when multiplied times an arbitrary frequency selected for the tonic, produce the frequencies for the notes in the scale. Because consonance is defined by frequency ratios, rather than absolute frequencies, Table I focuses on the intervals relative to the tonic associated with each tuning.
[16] Although proponents of just intonation can create a large number of pure tones by means of programmable synthesizers, this study uses only the 12 tones defined for just intonation by Table I in order to compare 12-note tuning methods.
[17] The term comma refers to a small interval that usually arises due to two different methods of determining a tone or interval. The syntonic comma is the difference between a major third in Pythagorean tuning, which has a ratio of 81/64, and a major third in 5-limit just intonation, which has a ratio of 5/4. This difference is the interval 81/80. The difference between

two ratios is calculated as one ratio divided by the other. In this case, (81/64)/(5/4) yields $(81/64)\times(64/80)=81/80$.

[18] The Pythagorean comma is calculated as $(3/2)^{12}/2^7=531441/524288$.

[19] D. E. Hall, *Musical Acoustics* (Brooks/Cole, Pacific Grove, CA, 2002).

[20] C. C. Wier, W. Jesteadt, and D. M. Green, ''Frequency discrimination as a function of frequency and sensation level,'' J. Acoust. Soc. Am. **61**, 178–1845 (1977).

[21] J. D. Harris, ''Pitch discrimination,'' J. Acoust. Soc. Am. **24**, 750–755 (1952).

[22] J. Vos, ''Purity ratings of tempered fifths and major thirds,'' Music Percept. **3**, 221–258 (1986).

[23] Just intonation was popular with theorists but was never in widespread use as a standard tuning.

[24] W. A. Sethares, *Tuning, Timbre, Spectrum, Scale* (Springer, London, 1998).

[25] J. Vos, ''Subjective acceptability of various regular twelve-tone tuning systems in two-part musical fragments,'' J. Acoust. Soc. Am. **83**, 2383–2392 (1988).

The dynamics and tuning of orchestral crotales

Bradley M. Deutsch, Cherie L. Ramirez, and Thomas R. Moore[a)]
Department of Physics, Rollins College, Winter Park, Florida 32707

(Received 8 June 2004; revised 30 June 2004; accepted 13 July 2004)

An experimental and theoretical investigation of the acoustic and vibrational properties of orchestral crotales within the range C_6 to C_8 is reported. Interferograms of the acoustically important modes of vibration are presented and the frequencies are reported. It is shown that the acoustic spectra of crotales are not predicted by assuming that they are either thin circular plates or annular plates clamped at the center, despite the physical resemblance to these objects. Results from finite element analysis are presented that demonstrate how changing the size of the central mass affects the tuning of the instruments, and it is concluded that crotales are not currently designed to ensure optimal tuning. The possibility of using annular plates as crotales is also investigated and the physical parameters for such a set of instruments are presented. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1788728]

PACS numbers: 43.75.Kk, 43.40.Dx [NHF] Pages: 2427–2433

I. INTRODUCTION

In the field of musical acoustics percussion instruments are understood especially well. To our knowledge, however, there is a nearly complete absence of discussion of orchestral crotales in the literature. The single exception appears to be a short mention of their acoustic properties by Fletcher and Rossing.[1] While the term *crotale* can be associated with several different types of percussion instruments, commercially available orchestral crotales offer little diversity; they are small cymbals with a central mass, as illustrated in Fig. 1. These instruments are commonly found in orchestras around the world and are commercially produced in the United States by at least two large manufacturers of percussion instruments. Each crotale in a set is tuned to one note of the Western musical scale and the note is stamped onto it for identification. The sound is usually produced by striking the instument with a mallet.

Crotales have a particularly pleasing sound, owing to the fact that the dominant partials are the second, fourth, and seventh harmonics of a nonexistent fundamental. The fortunate arrangement of these partials is clearly the result of the presence of the central mass, but to our knowledge there is no published discussion of the subject.

While the outer radii of the crotales become smaller as the pitch increases as one would expect, the radii of the central masses of the crotales do not change between C_6 and C_8. The invariability of the radius of the center mass leads one to question whether the tuning of each crotale is optimal. Here we report on an investigation of the acoustic and vibrational properties of a set of crotales in the two octaves from C_6 to C_8, identifying the vibrational modes and assessing their relative importance to the sound of the instruments. Evidence is presented demonstrating that the crotales are indeed not optimally tuned.

To assess the importance of the center mass on the normal modes of the crotale well-established thin plate theory is used to predict the normal mode frequencies. There is poor agreement between predicted and empirical values, and we conclude that the presence of the center mass is responsible for this discrepancy.

The center mass of a crotale causes it to physically resemble an annular plate that is free to vibrate at the outer radius and clamped at the inner radius. Rigid mounting through the center hole reinforces this resemblance. We therefore develop a model of the crotales as annular plates. Again, well-established theory is used to compare theoretical predictions with experimental results, and they are again found to be in poor agreement, although the agreement is better than is found when comparing experimental results to thin plate theory.

We finally turn to finite element analysis to facilitate an understanding of the effects of the center mass on the tuning of crotales. Within the model the height and radius of the central mass of the crotales are varied. The effect of these changes on the tuning of the crotales is then analyzed.

We conclude by presenting an alternative design for the manufacture of crotales based on annular plate theory. We present physical parameters for annular plates such that they have similar acoustic properties to crotales. The validity of this design as an alternative to commercially available crotales is confirmed using finite element analysis.

II. EXPERIMENT

A. Identification of the acoustically important modes

The crotales used in this investigation were manufactured by Zildjian Co. They span the octaves from C_6 to C_8 and have diameters ranging from 132.8 to 76.5±0.1 mm. All of the central masses are identical and have a diameter of 29.3±0.1 mm and a thickness of 13.1±0.1 mm. The thickness of the thin plate portion of the crotales is also uniform, measuring 4.7±0.1 mm. In order to determine the acoustic properties of the crotales each one was mounted on a one-inch diameter vibration-damping post that was secured to a vibration-isolated optical table in an anechoic chamber. The crotale was struck with a cork mallet and the sound was

a)Electronic mail: tmoore@rollins.edu

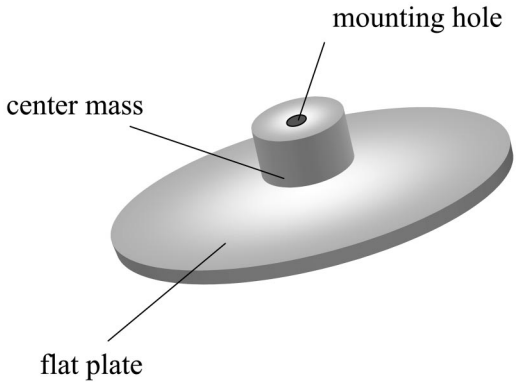


FIG. 1. Illustration of a crotale.

digitized. Two time series were recorded at a sampling rate of 40 kHz for each crotale in the set. The first series began at the strike time and had a duration of 0.25 s. A power spectrum of this time series was used to approximately identify the frequencies of many of the normal modes. The second time series was begun two seconds after the strike and had a duration of two seconds. Since this time series began after the transient modes had decayed to negligible relative power, the modes that are important in the steady-state sound of the crotales were evident.

Figure 2 is a typical example of the steady-state power spectrum of a struck crotale. Three modes are clearly visible, with most of the power being contained in the first two modes. As is commonly seen in other percussion instruments, the degenerate mode doublets are occasionally split due to slight asymmetries in the plate;[2] this is clearly evident in one of the modes shown in Fig. 2. Results from the other crotales within the set are similar, though in some cases the third mode is negligibly small and often none of the degenerate modes exhibit measurable splitting. We define an acoustically important mode as one that contains at least one percent of the total power, and using this definition there are at most three acoustically important modes for each crotale. In excess of 95% of the total power is contained within these three modes for all of the crotales, with no other single mode containing more than a small fraction of a percent of the total power.

Time-averaged electronic speckle pattern interferometry

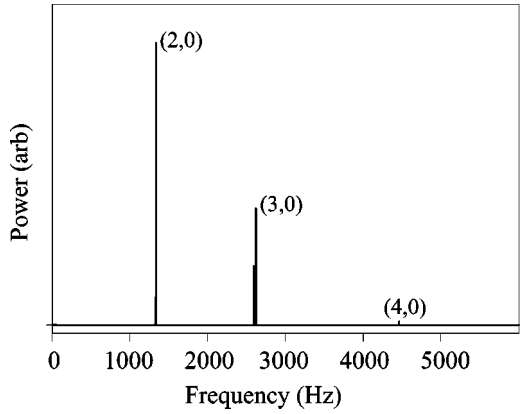


FIG. 2. Typical power spectrum of a crotale. The three acoustically important modes are clearly visible.

TABLE I. Diameters and frequencies of acoustically important modes of the two-octave set of crotales. Frequency uncertainties are ±0.25 Hz.

Note	Diameter ±0.1 mm	(2,0) mode	(3,0) mode	(4,0) mode
C_6	132.8	1055.5	2117.5	3667.5
$C^{\#}$	130.1	1117.0	2230.0	3849.5
D	125.9	1184.0	2348.0	4049.5
$D^{\#}$	123.8	1259.0	2483.0	4270.0
E	120.8	1333.5	2621.0	4502.5
F	117.8	1411.5	2754.5	4731.5
$F^{\#}$	114.6	1495.5	2879.0	4927.0
G	111.5	1585.5	3021.5	5166.5
$G^{\#}$	108.0	1682.0	3145.5	5361.0
A	106.6	1784.5	3357.5	2708.5
$A^{\#}$	104.6	1889.5	3548.5	6033.5
B	103.2	1668.5	3743.0	6340.0
C_7	101.6	2113.5	4043.5	6328.5
$C^{\#}$	98.0	2237.5	4214.5	6717.0
D	97.0	2361.0	4343.0	7082.5
$D^{\#}$	95.8	2521.5	4640.0	7554.5
E	91.6	2682.0	4777.5	8035.5
F	90.3	2832.5	5072.0	8492.0
$F^{\#}$	89.0	3004.0	5400.0	9011.5
G	86.2	3187.5	5609.0	9519.0
$G^{\#}$	84.4	3383.0	6106.0	10112.5
A	82.9	3584.5	6432.0	10792.0
$A^{\#}$	81.7	3801.0	6788.0	11224.0
B	79.5	4021.0	7144.0	11664.0
C_8	76.5	4241.0	7380.0	11992.0

was used to characterize the vibrational patterns of the modes of the crotales.[3] To drive the vibrations, a speaker was placed in the anechoic chamber containing the crotale. The speaker was driven by a high-quality sine-wave generator. It was demonstrated in a similar experiment that the location and orientation of the speaker in the chamber does not affect the modal structure of the vibrations of the object under investigation.[2] However, in order to drive the vibrations with the maximum possible efficiency, the speaker was oriented perpendicularly to the face of the crotales. Using this method the three acoustically important modes were identified as the (2,0), (3,0), and (4,0) modes, where the integers represent the number of diametric and circular nodes, respectively. This procedure was repeated for each crotale in the set. The frequencies of these modes for each crotale are shown in Table I. Note that the ratios of the frequencies of the (3,0) to (2,0) modes are approximately 2:1 while the ratios of the frequencies of the (4,0) to (2,0) modes are approximately 7:2 for each crotale. Typical electronic speckle pattern interferograms are presented in Fig. 3. The center mass is not visible

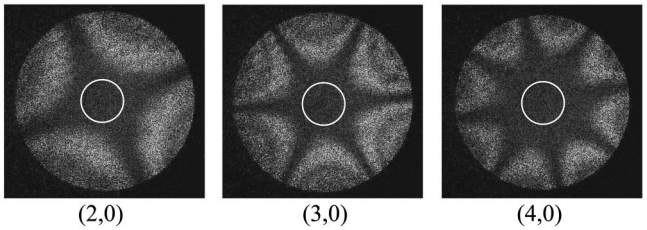


FIG. 3. Typical interferograms of the (2,0), (3,0), and (4,0) modes of a crotale. The light regions indicate places where the crotale is moving. Black regions indicate positions with little or no movement. The position of the center mass is indicated by a solid white line.

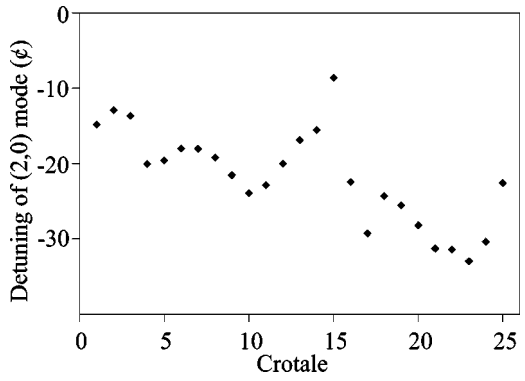


FIG. 4. Departure of the (2,0) mode from ideal tuning. If ideally tuned there would be no difference between the (2,0) mode and the frequency of the note according to the Western musical scale. The crotales correspond to numbers 1 through 25, respectively, with $C_6=1$ and $C_8=25$. Uncertainties are smaller than data points.

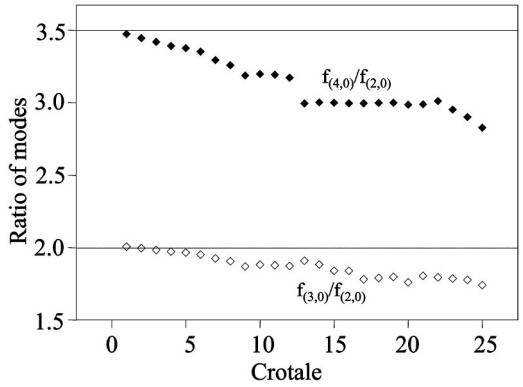


FIG. 5. Ratios of the frequencies of acoustically important modes. Open diamonds represent the ratios of the frequencies of the (3,0) to (2,0) modes. Closed diamonds represent the ratios of the frequencies of the (4,0) to (2,0) modes. The crotales correspond to numbers 1 through 25, respectively, with $C_6=1$ and $C_8=25$. Note the departure from ideal tuning of 2:1 and 7:2. Uncertainties are smaller than data points.

on the interferogram but is indicated by a solid white line in the figure.

B. Analysis of the tuning of crotales

Based on the data presented above, we define an ideally tuned crotale as one for which the (2,0) mode occurs at the frequency corresponding to the note name stamped on the crotale, and the frequency of the (3,0) mode is exactly one octave higher than that of the (2,0). Furthermore, the ratio of the frequencies of the (4,0) and (3,0) modes is 7:4, making a minor 7th. Using this definition, the tuning of each of the crotales was compared to the ideal. Figure 4 shows the detuning in cents of the (2,0) mode of each crotale from the frequency of its corresponding note on the usual chromatic scale.[4] Generally, the (2,0) mode becomes less accurately tuned as the scale is ascended. Although the ability of a person to perceive a mistuned interval varies, a good musician can discriminate a 5 cents mistuning. Hall opines that it is reasonable to insist that an organ be tuned to within 2 or 3 cents of the target pitch;[5] we see no reason for this standard not to be applied to crotales.

The ratios of the frequencies of the higher-order modes to the (2,0) mode show a similar trend as the scale is ascended. Figure 5 is a plot of the ratio of the frequencies of the (3,0) to (2,0) and (4,0) to (2,0) modes for each of the crotales, comparing each to the ideal 2:1 and 7:2 ratio. This demonstrates clearly that the crotales become increasingly detuned as the musical scale is ascended.

III. THEORY

A. Comparison to thin plate theory

Since crotales appear to be slightly modified thin plates fixed at the center it is reasonable to suspect that they can be accurately modeled using thin plate theory. Fortunately, thin plates have been studied for centuries and are well understood. Following the derivation presented by Leissa,[6] the solution to the general equation of motion for a thin circular plate is

$$z=[A_1J_n(kr)+A_2I_n(kr)+A_3Y_n(kr)+A_4K_n(kr)]\cos(n\theta)\sin(\omega t), \tag{1}$$

where n is an integer, z is the deflection of a point from the equilibrium plane of the plate, θ and r are polar coordinates, A_i is a constant, J_n and I_n are Bessel functions of the first and second kinds, respectively, and Y_n and K_n are modified Bessel functions of the first and second kinds, respectively. The constant k in the argument of the Bessel functions is defined by

$$k^4=\frac{12\rho\omega^2(1-\nu^2)}{Eh^2}. \tag{2}$$

Here, ν is Poisson's ratio, ρ is the volume mass density, E is Young's modulus, ω is the angular frequency, and h is the thickness of the plate. Terms with K_n and Y_n in Eq. (1) must be eliminated in this instance to prevent nonzero displacement at $r=0$, leaving only two unknown parameters in Eq. (1). Furthermore, for a plate fixed at the center, the $n=0$ terms are absent for the same reason. The boundary conditions at the edge of a plate of radius a that is free to vibrate are

$$M_r(a,\theta)=0 \tag{3}$$

and

$$V_r(a,\theta)=0, \tag{4}$$

where M_r is the bending moment, related to the displacement by

$$M_r(r,\theta)=-D\left[\frac{\partial^2 z}{\partial r^2}+\nu\left(\frac{1}{r}\frac{\partial z}{\partial r}+\frac{1}{r^2}\frac{\partial^2 z}{\partial\theta^2}\right)\right], \tag{5}$$

and V_r is the Kelvin–Kirchoff edge reaction, defined by

$$V_r(r,\theta)=-D\frac{\partial}{\partial r}(\nabla^2 z)+\frac{1}{r}\frac{\partial M_{r\theta}}{\partial\theta}. \tag{6}$$

Here, D is the flexural rigidity, defined as

TABLE II. The parameters used in all models to predict modal frequencies. The plate thickness (h) and density (ρ) were measured. Young's modulus (E) and Poisson's ratio (ν) are taken from Ref. 7.

Physical parameters of crotales	
E	10^{11} N/m^2
h	4.7 mm
ρ	8861 kg/m^2
ν	0.33

$$D=\frac{\rho\omega^2}{k^4 h}. \tag{7}$$

Upon applying these boundary conditions to Eq. (1), the eigenvalues determine the frequencies of the normal modes of the plates. We define a nondimensional frequency parameter as

$$\lambda = ka, \tag{8}$$

which can be used as a general solution for normal-mode frequencies, independent of the physical parameters of the plate.

In describing crotales as thin circular plates we assume that the physical parameters of every crotale are identical with the exception of the plate radius. Young's modulus and Poisson's ratio for yellow brass were taken from the literature;[7] all other physical parameters were measured. The values of the parameters used are given in Table II. Table III lists values of λ and predicted frequencies of the acoustically important modes for the C_6, C_7, and C_8 crotales, as well as the error that results by comparing them to measured values. Note that it is not only the absolute frequencies that show poor agreement, but the ratios of the frequencies of the modes also do not agree with the experimental values. Clearly, this model is insufficient to predict the normal modes of crotales.

One possible explanation for this discrepancy is that the thickness of the plates in question violates the assumption of a thin plate. The thin plate theory outlined above applies only to plates for which the thickness is much less than the plate diameter. Although this appears to be a valid approximation given the physical parameters of the crotales, it is possible that it is not. In order to confirm the validity of modeling the crotales as thin circular plates the center masses of two of the crotales ($D_6^\sharp$ and F_6) were milled off to create a thin plate. The frequencies of the normal modes were then experimentally determined and identified using the method described above and compared to predicted values for thin plates. The experimental values fall within 1.5% of the theoretical values for all but the (4,0) mode of the F_6 crotale, which deviates by approximately 2.4%. This supports the hypothesis that the crotales without the center mass may be modeled as thin plates. This also confirms that it is indeed the presence of the center mass that is responsible for the proper tuning of the crotales.

B. Comparison to annular plate theory

Since the crotales are physically clamped at the center when mounted for playing, and the interferograms shown in Fig. 3 indicate minimal movement of the central mass during play, one may suspect that crotales may be modeled as annular plates free to vibrate at the outer radius and clamped at the inner radius. To investigate this hypothesis further we model the crotale as an annular plate following the methods of Vogel and Skinner.[8] Using an approach similar to the thin plate theory described above, we begin with Eq. (1); this time, however, the terms containing K_n and Y_n are allowed, since $r=0$ is not included as a boundary condition. Vogel and Skinner define the nondimensional frequency parameter for an annular plate to be

$$\lambda' = \omega\left(\frac{4\rho a^4}{Eh^2}\right)^{1/2}; \tag{9}$$

the relationship between λ and λ' is therefore

$$\lambda' = \frac{\lambda^2}{\sqrt{3(1-\nu^2)}}. \tag{10}$$

The boundary conditions for an annular plate clamped at the inner edge and free to vibrate at the outer edge are

$$\frac{\partial^2 z}{\partial r^2}+\nu\left(\frac{1}{r}\frac{\partial z}{\partial r}+\frac{1}{r^2}\frac{\partial^2 z}{\partial\theta^2}\right)=0 \tag{11}$$

and

$$\frac{\partial}{\partial r}\left(\frac{\partial^2 z}{\partial r^2}+\frac{1}{r^2}\frac{\partial z}{\partial r}+\frac{1}{r^2}\frac{\partial^2 z}{\partial\theta^2}\right)+\frac{1-\nu}{r^2}\frac{\partial^2}{\partial\theta^2}\left(\frac{\partial z}{\partial r}-\frac{z}{r}\right)=0, \tag{12}$$

TABLE III. Predicted frequencies of the acoustically important modes for three crotales using thin plate theory. The error when compared to the actual values is also indicated.

Diameter (±0.1 mm) (Note)	Mode	λ	Predicted frequency (±0.25 Hz)	Actual frequency (±0.25 Hz)	% error	Ratio with (2,0) freq.
132.8	(2,0)	2.29	914.0	1055.5	13.40	1
(C_6)	(3,0)	3.50	2135.1	2117.5	−0.83	2.33
	(4,0)	4.65	3768.7	3849.5	−2.75	4.11
101.6	(2,0)	2.29	1561.6	2113.5	26.11	1
(C_7)	(3,0)	3.50	3647.8	4043.5	9.79	2.33
	(4,0)	4.65	6438.8	6328.5	1.74	4.11
76.5	(2,0)	2.29	2754.4	4241.0	35.05	1
(C_8)	(3,0)	3.50	6434.2	7380.0	12.82	2.33
	(4,0)	4.65	11357.1	11992.0	5.29	4.11

TABLE IV. Predicted frequencies of the acoustically important modes for three crotales using annular plate theory. The error when compared to the actual values is also indicated.

Diameter (±0.1 mm) (Note)	Mode	λ′	Predicted frequency (±0.25 Hz)	Actual frequency (±0.25 Hz)	% error	Ratio with (2,0) freq.
132.8	(2,0)	4.04	1150.7	1055.5	−9.02	1
(C_6)	(3,0)	7.64	2178.4	2117.5	−2.87	1.89
	(4,0)	13.19	3758.9	3849.5	−2.49	3.27
101.6	(2,0)	4.68	2278.6	2113.5	−7.81	1
(C_7)	(3,0)	7.92	3858.0	4043.5	4.59	1.69
	(4,0)	13.28	6466.2	6328.5	−2.18	2.84
76.5	(2,0)	6.03	5175.9	4241.0	−22.04	1
(C_8)	(3,0)	8.78	7543.6	7380.0	−2.22	1.46
	(4,0)	13.71	11773.9	11992.0	1.82	2.27

for the free outer edge, and

$$z = 0 \tag{13}$$

and

$$\frac{\partial z}{\partial r} = 0, \tag{14}$$

for the clamped inner edge. When these conditions are applied to Eq. (1), the eigenvalues can be used to find the frequencies of the normal modes of the plates. The only physical parameters required to determine the eigenvalues are the ratio of the inner to outer radii and Poisson's ratio, assumed here to be 0.33 (Vogel and Skinner conclude that the eigenvalues are not sensitive to the value of Poisson's ratio).

Using this theory, we have predicted the frequencies of the (2,0), (3,0), and (4,0) modes for a set of free-clamped annular plates. The inner radii were chosen to be equal to the radii of center masses of the crotales and the outer radii were chosen to be equal to that of the indicated crotales. The values for λ′ and predicted frequencies are shown in Table IV. Comparing these predictions to the actual frequencies of the crotales shows better agreement than was found for the model of a flat plate, indicating the superiority of the annular plate model. This is not surprising given the physical similarity between crotales and free-clamped annular plates; however, the poor agreement between the predicted ratios of the frequencies of the modes and the actual ratios of the frequencies demonstrates clearly that annular plate theory is also insufficient as a model for crotales.

Since these simple theories proved to be insufficient for predicting the behavior of crotales, we turned to finite element analysis to understand the effect of the center mass on the tuning.

C. Modeling the crotales using finite element analysis

A finite element model was developed using Solidworks, a commercially available software program. The model contained up to 127 620 nodes, with the actual number of nodes depending upon the size of the center mass relative to the plate. The software calculated the resonant frequencies of the normal modes as well as providing a visual confirmation of the mode shapes. This allowed us to compare predicted and observed mode shapes as well as the resonant frequencies.

The accuracy of the program was verified by modeling a thin flat circular plate. The predicted values matched those derived from Eq. (1) to within 0.5%. We then created a crotale model using the physical parameters of the C_6 crotale and compared the predicted resonant frequencies to those measured in the laboratory. The predicted frequency of the (2,0) mode using this model agrees to within 6% of the corresponding experimental value, and we believe that this error results from not knowing the exact values for Young's Modulus, and to a lesser extent Poisson's ratio. However, the ratios of the frequencies of the modes agree to within 0.7%.

The first physical parameter investigated within the model was the height of the center mass. A series of models of the C_6 and C_8 crotales were created with the height of the center mass ranging from a flat plate to twice the height of the actual center mass. The ratios of the acoustically important modes were then determined, and are plotted in Figs. 6 and 7. It is evident that increasing the height of the center mass has little effect on the ratios of the frequencies of the acoustically important modes of the C_6 crotale once it has reached 100% of the mass height as manufactured. However, the data in Fig. 7 indicate that any central mass detunes the smaller C_8 crotale, but with little extra effect after 100%.

The fact that increasing the height of the center mass beyond a certain point produces little or no change in the

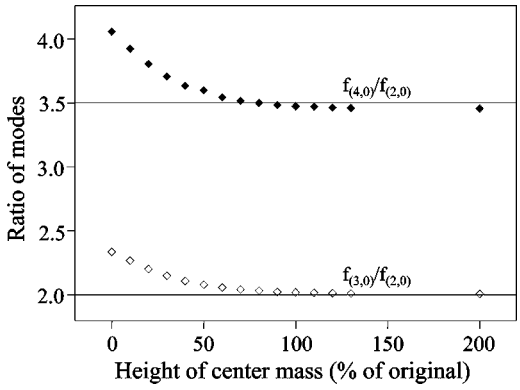


FIG. 6. Frequency ratios of the acoustically important modes of the C_6 crotale as a function of height of center mass according to the FEA model. Open diamonds represent the ratios of the frequencies of the (3,0) to (2,0) modes. Closed diamonds represent the ratios of the frequencies of the (4,0) to (2,0) modes. The ideal ratios of 2:1 and 7:2 are indicated by horizontal lines. After approximately 100%, the frequency ratios of the modes changes very little.

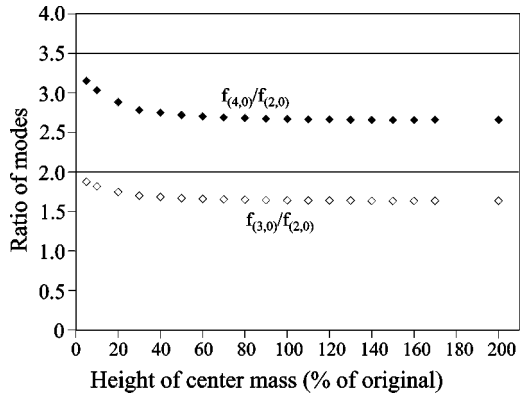


FIG. 7. Frequency ratios of the acoustically important modes of the C_8 crotale as a function of the height of the center mass according to the FEA model. Open diamonds represent the ratios of the frequencies of the (3,0) to (2,0) modes. Closed diamonds represent the ratios of the frequencies of the (4,0) to (2,0) modes. The ideal ratios of 2:1 and 7:2 are indicated by horizontal lines. Note that tuning worsens with the departure from a flat plate.

frequencies of vibration indicates that the center mass does indeed act as a clamping mechanism as was postulated above. However, clearly the boundary conditions are not equivalent to those of an annular plate clamped at the center.

The second investigation within the context of the finite element model entailed changing the radius of the center mass and determining the modal frequencies as described above. In this set of models the radius ranged from approximately 30% to 200% of the original center mass radius for the C_6 and C_8 crotales. The results of this investigation are plotted in Figs. 8 and 9. The radius of the center mass of crotales clearly has a significant effect on their tuning, as one would expect; however, these simulations indicate that the manufacturer could choose to make more ideally tuned crotales by choosing the height and radius for each crotale individually rather than a single size as they are currently manfactured. While the C_6 crotale is well-tuned as manufactured, both investigations of the C_8 crotale indicate that a flat plate of the same radius would be better tuned than the actual crotale.

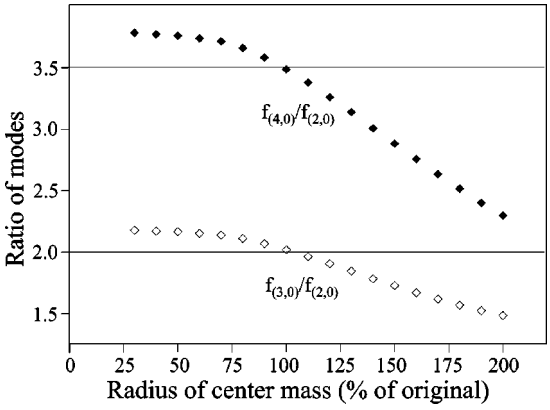


FIG. 8. Frequency ratios of the acoustically important modes of the C_6 crotale as a function of radius of center mass according to the FEA model. Open diamonds represent the ratios of the frequencies of the (3,0) to (2,0) modes. Closed diamonds represent the ratios of the frequencies of the (4,0) to (2,0) modes. The ideal ratios of 2:1 and 7:2 are indicated by horizontal lines.

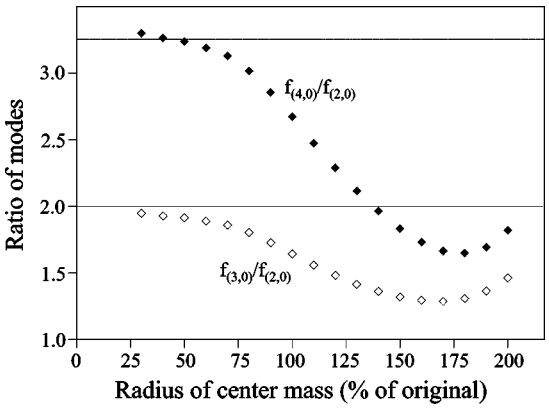


FIG. 9. Frequency ratios of the acoustically important modes of the C_8 crotale as a function of radius of center mass according to the FEA model. Open diamonds represent the ratios of the frequencies of the (3,0) to (2,0) modes. Closed diamonds represent the ratios of the frequencies of the (4,0) to (2,0) modes. The ideal ratios of 2:1 and 7:2 are indicated as horizontal lines. Again, tuning worsens with the departure from a flat plate.

D. An alternate solution for tuning crotales

While the annular plate model was shown to be insufficient for describing crotales, certain aspects of the theory deserve more careful consideration. The work of Vogel and Skinner implies that the ratio of the frequencies of the (3,0) to (2,0) mode becomes fixed when the ratio of the inner to outer radii is chosen. In most circumstances, the ratio of the frequencies of these modes corresponds to a unique value for the ratio of radii. This is true in the case of the (3,0) to (2,0) mode frequency ratio. Building on the work of Ref. 8, we have defined the parameters for a series of free-clamped annular plates that have similar acoustic properties to crotales. They have the added advantages of being more ideally tuned and containing less metal, presumably leading to lower production costs.

To meet the criteria for ideal tuning the (2,0) modes of the annular plates must occur at the frequencies corresponding to the desired note, a ratio of 2:1 must exist between the frequencies of the (3,0) and (2,0) modes, and a ratio of 7:2 must exist between the frequencies of the (4,0) and (2,0) modes. Since the (2,0) and (3,0) modes contain the most power, their nondimensional frequency parameters were chosen to optimize tuning. Using these criteria, the optimal relationship between the inner and outer radii was determined to be

$$\frac{b}{a} = 0.185, \tag{15}$$

where b is the inner radius. We note that this is a unique relationship and that it indicates a smaller ratio than exists for any of the crotales investigated if the inner radius is taken as the radius of the center mass. Once the ratio of radii is chosen, the ratio of nondimensional frequency parameters between any two modes is uniquely specified. The ratio of 0.185 corresponds to values of λ' for the acoustically important modes of

$$\lambda'_{2,0} = 3.79, \tag{16}$$

$$\lambda'_{3,0} = 7.57, \tag{17}$$

and

$$\lambda'_{4,0} = 13.18. \tag{18}$$

This ratio and set of nondimensional frequency parameters ensures that the ratio of the frequencies of the (3,0) to (2,0) modes is 2.00, and that the ratio of the (4,0) to (2,0) modes is 3.48. This is fortunate, since the latter results in an almost perfect minor seventh relationship and cannot be changed without affecting the ratio of the frequencies of the (3,0) to (2,0) modes. Note that since the nondimensional frequency parameter is related linearly to the frequency of its corresponding mode, the ratios of the frequencies of the modes are equal to the ratios of the nondimensional frequency parameters.

To finish the design all that remains is to choose the value of the outer radii of the plates such that the (2,0) modes occur at the correct frequencies. This can be accomplished by rearranging Eq. (9) to yield

$$a = \left(\frac{\lambda' h}{4\pi f} \sqrt{\frac{E}{\rho}} \right)^{1/2}, \tag{19}$$

where f is the desired frequency of the (2,0) mode.

Using Eqs. (15) and (19), the radii of a set of free-clamped annular plates with tuning similar to that of an ideal crotale can be found. Values for the inner and outer radii of such a set of plates were determined in this manner, and finite element models were used to confirm the validity of these parameters. For both octaves, the ratios of the frequencies of the acoustically important modes agrees extremely well with the ideal. The ratio of the frequencies of the (3,0) to (2,0) modes agrees to within 1%, and the ratio of the frequencies of the (4,0) to (2,0) modes agrees to within 4% for the entire set. As one ascends the scale into the upper octave, however, the frequencies of the (2,0) modes are no longer accurately predicted by Eq. (19). This is apparently due to the thickness of the plates becoming large relative to the diameter. One may counter this effect by making the plates correspondingly thinner or decreasing the outer radius while keeping the same inner to outer radius ratio of 0.185.

IV. CONCLUSION

We have investigated the dynamics of orchestral crotales both theoretically and experimentally and have determined the important physical parameters in creating the sound of the instruments. We have shown that orchestral crotales typically have three acoustically important modes, which have been identified as the (2,0), (3,0), and (4,0) modes. However, in some cases only the (2,0) and (3,0) modes were found to be acoustically important. The frequencies of these modes were reported for each crotale in a two-octave set. The crotales become increasingly detuned as the Western musical scale is ascended, which may be explained by the invariance of the magnitude and radius of the central mass.

Empirical values of the frequencies of the normal modes of the crotales were compared to predicted values derived from a model of a thin plate clamped at the center and a model of an annular plate clamped at the center. Since both of these models were shown to be inadequate for describing the behavior of crotales, a finite element model of the crotales was used to investigate the importance of the central mass. It was found that decreasing the height of the center mass increases the ratios of the modal frequencies while increasing the height has little effect. Additionally, increasing or decreasing the radius of the center mass has a large effect on the tuning of a crotale. It has been shown that the physical parameters of the center mass have been chosen well for the lowest crotales, but that the highest crotale would be better tuned if the center mass were absent entirely.

Finally, a design for a more ideally tuned instrument was presented. This instrument consists of clamped annular plates with a ratio of inner to outer radii of 0.185.

ACKNOWLEDGMENTS

This work was supported by a grant from the John R. and Ruth W. Gurtler Foundation. The authors thank E. Gottleib and N. Hall for their support and expertise.

[1] N. Fletcher and T. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer-Verlag, New York, NY, 1998).
[2] B. Deutsch, A. Robinson, R. Felce, and T. Moore, ''Nondegenerate normal-mode doublets in vibrating flat circular plates,'' Am. J. Phys. **72**, 220–225 (2004).
[3] R. Jones and C. Wykes, *Holographic and Speckle Interferometry*, 2nd ed. (Cambridge University Press, New York, NY, 1989).
[4] J. R. Pierce, *The Science of Musical Sound* (Freeman, New York, NY, 1992).
[5] D. E. Hall, *Musical Acoustics* (Brooks/Cole, Pacific Grove, CA, 2002), p. 411.
[6] A. Leissa, *Vibrations of Plates* (Acoutical Society of America, Melville, NY, 1993).
[7] http://www.matweb.com, 2004.
[8] S. M. Vogel and D. W. Skinner, ''Natural frequencies of transversely vibrating uniform annular plates,'' J. Appl. Mech. **32**, 926–931 (1965).

Vocal tract resonances in singing: The soprano voice[a)]

Elodie Joliveau, John Smith,[b)] and Joe Wolfe[c)]
School of Physics, University of New South Wales, Sydney NSW 2052, Australia

(Received 20 January 2004; revised 18 June 2004; accepted 15 July 2004)

The vocal tract resonances of trained soprano singers were measured while they sang a range of vowels softly at different pitches. The measurements were made by broad band acoustic excitation at the mouth, which allowed the resonances of the tract to be measured simultaneously with and independently from the harmonics of the voice. At low pitch, when the lowest resonance frequency $R1$ exceeded f_0, the values of the first two resonances $R1$ and $R2$ varied little with frequency and had values consistent with normal speech. At higher pitches, however, when f_0 exceeded the value of $R1$ observed at low pitch, $R1$ increased with f_0 so that $R1$ was approximately equal to f_0. $R2$ also increased over this high pitch range, probably as an incidental consequence of the tuning of $R1$. $R3$ increased slightly but systematically, across the whole pitch range measured. There was no evidence that any resonances are tuned close to harmonics of the pitch frequency except for $R1$ at high pitch. The variations in $R1$ and $R2$ at high pitch mean that vowels move, converge, and overlap their positions on the vocal plane $(R2,R1)$ to an extent that implies loss of intelligibility. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1791717]

PACS numbers: 43.75.Rs [NHF] Pages: 2434–2439

I. INTRODUCTION

During normal voiced speech, the vibrating vocal folds generate a harmonically rich signal with pitch frequency f_0, which is transmitted via the vocal tract into the surrounding air (Fant, 1973). Resonances in the tract are controlled almost independently of f_0 by varying the position of the tongue, jaw, and lips. These resonances produce broad peaks in the spectral envelope of speech. Historically, the word ''formant'' has been used to describe both a resonance of the tract and a consequent peak in the spectrum of the output sound. However, these are physically quite distinct phenomena. To avoid confusion, the term ''formant'' will be used in this paper to describe a broad peak in the sound spectral envelope and Fi will be used to describe the frequencies at which these maxima occur. The term resonance will be used to denote an acoustic resonance of the vocal tract with resonance frequency Ri. Western vowels are generally identified by the frequencies of the first two formants $(F1,F2)$ or those of their associated resonances $(R1,R2)$.

Singers trained in the Western classical tradition often need to be heard in large auditoria, sometimes with loud orchestral accompaniment. Some, especially male singers, learn to produce a strong ''singers' formant'' and, thus, without extra effort, to produce greater power in the range around 3 kHz, a range where the competition from orchestras is reduced (Sundberg, 1974). This technique would be less effective for the high soprano range because the large spacing between the vocal harmonics means that few or no harmonics may coincide with such a resonance. In the high soprano range, however, the fundamental begins to enter the range at which human hearing sensitivity is greatest, so tuning $R1$ close to f_0 is a possibility that could produce a sound whose loudness varies less with pitch, and which is louder for constant effort (Sundberg, 1975, 1977, 1987). Indeed tuning $R1$ slightly above f_0 could maintain an inertive load on the vocal folds and consequently might enhance their vibration (Titze, 1988). As vowel identifiability is inevitably compromised once f_0 exceeds $R1$, this should not (further) reduce comprehensibility greatly. Sopranos are often taught to lower the jaw, to ''smile'' or to yawn as they ascend a scale; these actions increase mouth opening, which increases $R1$. Indeed, Sundberg and Skoog (1997) measured mouth openings increasing with f_0, consistent with the resonance tuning hypothesis.

It has proved difficult, however, to determine the degree to which this resonance tuning actually occurs during singing. The problem is that it is difficult to determine reliably the resonance frequencies of the tract from the sound alone, using either spectral analysis or linear prediction, once f_0 exceeds 350 Hz (Monson and Engebretson, 1983), and essentially impossible once f_0 exceeds 500 Hz.

Consequently, several indirect methods have been employed to see whether this tuning occurs. These have included use of an external vibrator held at the throat while the singer mimed singing (Sundberg, 1975), matching of sound spectra to various source-filter models (Sundberg, 1975), measurement of lip area and jaw opening followed by an articulatory model (Sundberg, 1975; Lindblom and Sundberg, 1971), and the use of various nonperiodic phonations such as vocal fry (Miller *et al.*, 1997). However, none of these techniques are capable of measuring the resonance frequencies precisely during natural singing (e.g., see Erickson and D'Alfonso, 2002).

Recently we have developed a new technique that uses an external broadband acoustic current source to excite the vocal tract resonances independently of the voice (Epps

[a)]A Brief Communication covering the tuning of only the average data for the lowest resonance has been published in Nature (London) **427**, 116 (2004).

[b)]Author to whom correspondence should be addressed. Electronic mail: john.smith@unsw.edu.au
[c)]Electronic mail: j.wolfe@unsw.edu.au

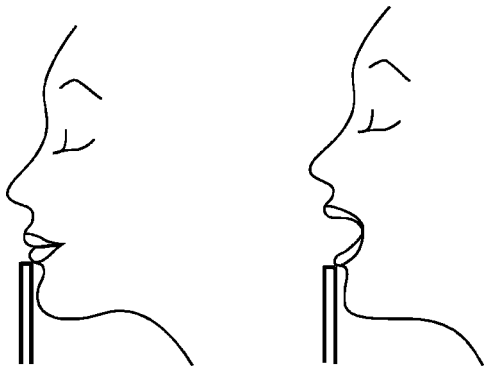


FIG. 1. The geometry (approximately to scale) in which the measurements were made showing how an acoustic current source and a microphone are placed so as to touch the singer's lower lip.

et al., 1997; Joliveau *et al.*, 2004). The present work aims to use this technique to determine the extent to which soprano singers tune their several vocal tract resonances to match harmonics of the sung pitch during normal singing.

II. MATERIALS AND METHODS

A. Measurements of vocal tract resonances

The resonances of the tract are measured directly, during sustained sung notes, using a technique described previously (Dowd *et al.*, 1997; Epps *et al.*, 1997). Briefly, a microphone and a small source of acoustic current, side by side on a flexible mounting, are placed just below the singer's mouth so that they gently touch the lower lip throughout the experiment—see Fig. 1. A computer (Mac IIci with analog interface card—National Instruments NB-A2100) synthesises the broad band signal from frequencies spaced at 5.38 Hz over the range of interest, here 0.2–4.5 kHz. The microphone is used to measure simultaneously the harmonics in the voice signal and the acoustic pressure produced when the broadband acoustic current interacts with the vocal tract. This current acts on the parallel combination $\mathbf{Z}_{\|}$ of the acoustical impedance of the vocal tract $\mathbf{Z}_t$ and that of the radiation field $\mathbf{Z}_r$ of the surrounding air. $\mathbf{Z}_r$ is inertive: it is a largely imaginary impedance whose value is low but which increases with frequency. It is measured for each singer by conducting a calibration experiment with her mouth closed, during which the acoustic current is adjusted so that the measured pressure spectrum p_{closed} is independent of frequency (see Fig. 2). Plots of the magnitude of the ratio $\gamma \equiv \mathbf{Z}_{\|}/\mathbf{Z}_r$ show peaks at the resonance frequencies of the tract (Dowd *et al.*, 1997). The acoustic current source has an output impedance much higher than the impedance of the load, so a plot of the magnitude of the ratio of the pressure measured during singing (p_{open}) to that measured during calibration (p_{closed}) shows peaks in the broad band signal corresponding to the vocal tract resonances, whose frequencies can be measured with a precision of order ± 11 Hz—see Fig. 2. It also shows the harmonics of the voice. Measurements with simple models of the vocal tract show that the resonance frequency measured in this fashion corresponds well with the resonances in the transpedance of the vocal tract (unpublished).

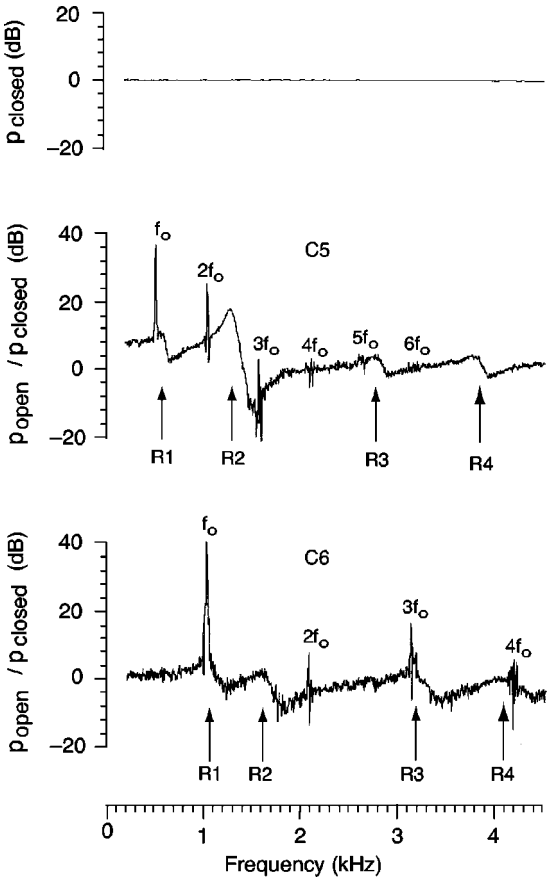


FIG. 2. The upper figure presents the pressure spectrum measured for a soprano without singing and with the mouth closed (p_{closed}). The spectrometer had been previously calibrated in this configuration by adjusting the acoustic current so that the pressure spectrum measured with the mouth closed was independent of frequency with nominal value of 0 dB. The lower two figures present the ratio of the pressure spectrum measured with the mouth open to that measured with the mouth closed ($p_{\text{open}}/p_{\text{closed}}$) when the subject sang the vowel /u/ (in *who'd*) on the notes C5 and C6. The harmonics of the (periodic) voice signal are indicated. The peaks in the broad band signal indicated by arrows correspond to the resonances $R1$, $R2$, $R3$, and $R4$.

The presence of the current probe and microphone (total cross-sectional area approximately 120 mm^2) immediately below the lip slightly reduces the area for radiation and consequently the measured resonance frequency is expected to underestimate slightly the correct value. The magnitude of this difference was estimated by measuring the shift in resonance frequency that occurred when an additional current probe plus microphone was placed next to the first pair using a geometrically simplified, rigid model of a face and vocal tract. It was thus found that the error caused by the presence of the current probe lay within the resolution of these experiments (± 11 Hz).

B. The subjects

Nine sopranos participated (five were professionals with an average of 12 years classical training, four were students with an average of 7 years classical training). They all described their singing style as ''Western classical'' and sang predominantly in opera and choirs. All were born in Austra-

lia or had lived in Australia for at least 10 years and were judged by the investigators to speak educated, metropolitan, Australian English.

C. The experiments

Four vowels were chosen (ɑ,ɔ,u,ɜ) to ensure ease of singing and measurement, sampling of the phoneme space, and the effects of lip rounding. The word to be sung was presented in writing and had the form $h\langle\text{vowel}\rangle d$ (ɑ—*hard*; ɔ—*hoard*; u—*who'd*; ɜ—*heard*). Each subject then sang a note without vibrato that was sustained for 4 s. These notes comprised an ascending diatonic scale that covered their entire comfortable range for each vowel. The target pitch was provided by a glockenspiel before each note. They were asked to sing ''softly, but in their trained singing style.'' They were asked to sing softly for the following reason. Singers can produce very high sound pressure levels immediately outside the mouth. The technique used requires a good signal-to-noise ratio and this requires that the sound pressure level produced by the injected broadband signal, at each injected frequency, should be comparable with that produced by the singing. As the broad band signal has many frequency components, the sound pressure level may become high enough to cause distortion in the microphones used. Although the distortion is small, the technique is rather sensitive to it. Subjects were asked to sing without vibrato because this causes a smearing of the harmonics of the voice signal that can obscure the tract resonances. No subjects complained or appeared to be worried by the presence of the broadband signal nor the request to sing *piano* and *senza vibrato*. Measurements were made in a quiet room with low reverberance. Reproducibility was measured by asking each singer to sing the vowel /ɑ/ in *hard* at pitch A4 before and after each scale. The standard deviation of the resonance frequencies of these reproducibility measurements, averaged across all singers, was thus found to be ± 25 Hz ($R1$), ± 60 Hz ($R2$), ± 60 Hz ($R3$), and ± 90 Hz ($R4$). The reproducibility for an individual singer was better than the variation among different singers.

III. RESULTS AND DISCUSSION

Figure 2 provides two examples of the measured ratio $p_{\text{open}}/p_{\text{closed}}$. One of the subjects was asked to sing the vowel /u/ (in *who'd*) at pitches C5 (523 Hz—near the middle of her range) and C6 (1046 Hz—near the top of her range). For C5, the first six harmonics of the voice are visible. It is apparent that, even at this modest pitch, it is difficult to estimate the tract resonances from the voice signal, simply because the spacing between harmonics (here 523 Hz) is too great. However, the resonances of the vocal tract are clearly seen in the broad band signal. $R1$ for this vowel and singer was about 420 Hz at low pitches around C4 (261 Hz). When this vowel was sung at C5, $R1$ has increased to a value slightly higher than the f_0 for C5. When the subject was asked to sing the same vowel at C6, her values of $R1$ to $R4$ have all increased. $R1$ and f_0 now coincide to within ~ 20 Hz, thus suggesting that it is possible for them to match quite closely. Again it is apparent that the resonances of the tract could not be determined from the voice signal at this pitch. From curves like those in Fig. 2, $R1-R5$ were determined for each singer for the four vowels over their comfortable pitch range. However, $R5$ was not always apparent in the frequency range studied, and $R1$ was not strong enough for reliable measurement on one singer.

Figure 3 shows the values for the vocal tract resonances for each vowel and pitch frequency f_0, averaged across all singers. The dashed lines represent $R=nf_0$, where $n=1,2,...,6$. When a resonance coincides with the nth of these lines, the resonance is in tune with the nth harmonic of the voice. Consistent coincidence of the average resonance with one of the harmonics would suggest matching or tuning to that harmonic.

If there are several possible harmonics nearby, then it is insufficient to compare the resonance averaged over all singers with a single harmonic. This is because some singers might be tuning their resonance to the nth harmonic and other singers to the $(n+1)$th harmonic, and consequently the average would not coincide with a harmonic. This problem can be overcome by examining $|\Delta f|$, the absolute difference between a resonance and the nearest harmonic. Once $R1$ exceeds f_0, a resonance will always lie within $\pm f_0/2$ of a harmonic. If the resonance frequencies are distributed randomly with frequency rather than tuned, then the average $|\Delta f|$ would be expected to be close to $f_0/4$. On the other hand, where resonance tuning occurs, the average $|\Delta f|$ would be expected to be small with negligible dependence on f_0.

A. The tuning of *R*1

For each vowel and over the lower pitch range, f_0 is less than the value of $R1$ and the resonances $R1$ and $R2$ for each vowel are held approximately constant, independent of pitch (Fig. 3). This is the result that one would expect for speech, because ($R1$,$R2$) characterizes vowels. However, once f_0 exceeds this value of $R1$, the value of $R1$ for the individual singers, and also for the average over all singers, increases with increasing f_0. This trend continues to 1 kHz for the vowels that do not use lip rounding (/ɑ/ in *hard* and /ɜ/ in *heard*), but for the vowels that use lip rounding (/ɔ/ in *hoard* and /u/ in *who'd*), the data fall below the tuning line near 1 kHz. This may be because, with the lips rounded, it is uncomfortable or anatomically impossible to raise $R1$ to 1 kHz.

The tuning line $R1=f_0$ lay within the standard deviations of the data in the approximate range 600 to 1000 Hz. There was variation between the singers; some consistently tuned $R1$ to a frequency just above f_0 whereas others displayed no obvious pattern. The average value of $R1$ was, however, consistently slightly higher than f_0 and this supports the idea that vocal fold vibration is enhanced when the vocal tract presents an inertive load (Titze, 1988). There was no systematic difference between professionals and students—however, the lack of precise tuning in some subjects might be a partial consequence of singing softly.

It is possible that singers might tune $R1$ to match the second harmonic of f_0 at lower pitches. However, examination of $|\Delta f|$ over the range pitch A3 (220 Hz) to A4 (440 Hz)

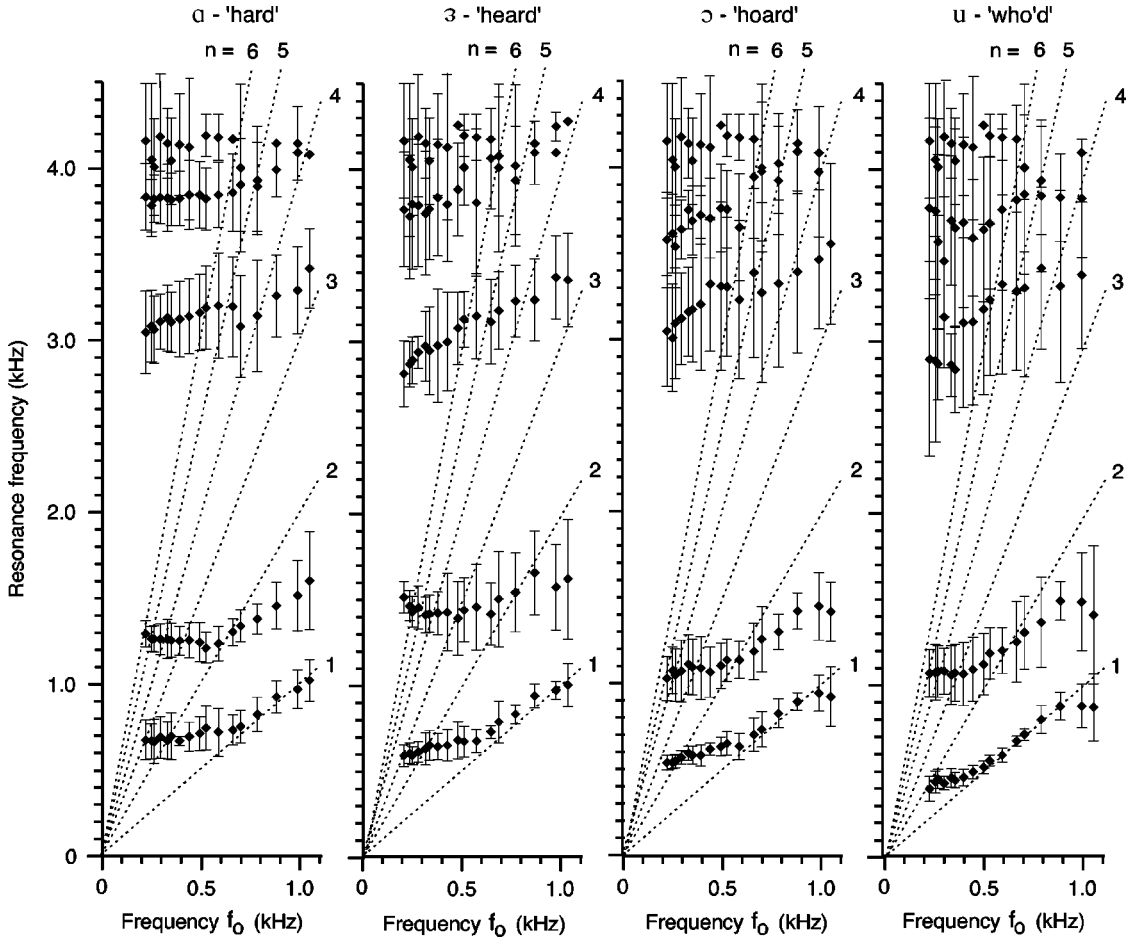


FIG. 3. The average vocal tract resonances for the four vowels studied, as a function of f_0. The vertical bars indicate the standard deviations. The dotted lines indicate the relationship $R = nf_0$, where n is the integer indicated.

found no evidence of systematic tuning for either the professionals or the students.

B. The tuning of R2

Once f_0 exceeds the normal value of $R1$, there is a systematic increase of $R2$ with f_0, but it is proportionally smaller than that measured for $R1$. This may be simply the result of the fact that, while $R2$ depends primarily on tongue configuration, it also depends somewhat on mouth opening (Fant, 1973): consequently the tuning of $R1$ can also vary $R2$. Figure 4 indicates that the overall trend is an increase in $|\Delta f|$ with f_0: as the separation of harmonics (f_0) increases the resonance tends to lie further from the nearest harmonic, as would be expected in the absence of tuning. The local

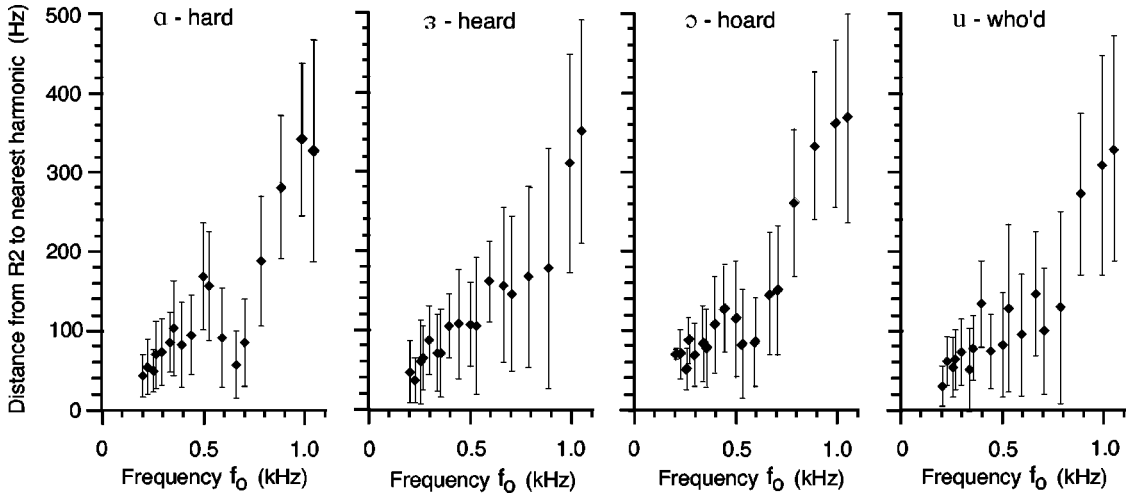


FIG. 4. The average value of $|\Delta f|$, the *absolute* value of the difference between $R2$ and the closest harmonic of f_0, plotted vs f_0 for each vowel. The vertical bars indicate the standard deviations.

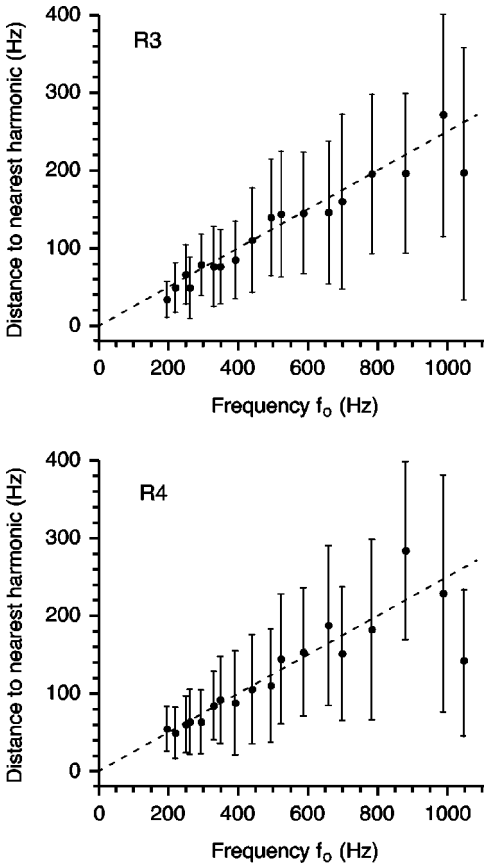


FIG. 5. The average value of $|\Delta f|$, the *absolute* value of the difference between $R3$ or $R4$ and the closest harmonic of f_0, plotted vs. f_0 and averaged over all vowels. The vertical bars indicate the standard deviations. The relationship $|\Delta f| = f_0/4$ that would be expected in the absence of tuning is indicated by a dashed line.

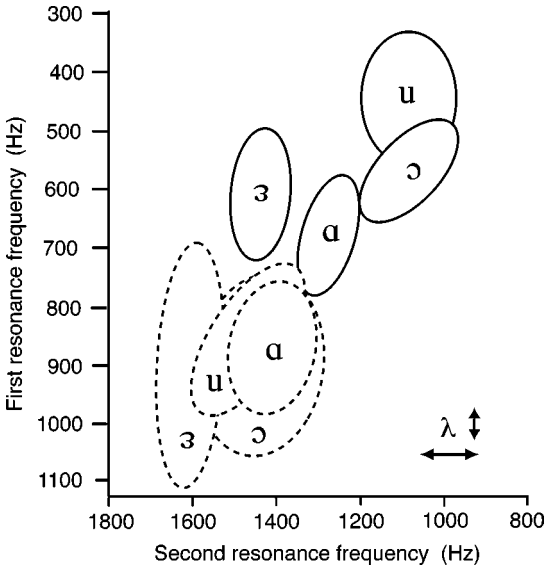


FIG. 6. Changes in position on the vowel plane as pitch is varied. The center of each ellipse gives the mean values of $R1$ and $R2$, and the slope of the major axis is the correlation between $R2$ and $R1$ for that vowel. The lengths of the semi-axes indicate the standard deviations calculated in those directions. The ellipses with solid borders are for the pitch range A3 to A4. At these low pitches, the vowels are well separated and their relative positions lie within the range typical for Australian English. The ellipses with dashed borders are for the pitch range G5 to B5. λ is a measure of the confusion length, the characteristic separation in the vowel plane at which vowels become confused.

minima around 700 Hz for ɑ and 600 Hz for ɔ correspond to the frequency at which the line $R = 2f_0$ crosses the curve $R2(f_0)$.

C. The tuning of *R*3–*R*5

The proportional variation in $R3$, $R4$, and $R5$ with frequency is not very strong (see Fig. 3), and showed no evidence of resonance tuning to match harmonics of f_0. Because the results were similar for all vowels studied, Fig. 5 shows the average $|\Delta f|$ as a function of f_0 for $R3$ and $R4$. In these data, the lines of best fit indicate that $|\Delta f| = 0.24 f_0$, $0.23 f_0$, and $0.20 f_0$ for $R3$, $R4$, and $R5$, respectively. This suggests that the sopranos in this study did not significantly tune their higher resonances when singing softly.

Despite the absence of tuning in the higher resonances, there is a small but systematic increase in $R3$ and $R4$ for all vowels. $R3$ increased with increasing f_0 for all singers and vowels, with an average value of $R3$ on $f_0 = 0.48 \pm 0.39$. The variation of $R4$ with f_0 was more varied, and in some measurements $R4$ even decreased slightly with increasing f_0. However, the average regression of $R4$ on f_0 was similar and equal to 0.46 ± 0.38. These increases are observed across the whole range studied, and not merely across the range over which $R1$ is tuned. Perhaps these variations are related to changes in the vicinity of the larynx that are related to producing high f_0.

The singers' formant refers to a prominent peak in the sound spectrum around 3 kHz (Sundberg, 2001; Weiss *et al.*, 2001). This corresponds to the average measured value of $R3$. We have not performed a study of speech for sopranos and so cannot comment on whether $R3$, $R4$, and $R5$ are clustered differently during singing. Nor can we determine whether this peak is associated with a trained voice.

It must be remembered that the subjects in this study, although instructed to sing in their trained Western classical singing style, were singing softly without vibrato. The possibility thus exists that sopranos singing at full volume will indulge in additional and more precise control of resonances to increase the sound level produced. Such effects might be more important for singers at the very top of their profession.

D. Consequences for vowel intelligibility

Vowels are commonly represented on the $(F2,F1)$ or $(R2,R1)$ plane, where, in accordance with phonetic traditions, the axes are reversed. Figure 6 shows how tuning $R1$ close to f_0 moves the vowels studied in this plane. In the low range of f_0, the resonances of the vowels are well separated and their relative positions are consistent with those for normal speech (Donaldson *et al.*, 2003). In the high pitch range, all are shifted in the direction of increasing $R1$ and $R2$ and their separation is reduced, especially in the $R1$ direction. At high pitch, the vowels converge and overlap. Their separations become comparable with the confusion length λ, the characteristic separation in the vowel plane at which vowels become confused. [To date, this distance has only been mea-

sured for speech and for French vowels (Dowd *et al.*, 1997), whence the λ shown here. It is also likely that our perceptual categorization is modified when listening at high pitch.] This movement and convergence of vowels may contribute to the well-known difficulty in identifying vowels sung in the high range by sopranos (Berlioz, 1844; Scotto di Carlo and Germain, 1985; Benolken and Swanson, 1990) and might even be one of the reasons why opera houses often use surtitles even for operas in the native language of their audiences. Difficulty in comprehending the vowels of the high soprano voice cannot, however, be blamed on resonance tuning alone: once the spacing of harmonics approaches 1 kHz, it is virtually impossible for a human ear (or signal analysis) to determine features of the spectral envelope with the precision required to resolve vowels in the ($F1$,$F2$) plane. This impossibility or near impossibility is appreciated by listening to recordings of ''different'' vowels at high pitch and, for this purpose, sound files have been placed on the web at http://www.phys.unsw.edu.au/~jw/soprane.html.

Because the price of (further) reduced intelligibility is not great, it is not surprising that sopranos elect to tune $R1$ close to f_0, thereby obtaining greater radiated power for a given effort, and also perhaps avoiding the effects on vocal timbre of having a fundamental whose amplitude varied strongly from note to note or vowel to vowel.

IV. CONCLUSIONS

A technique is demonstrated that can precisely measure the vocal tract resonances of sopranos during singing. The results presented are, however, subject to the constraint that the singers should sing softly without vibrato. In the lower part of their range, the vocal tract resonances of sopranos trained in the Western classical tradition have frequencies that vary little with pitch. However, once f_0 exceeds this normal value of $R1$, $R1$ is tuned so that $R1$ is close to f_0, except at very high pitches for vowels that involve lip rounding. Over the pitch range in which $R1$ is tuned, $R2$ increases a little, probably as an incidental consequence of the $R1$ tuning. $R3$ and $R4$ increase systematically, but not strongly over the whole soprano range. There is no evidence for the tuning of $R3$, $R4$, or $R5$ to harmonics of f_0.

ACKNOWLEDGMENTS

We thank our volunteer subjects and the Australian Research Council for support. We would also like to thank Professor Ron Scherer for his helpful comments.

Benolken, M., and Swanson, C. (**1990**). ''The effect of pitch-related changes on the perception of sung vowels,'' J. Acoust. Soc. Am. **87**, 1781–1785.

Berlioz, H. (**1844**). *Grand traité d'instrumentation et d'orchestration modernes*, translated by M. C. Clarke (Novello, London, 1882).

Donaldson, T., Wang, D., Smith, J., and Wolfe, J. (**2003**). ''Vocal tract resonances: a preliminary study of sex differences for young Australians,'' Acoust. Austral. **31**, 95–98.

Dowd, A., Smith, J. R., and Wolfe, J. (**1997**). ''Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time,'' Lang. Speech **41**, 1–20.

Erickson, M. L., and D'Alfonso, A. E. (**2002**). ''A comparison of two methods of formant frequency estimation for high-pitched voices,'' J. Voice **16**, 147–171.

Epps, J., Smith, J. R., and Wolfe, J. (**1997**). ''A novel instrument to measure acoustic resonances of the vocal tract during speech,'' Meas. Sci. Technol. **8**, 1112–1121.

Fant, G. (**1973**). *Speech Sounds and Features* (MIT, Cambridge, MA).

Joliveau, E., Smith, J., and Wolfe, J. (**2004**). ''Tuning of vocal tract resonance by sopranos,'' Nature (London) **427**, 116.

Lindblom, B. E. F., and Sundberg, J. E. F. (**1971**). ''Acoustical consequences of lip, tongue, jaw, and larynx movement,'' J. Acoust. Soc. Am. **50**, 1166–1179.

Miller, D. G., Sulter, A. M., Schutte, H. K., and Wolf, R. F. (**1997**). ''Comparison of vocal tract formants in singing and non-periodic phonation,'' J. Voice **11**, 1–11.

Monsen, R. B., and Engebretson, A. M. (**1983**). ''The accuracy of formant frequency measurements: a comparison of spectrographic analysis and linear prediction,'' J. Speech Hear. Res. **26**, 89–97.

Scotto di Carlo, N., and Germain, A. (**1985**). ''A perceptual study of the influence of pitch on the intelligibility of sung vowels,'' Phonetica **42**, 188–197.

Sundberg, J. (**1974**). ''Articulatory interpretation of the 'singing formant,''' J. Acoust. Soc. Am. **55**, 838–844.

Sundberg, J. (**1975**). ''Formant technique in a professional female singer,'' Acustica **32**, 89–96.

Sundberg, J. (**1977**). ''The acoustics of the singing voice,'' Sci. Am. **March**, 82–91.

Sundberg, J. (**1987**). *The Science of the Singing Voice* (Northern Illinois U.P., De Kalb, IL, 1987).

Sundberg, J. (**2001**). ''Level and centre frequency of the singer's formant,'' J. Voice **15**, 176–186.

Sundberg, J., and Skoog, J. (**1997**). ''Dependence of jaw opening on pitch and vowel in singers,'' J. Voice **11**, 301–306.

Titze, I. R. (**1988**). ''The physics of small-amplitude oscillations of the vocal folds,'' J. Acoust. Soc. Am. **83**, 1536–1552.

Weiss, R., Brown, W. S., and Morris, J. (**2001**). ''Singer's formant in sopranos: fact or fiction?'' J. Voice **15**, 457–468.

The effect of surface agitation on ultrasound-mediated gene transfer *in vitro*

Wen-Shiang Chen
Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708 and Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital and National Taiwan University College of Medicine, Taipei, Taiwan 100

Xiaochun Lu
Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708 and Department of Biochemistry and Molecular Biology, Peking University Medical Center, Beijing, China 100083

Yunbo Liu and Pei Zhong[a)]
Department of Mechanical Engineering and Materials Science, Duke University, Durham, North Carolina 27708

(Received 9 March 2004; revised 27 May 2004; accepted 7 June 2004)

This article reports the effect of surface agitation of culture medium on ultrasound-mediated gene transfection *in vitro* and its possible mechanisms. The possibility of active induction of bubbles without using contrast microbubbles for effective gene transfer was also demonstrated. Cultured HeLa cells mixed with green fluorescent protein plasmid were exposed to 1.0 MHz ultrasound in 24-well culture plates. Up to 26% transfection efficiency in the survival cell population was achieved in samples exposed to 0.44 MPa ultrasound pulses with the presence of surface agitation. Inertial cavitation and bubble generation were observed throughout the ultrasound exposure. When surface agitation was suppressed by covering the medium surface with a thin membrane, bubble generation and gene transfection were significantly suppressed. Interestingly, transfection efficiency could be partially resumed by adding a small amount of culture medium onto the covering membrane to rebuild the surface agitation and bubble generation. Pressure fluctuation and transient high-pressure loci were found in samples with surface agitation. Numerical simulations of bubble dynamics showed that transient high pressures above the inertial cavitation threshold could generate bubbles, which might be subsequently stabilized at lower pressures by rectified diffusion, and exert strong shear forces that might create transient pores on cell membranes to facilitate gene transfer. [DOI: 10.1121/1.1777855]

PACS numbers: 43.80.Cs, 43.80.Gx, 43.25.Yw [FD] Pages: 2440–2450

I. INTRODUCTION

Gene therapy holds great promises for the treatment of a variety of inherited and acquired diseases, such as severe combined immunodeficiency (SCID), atherosclerosis, and certain types of malignancy.[1–3] For example, SCID, which disrupts the development of the immune system, can be cured by gene therapy.[4] However, despite of its great potential, gene therapy also has several drawbacks, including insertional mutagenesis and adverse systemic immunogenicity.[5–9] Recent news that a child participating in a SCID gene therapy trial using retrovirus vector had developed leukemia forced the US National Institutes of Health (NIH) to recommend that gene therapy should be used only as a last resort.[7–9] The use of viral vectors to deliver genetic materials in patients is believed to be responsible for some of the major side effects of gene therapy. Therefore, there is a growing interest to develop new gene transfer technologies based on nonviral vectors, which are less immunogenic than viral-based delivery systems.[10]

Currently, lipofection and electroporation are the most frequently used nonviral methods for gene delivery. However, similar to viral vectors, gene delivery via lipofection often suffers from lack of site specificity. Electroporation has been widely used in *in vitro* and *ex vivo* gene transfer for years. Yet, because of its invasive nature, electroporation is difficult to apply to deep-seated tissue or organs. In contrast, the use of ultrasound for gene delivery has some unique advantages because of its noninvasiveness and ability to be focused on internal organs. Furthermore, ultrasound energy at the output level of clinical diagnostic and physical therapy systems has been shown to be effective for *in vitro* and *in vivo* gene transfection.[11–13] However, similar to other nonviral based gene transfer methods, the most critical disadvantage of ultrasound-mediated gene transfer at present is the low transfection efficiency. To combat this problem, ultrasound contrast agent (UCA) has often been added in the transfection medium to enhance ultrasound-induced cavitation activity and resultant gene transfection both *in vitro*[14–18] and *in vivo*.[11,12]

For example, Greenleaf and associates reported that the combination of ultrasound with UCA increased the transfection efficiency of immortalized chondrocytes *in vitro* from

[a)]Author to whom correspondence should be addressed; electronic mail: pzhong@duke.edu

2.4% to 50% in the surviving cells following exposure to 1 MHz continuous wave (cw) at 0.41 MPa peak pressure for 20 s.[14,15] Approximately 300-fold increments in reporter gene expression were found in cultured vascular smooth muscle cells in the presence of contrast microbubbles.[18] *In vivo* study also demonstrated that the use of UCA could dramatically increase the transfection efficiency of ultrasound-mediated gene transfer in skeletal muscle and myocardium.[11,12]

Cavitation, the inception and oscillation of gas cavities, is believed to be a primary mechanism for ultrasound-induced bioeffects, such as gene transfection, hemolysis, and thrombolysis. With the advent of gas-based UCA, it has been recognized that the potential for ultrasound-induced cavitation activity and resultant bioeffects will increase significantly (see AIUM 2000 for a comprehensive review).[19] However, the use of ultrasound contrast agent to facilitate gene transfection has several drawbacks. First of all, high dose of contrast agent, which is typical for current *in vitro* study, may not be realistic for *in vivo* applications. Second, strong inertial cavitation induced by contrast microbubbles may lead to more cell death and tissue injury than gene transfection. Furthermore, contrast agent is expensive, and is not readily available everywhere. Therefore, we are interested in developing effective methods of gene transfection based on *in situ* induction of cavitation bubbles without the use of ultrasound contrast agents. Our hypothesis is that gene transfection can be effectively induced by cavitation originated from *in situ* gas nuclei, instead of from artificially introduced contrast agent microbubbles, if adequate acoustic environment is created. We are also interested in the physical mechanisms of bubble induction that facilitates gene transfection. In this study, we demonstrated the effect of surface agitation on creating transient high-pressure loci and resultant *in situ* bubble generation that is essential for ensuring the efficiency and effectiveness of ultrasound-mediated gene transfection *in vitro*.

II. THEORY

Similar models of bubble dynamics and implications to ultrasound-induced bioeffects have been described by various authors listed in the following sections. The principles of our simulation are summarized later. Main equations are listed in the Appendix.

A. Nonlinear oscillation and inertial collapse

The spherical oscillation of a free air bubble in an acoustic field can be modeled using the Rayleigh–Plesset equation.[20–24] The general form used in our simulation is shown in Appendix A. For a shelled UCA microbubble, the viscous and elastic forces caused by motion and tension in the shell should be considered in the simulation model. An equation [Eq. (A3)] based on the work of de Jong, Church, and Hoff was used for modeling the oscillation of a shelled microbubble in various acoustic fields.[24,25]

The threshold for inertial cavitation (IC) (also known previously as transient cavitation) was determined by the supersonic collapse criterion (i.e., the collapse speed of the bubble wall is greater than the sound speed in water).[26] The model calculations were carried out using programs written in MatLab (Math Works, Natick, MA). By selecting an initial radius (R_0) and an incident pressure level (P_a), the variations of bubble radius (R) and the bubble wall velocity ($\dot{R}$) over time (t) to 40 cycles of incident acoustic wave were calculated.

B. Stable cavitation, microstreaming, and shear stresses

The microstreaming established around a pulsating bubble near a fixed boundary (platted cells in our case) was first modeled by Nyborg in 1958.[27,28] The existence of a boundary resulted in velocity gradients and corresponding hydrodynamic shear stresses (S) which may generate reparable sonoporation or lethal cell damage.[29] The detailed equations are listed in Appendix B.

C. Rectified diffusion

Small gas bubbles undergoing stable cavitation may grow (or shrink) slowly in the presence of an acoustic field.[30,31] In the frequency range of medical ultrasound (1–20 MHz), this process is called rectified diffusion, which may take thousands of cycles or several seconds before the pulsating bubble can be stabilized. The active pumping of dissolved gas from the medium into the bubbles competes with the passive diffusion of the gas out of the bubble into the surrounding medium by either an ''area'' or a ''shell'' effect.[28] When the acoustic pressure amplitude becomes greater than the rectified diffusion threshold, the pulsating bubbles will not dissolve in the presence of the acoustic field.

The process of rectified diffusion can be modeled mathematically.[30,31] The main equations that describe the rate of change of the equilibrium bubble radius (R_0) with time during rectified diffusion are summarized in Appendix C.

D. Static diffusion

In the experiments described later, free air bubbles were subjected to a train of pulsed acoustic waves. During the long ''on'' period (e.g., 2500 cycles in the standard condition), the pulsated bubbles might grow by rectified diffusion or shrank by dissolution, depending on their initial size. When the acoustic wave was temporarily stopped during the ''off'' period (e.g., 5 ms), the unforced bubbles gradually dissolved, obeying the equations of the free bubble dissolution. For a free air bubble in water, the dissolution rate depends on its initial size R_0 and the gas concentration in the solution. The simulation was based on a classic equation described by Epstein and Plesset,[21,32] which is summarized in Appendix D.

III. EXPREIMENTAL MATERIALS AND METHODS

A. Ultrasound apparatus

All experiments were performed in an acrylic tank (14 cm L$\times$15 cm W$\times$28 cm H) containing de-ionized and degassed water (O_2 concentration <4 ppm). The temperature was kept constant at 37 °C during each experiment. A 1 MHz flat ultrasound transducer (3.8 cm in diameter, A392S,

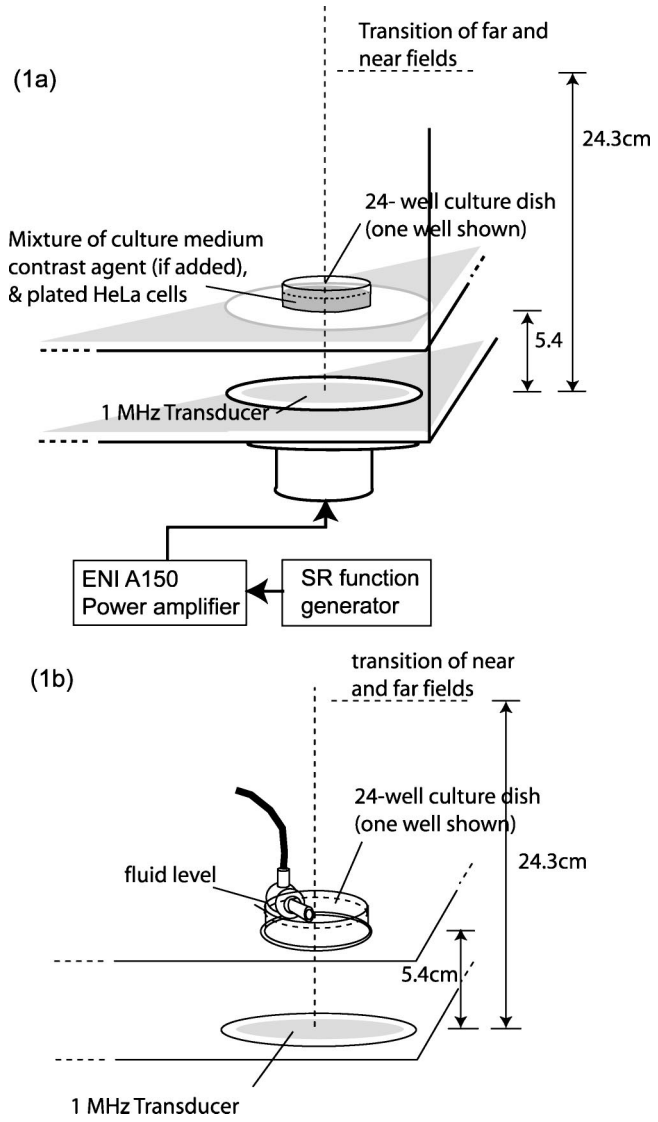


FIG. 1. (a) Experimental setup: A sample well of a 24-well plate containing cultured HeLa cells was placed in the near field of a flat 1 MHz ultrasound transducer; (b) the setup of the ACD system: A 15 MHz ultrasound transducer was placed through a hole on the lateral wall of a culture well with the tip of the transducer protruding into the well.

Panametrics, Waltham, MA) was mounted on the bottom of the tank. A 24-well plate was placed on a stand 5.4 cm above the transducer surface, while the center of the plate's first or 24th well was aligned to the center of the transmitting transducer. The diameter of each well is 16 mm. The well containing 1 ml culture medium (5 mm in depth) was thus situated in the near field of the transducer in order to have relatively even distribution of the acoustic pressure across the whole well [Fig. 1(a)].

Preselected pulse length, pulse repetition frequency (PRF), and amplitude of sinusoidal waveforms were supplied by a function generator (DS345, Stanford Research Systems, Sunnyvale, CA). The typical values were 2500 cycles in pulse length, 200 Hz PRF (thus 50% duty cycle), and 0.3 V_{pp} output voltage. The output signal from the function generator was amplified by a power amplifier (A150, ENI, Rochester, NY) at a fixed gain of 55 dB before driving the transmitted transducer. The pressures along the central axis was first calibrated by a needle hydrophone (NTS Systems, Seattle, WA) in a water tank using short pulses (20 cycles). The peak-negative pressure 5.4 cm away form the transducer surface was 0.44 MPa. The actual pressure during experiments was measured in the presence of a 24-well plate by inserting the needle hydrophone vertically from above into the culture medium and keeping at 0.5 mm above the well bottom (see Fig. 5). It should be noted that standing waves might exist in the experiment system since the wavelength and the depth of the culture medium are comparable.

The contrast agent used in this study was Definity® (Bristol-Myers Squibb Company, North Billerica, MA). Definity microbubbles are composed of perfluoropropane gas (C_3F_8) and a phospholipids shell. Based on manufacture's data, the mean diameter range is between 1.1 and 3.3 μm with 98% of Definity microbubbles less than 10 μm. After activating the contents of the Definity vial using an agitator (Vialmix™), the milky white suspension contains a maximum of 1.2×10^{10} bubbles/ml (the number was provided by the manufacturer, but the actual concentration may be lower). The concentration used in this study was 10 μl/ml.

B. Sample preparation and bioassays

1. Plasmid preparation

The green-fluorescent protein (GFP) reporter gene was used in our studies mainly because of its high sensitivity and convenience in quantification of gene expression. The GFP plasmid deoxyribose nucleic acid (DNA) was purchased from Clontech (Palo Alto, CA) and amplified by E. coli cultures. The final products of the GFP plasmids in bacteria were purified by the Qiagen Giga kit (Qiagen Inc., Chatsworth, CA). Agarose gel electrophoresis was performed before and after restriction endonuclease digestion to verify the identity and purity of the extracted plasmid DNA. A concentration 10 μg/ml of GFP plasmid was added to the HeLa cell culture medium before ultrasound exposure.

2. Cell culture

Cervical cancer cells (HeLa) were routinely cultured as a monolayer in culture flasks covered with minimal essential medium (MEM, GIBCO, Grand Island, NY) supplemented with 10% fetal bovine serum (GIBCO), sodium pyruvate, nonessential amino acids, penicillin, and streptomycin in a 37 °C humidified incubator (5% CO_2). One day before the ultrasound exposure, the cultured cells were trypsinized and seeded at a concentration of 5×10^4 cells in the corner wells of 24-well plates (Becton Dickinson Company, Franklin Lakes, NJ). The cells were then incubated for 24 h until reaching approximately 50% confluence. Right before ultrasound exposure, the cultured cells were rinsed twice with phosphate buffered saline and immersed in 0.5 ml serum-free MEM solution (without additives) mixed with GFP plasmid. Following ultrasound treatment, some cells were dislodged and some lysed. 4 h after the ultrasound exposure, the dislodged cells regained their attachment. The dead or lysed cells were carefully washed away by changing the culture medium three times, while the viable cells remained attached

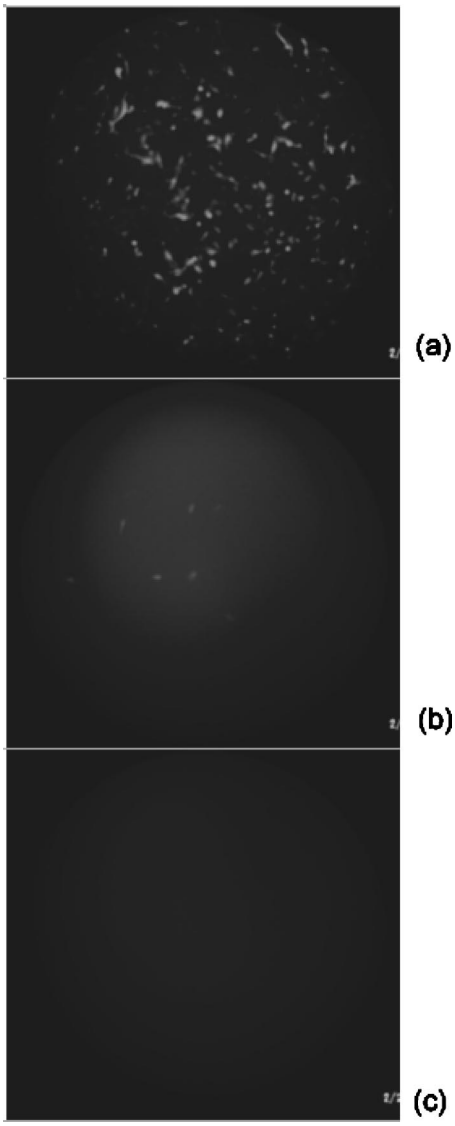


FIG. 2. Transfection results observed by fluorescent microscopy: (a) high transfection ratio (about 15%); (b) low (about 2%); and (c) none (0%), for no ultrasound exposure control.

to the bottom of the culture well. The viable cells were cultured for another 20 h before performing gene expression analysis and cell counting.

3. Gene expression and survival assays

A flow cytometer (FACScan, Becton Dickinson, San Jose, CA) in Duke University Medical Center was used for counting the percentage of cells with fluorescence above a predefined threshold according to base line fluorescence reading from a control group. The percentage of transfected cells was generally determined by counting 10 000 viable cells. Three sample cases of the transfection results are shown in Fig. 2, 15% of transfection [Fig. 2(a)], 2% [Fig. 2(b)], and no exposure control [Fig. 2(c)].

The viability of the cultured cells after ultrasound or sham exposure was analyzed by using Dijindo's cell counting kit-8 (Dojindo Molecular Technologies, Gaithersburg, MD) following the manufacturer's protocol.

C. Experimental procedures

1. Gene transfection using ultrasound

(a) Samples containing 10 μl/ml of GFP plasmid were exposed to 0.44 MPa, 50%-duty-cycle acoustic waves for 20 s. The transfection efficiency was analyzed by a flow cytometer for samples with or without the addition of UCA.

(b) In order to block the possibility of air entrapment by surface agitation, a surface membrane (Parafilm, VWR, West Chester, PA) was placed on top of the fluid surface to block the contact between air and fluid surface. Effort was made to ensure that there was no leakage of fluid or air through the surface membrane. Transfection efficiency was compared for samples with or without a surface membrane.

(c) 0.5 ml of culture medium was added on top of the surface membrane to resume surface agitation, but without the possibility of air entrapment. Transfection efficiency was compared for samples with or without the addition of surface fluid.

(d) Degassing of the culture medium was employed with the intention to remove as much *in situ* pre-existing cavitation nuclei as possible. After 1 h degassing, about 50% of dissolved air could be removed. Experiments were performed to compare the transfection efficiency of samples with or without degassing.

2. Active cavitation detection

The active cavitation detection (ACD) system used in this study comprises of two transducers: the 1.0 MHz source transducer described in Sec. III A and a 15 MHz interrogating transducer (4 mm in diameter, M2062, Panametrics, Waltham, MA), positioned orthogonally to each other [Fig. 4(b)]. The interrogating transducer was mounted through a hole on the sidewall of the culture well with the transducer tip immersed in the culture medium.

The high-frequency interrogating transducer was driven by a pulser/receiver (5052PR, Panametrics), which also triggered simultaneously a function generator (DS345, Stanford Research Systems, Sunnyvale, CA) and an oscilloscope (9310A, LeCroy, Chestnut Ridge, NY). The pulser/receiver enabled the transmission of a single burst from the 15 MHz transducer to the focal area. In receiving mode, the pulser/receiver employed a fixed 36 dB gain preamplifier for the received signal. The received radio frequency wave forms were collected by the digital oscilloscope and stored on a personal computer via the general purpose interface bus interface using the SCOPE EXPLORER program (LeCroy) for offline processing.

For each echo from the 15 MHz transducer, a 2 μs window was opened to sample the received wave form after a delay of 6 μs. That is, we were interested in the bubble activity between 4.5 and 6 mm in front of the ACD transducer (near the center of the culture well). The sampling resolution was 20 ns/pt. A LabView program (v.5.1, National Instruments, Austin, TX) was written to control the insonifying and recording processes of each ultrasound exposure.

In addition, fast Fourier transform (FFT) was applied to 100 wave forms collected from the ACD system and averaged. The rational was: if there are any bubbles present in

our selected acoustic window, either by surface mixing or *in situ* nucleation, high frequency components (15–20 MHz) in the spectrum after FFT should be present.

3. Passive cavitation detection

To investigate the IC activity of the generated bubbles, the same setup shown in Fig. 1(b) was used, but the pulser/receiver was changed to pure listening mode. FFT was performed and averaged for each segment to determine the amplitude of the broadband noise between 0 and 25 MHz. The averaged results of 100 wave forms after performing FFT were compared for experiments with or without the use of a surface membrane.

D. Statistical analysis

Data are presented as means ± standard deviation of at least three replicates. All statistical comparisons were performed by using the student's *t* test embedded in the Sigma Plot software (SPSS Science, Chicago, IL). The significant level was set at $p<0.05$.

IV. RESULTS

A. Gene transfection efficiency

High transfection efficiency (up to 26% of the survival cells) was achieved *in vitro* using our ultrasound system without the presence of a surface membrane. During each experiment, significant surface agitation with resultant visible bubble generation and growth could be observed in the 24-well exposure chamber. To investigate the effect of surface agitation on gene transfection, an acoustically transparent membrane was used to cover the culture medium surface to block the contact between fluid and air, and to suppress surface agitation. In each case, the transfection experiments were performed either with or without the addition of UCA in the culture medium. With surface agitation, effective transfection efficiency (>12%) was observed in the surviving cells; yet, there was no statistically significant difference between samples with or without UCA (Fig. 3). In contrast, when surface agitation was blocked, visible bubble generation and growth in the medium disappeared and the transfection efficiency was found to drop substantially, from more than 12% without the membrane (No UCA, 20 s exposure, 50% duty cycle) to about 1% with the membrane (Fig. 3). Without surface agitation, the addition of UCA in the culture medium could increase the transfection efficiency from about 1% to 4%. However, when the culture medium was partially degassed, the addition of UCA microbubbles had no effect on transfection, probably because of the rapid dissolution of the UCA microbubbles in a degassed environment. Interestingly, in a separate set of experiments when a small amount of fluid (0.5 ml) was added on top of the surface membrane, surface agitation could be re-established and the transfection efficiency was partially resumed (Fig. 3, inset plot).

The typical cell viability following 20 s of ultrasound exposure was 50%–60%. However, cell viability decreased to about 20%–25% after adding the UCA, and the majority of the cell death was found to occur within the initial 5 s of the ultrasound exposure (data not shown).

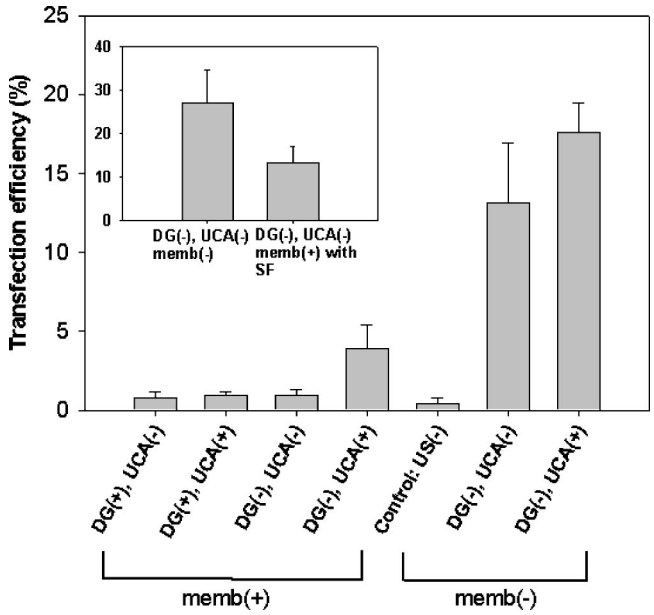


FIG. 3. Comparison of gene transfection efficiency in samples with or without surface membranes, with or without degassing, and with or without UCA. The inset plot shows the transfection efficiency of samples with an open surface (left column) and samples covered by a surface membrane but with additional 0.5 ml culture medium added on top of the membrane. Test to test variation of the transfection results existed, which is shown by the high transfection efficiency (mean: 28%) in the left column of the inset compared to the main plot (14%, 2nd column from the right). The variation might be caused by the differences in the plating generation of the HeLa cells used in these experiments. (DG: degassed; UCA: ultrasound contrast agent; SF: surface fluid; US: ultrasound)

B. Bubble generation during insonification

Inception of detectable bubbles from *in situ* nuclei was detected by ACD primarily in samples without a surface membrane. Figure 4(a) shows the peak-to-peak amplitude of the received signals (15 MHz) sampled around a full 18 s exposure, which is marked by two vertical dashed lines. Clearly, before the 1 MHz transducer was turned on, there was no detectable high-amplitude signal, i.e., no large bubbles existed to scatter the interrogating pulses. It is most likely that any pre-existing cavitation nuclei in the culture medium were too small to be detected. After the 1 MHz transducer was switched on, high-amplitude signals were observed only when the medium surface was open to air (solid line), while the ACD signal remained low when the medium surface was covered by a surface membrane (dashed line). When surface agitation was not suppressed during insonification, some residual high-amplitude signals could still be observed even after the 1 MHz transducer had been switched off (second vertical dashed line), indicating the presence of sizable bubbles in the medium. Furthermore, results from passive cavitation detection (PCD) measurements with (dotted line) or without (solid line) a surface membrane revealed obvious difference in broadband noise spectrum [Fig. 4(b)]. Broadband noise associated with inertial cavitation (large expansion and violent collapse of bubbles from cavitation nuclei, and the generation of daughter bubbles upon inertial collapse) was not seen in samples with a surface membrane. Moreover, ACD data in the frequency domain revealed strong signals around the resonance frequency of the interro-

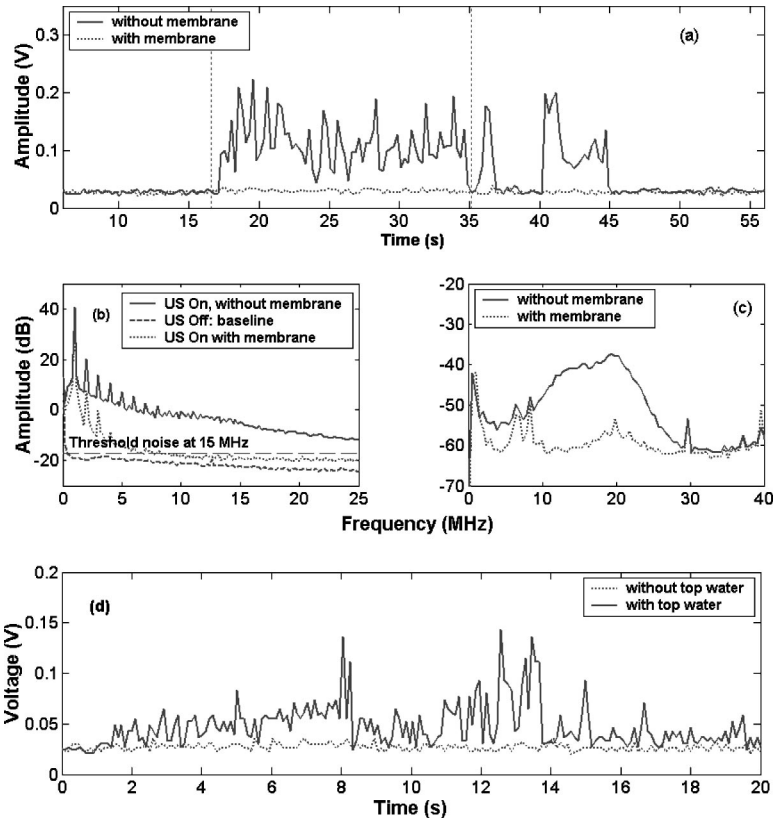


FIG. 4. ACD and PCD results obtained from for degassed water with (dashed line) or without (solid line) a surface membrane: (a) peak-to-peak amplitudes of the bubble signals recorded before (prior to the first vertical dashed line), during (in between two vertical dashed lines) and after (after the second vertical dashed line) the ultrasound exposure; (b) averaged frequency spectrum from a PCD system; (c) averaged frequency spectrum from a ACD system; and (d) bubble generation was evident by ACD after adding fluid on top of the covering surface membrane.

gating transducer (15 MHz) when the 1 MHz transducer was on, indicating the presence of bubbles in the medium [Fig. 4(c), solid line]. Covering the culture medium surface with a thin plastic membrane effectively eliminated the generation of detectable inertial cavitation bubbles [Fig. 4(c), dotted line]. However, adding 0.5 ml of culture medium on top of the surface membrane could partially re-establish the surface agitation and resumed bubble generation [Fig. 4(d)], as well as the transfection efficiency (see inset of Fig. 3).

C. Pressure field inside the culture medium

Figure 5 shows the time evolution of pressures at a position 0.5 mm above the bottom of a culture well along its central axis. Without the surface membrane (solid line), the pressure level fluctuated continuously with the maximal pressure ($\sim$0.8 MPa, peak negative) exceeding the inertial cavitation threshold for bubbles of a few micrometers in radius (see Fig. 6). Higher transient pressures ($>$1 MPa) could be found in other locations in the medium (data not shown). However, when surface agitation was suppressed by a thin membrane (dotted line), the pressure fluctuation was reduced to be within a maximum of 0.45 MPa, below the threshold for IC.

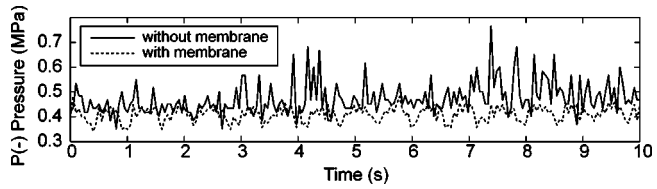


FIG. 5. The time evolution of the pressure change for a selected point at the central axis of the sample dish. Solid line: without a surface membrane. Dotted line: with surface membrane.

D. Inertial cavitation and rectified diffusion thresholds

To better understand the bubble dynamics in our experimental system, the IC and rectified diffusion thresholds for microbubbles ($<10\ \mu$m in radius) in response to 1 MHz ultrasound with a peak negative pressure in the range of 0–0.9 MPa were determined numerically (Fig. 6). Depending on the initial size and the pressure field they experienced, bubbles could dissolve, grow by rectified diffusion, or undergo IC. The pressure levels in our uncovered transfection system ranged between 0.44 and 0.80 MPa (horizontal dashed lines). Two vertical dashed lines in the graph represent the mean size distribution of the shelled Definity bubbles.

For the rectified diffusion threshold, the right arm of the dashed ''V'' shape curve represents the stabilized equilibrium radii at various pressure levels, or an equilibrium ''valley.'' In theory, when a free bubble falls either inside the interior region of the V curve or on the right side of the right arm, the bubble size will eventually reach the stabilized radius at the right arm by either rectified diffusion or gradual dissolution. For example, a 5.5 μm bubble in a 0.5 MPa ultrasound field will gradually grow in size due to rectified diffusion (left empty circle, moving toward right), whereas a 9.5 μm bubble will shrink gradually (right empty circle, moving toward left). Both bubbles will eventually reach the stabilized radius of 6.9 μm.

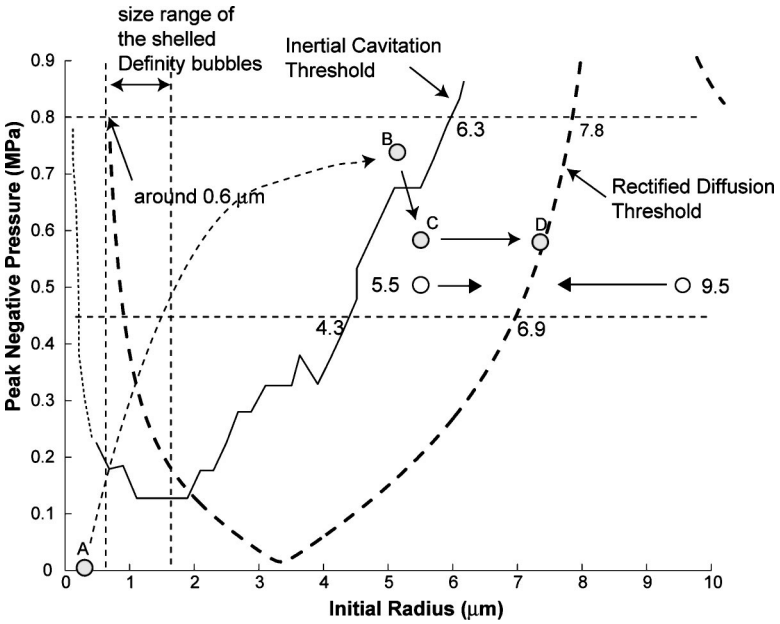


FIG. 6. This plot shows the inertial cavitation (solid line) and rectified diffusion (dashed line) thresholds of a single air bubble of various initial radii oscillating in an acoustic field of different pressure amplitudes.

The rectified diffusion threshold on the left arm of the V curve is a ''quasistatic'' equilibrium, or an equilibrium ''ridge.'' All bubbles on the right side of this arm will increase in size and move toward the right arm (due to rectified diffusion). In contrast, bubbles reside on the left side of the left arm will gradually dissolve.

The interior of the truncated ''U'' curve (solid line) is the region where bubbles will undergo IC with a collapsing speed of the bubble wall greater than 1500 m/s. Bubbles with initial size outside the U shape region will undergo stable cavitation once the acoustic field is present. It is important to note here that the process of rectified diffusion requires a time scale of seconds, while IC usually occurs in microseconds. Therefore, if pressure is sufficiently high, IC always occurs prior to rectified diffusion. For instance, if a bubble initially falls inside both the interior of the V and U curves, it will undergo inertial collapse first and presumably break into smaller daughter bubbles. Subsequently, if the size of the daughter bubbles falls outside the V curve but inside the U curve, the bubbles will undergo IC again, breaking into even smaller sizes and gradually dissolved.

Altogether, in our gene transfection system with an *in situ* pressure amplitude between 0.44 and 0.8 MPa, any free air bubbles with an initial radius greater than 6.3 μm would either grow gradually by rectified diffusion or dissolved progressively to reach a stabilized radius between 6.9 and 7.8 μm. If the bubbles were less than 4.3 μm in radius, they would undergo IC, break into smaller daughter bubbles, and dissolve once they were outside the left size limit of the IC threshold curve.

E. Shear stresses generated by shelled and free bubbles

Figures 7(a) and (b) show, respectively, the changes in radius and shear stresses generated by the nonlinear oscillation of a free air bubble and a shelled Definity bubble in response to 1 MHz ultrasound with a peak negative pressure of 0.44 MPa. For the shelled Definity bubble, which has a size distribution primarily in the range between 0.55 and 1.65 μm [vertical dashed lines in Fig. 7(b)], the radius changes were small, and thus low shear stresses would be expected. Based on model calculation, the maximal shear stress was only about 2 kPa at this pressure level. In comparison, the amplitude of the radial oscillation of a free air bubble was much larger, leading to the generation of significantly higher shear stresses within the same size range. The discontinuity parts of both solid and dashed lines in Fig. 7 represent the region of IC, where shear stresses cannot be calculated using equations of stable cavitation. It should be noted that the shear stress generated by the inertial collapse of the bubble itself was much larger but short-lived. On the other hand, for free bubbles larger than 4.3 μm in radius, they would undergo stable cavitation, persist in the acoustic field, and exert steady and large shear stresses. A maximal shear stress of about 50 kPa was predicted at a bubble radius of 6.9 μm, the equilibrium radius of 1 MHz ultrasound at

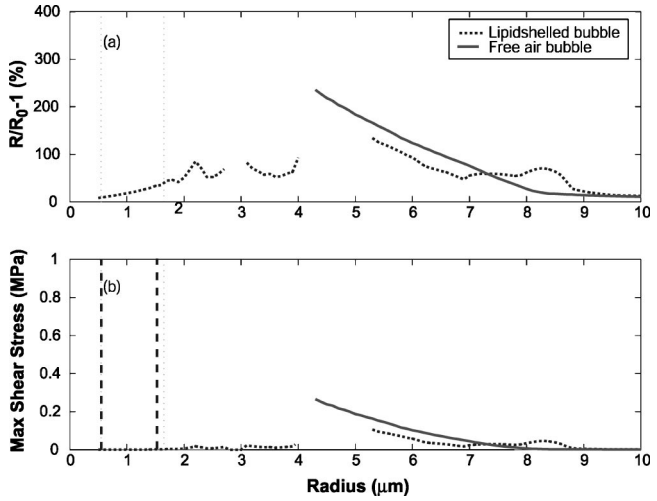


FIG. 7. The simulation results of (a) radius change and (b) shear stresses generated by either a shelled Definity bubble (dashed line) or a free air bubble (solid line) with different initial radii.

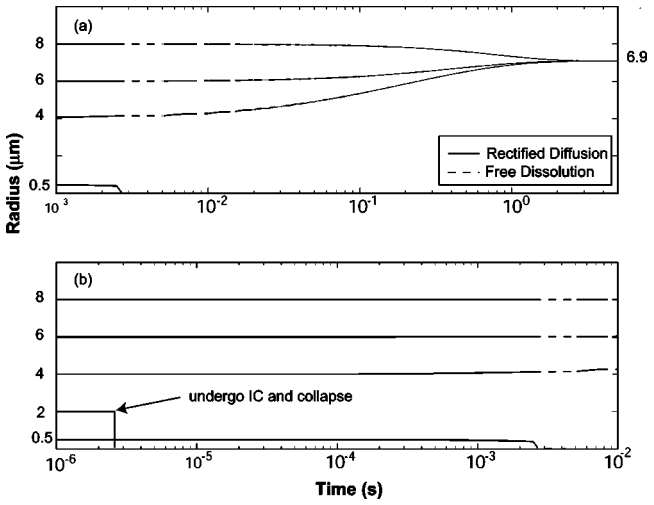


FIG. 8. The temporal size changes of bubbles with different initial radii ((a) 0.5, 4, 6, and 8 μm and (b) 0.5, 2, 4, 6, and 8 μm from the bottom line to the top one) insonified by 1 MHz pulsed ultrasound waves with peak negative pressure = 0.44 MPa, PRF = 200 Hz, and pulse length = 2500 cycles. Plot (b) shows the changes of bubble size of plot (a) in the initial 0.01 s.

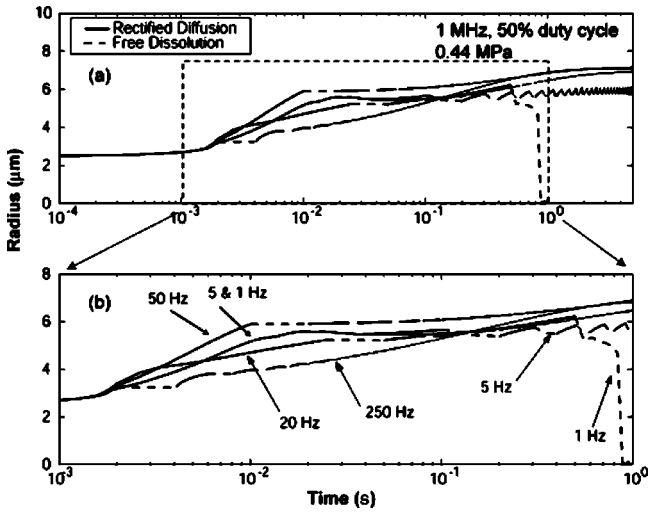


FIG. 9. The temporal size variation of a 2.5 μm bubble insonified by 1 MHz pulsed ultrasound waves of 0.44 MPa at different off periods or pulse repetition frequencies (in Hz, indicated in each line). Plot (b) shows the detailed changes of bubble size of plot (a) from 0.001 to 1 s.

0.44 MPa [Fig. 7(b)]. Overall, the model calculation shows that shelled contrast bubbles, if persisted, yielded much lower shear stresses than free air bubbles.

F. Rectified diffusion and dissolution of free air bubbles

In our gene transfection experiments, pulsed ultrasound of 50% duty cycle was used to avoid rapid heat accumulation. In such an acoustic field, a free bubble may undergo rectified diffusion during the ''on'' period of the acoustic cycle, and dissolve freely during the ''off'' period. Figure 8 shows the model prediction for five bubbles of different initial sizes at a peak negative pressure of 0.44 MPa. Bubbles with initial radii of 4 and 6 μm grew, while the 8 μm one shrank. Yet, when sufficient exposure time was allowed they all eventually reached the stabilized equilibrium radius of 6.9 μm. The 2 μm bubble underwent inertial collapses at the second acoustic cycle and disappeared. The 0.5 μm bubble, since its size fell outside the threshold of IC, dissolved. Figure 8(b) shows the initial stage of Fig. 8(a).

Figure 9 shows the changes in bubble radius of a 2.5 μm bubble exposed to a 1 MHz sinusoidal wave of 0.44 MPa in peak pressure, but with different length of off time (same duty cycle but different PRF). Figure 9(b) is the enlargement of Fig. 9(a). The longer the off period (lower PRF) was, the longer the growth time would be needed for the bubble to reach the stabilized size. At an extremely low PRF of 1 Hz, bubble dissolved during the long off period. Therefore, a critical off period exists for bubbles of different sizes to reach their stabilized radii by rectified diffusion. For example, at 50% duty cycle, the maximal off time allowed is 0.1 s, which corresponds to a PRF of 5 Hz.

V. DISCUSSION

In this study, a series of experiments were carried out to assess gene transfection *in vitro* induced by 1 MHz ultrasound. Following exposure to pulsed ultrasound of 50% duty cycle for 20 s, cultured HeLa cells showed about 15% gene transfection in the surviving population. The addition of Definity contrast agent did not yield more transfection in our setup. It was found that surface agitation is necessary for producing effective gene transfection in our exposure setup. The surface agitation created a fluctuated pressure field with the generation of transient high-pressure foci and detectable free bubbles in the fluid (Fig. 5). Broadband noise, which is characteristic of IC, was also observed during the ultrasound exposure. These bubbles might subsequently undergo further IC, or increase in size by rectified diffusion.[30,31] The accompanying microstreaming and shear stresses during nonlinear oscillation of these bubbles might cause a compromise of cell membranes and lead to plasmid DNA transportation.

The culture medium used in this study was saturated with air, but the originally dissolved bubbles or nuclei were too small to be detected by our ACD system. The resonant size of a free air bubble at 15 MHz (the interrogating ultrasound frequency) is about 0.4 μm in radius and the backscattering cross sections of both free air and lipid-shelled bubbles at this frequency are similar. Because the scattering signals from the lipid-shelled Definity bubbles could be easily detected, we believe that any pre-existing cavitation nuclei in the culture medium should be smaller than 0.4 μm in radius. Based on our model calculation, these small nuclei were around the left border of the IC threshold and outside the left limit of the rectified diffusion threshold (Fig. 6). Therefore, these small nuclei, if not stabilized by impurities or boundaries in the exposure system, would simply dissolve.

After switching on the therapeutic transducer, significant amount of bubble and IC signals were detected by our ACD system in samples with an open medium surface. Transient high-pressure loci might be produced by focusing of reflected waves from the convex fluid surface produced by surface agitation, and/or wave summation inside the medium. The resultant pressure could exceed the IC threshold for pre-existing cavitation nuclei in the culture medium, and

thus induced IC. Because the high-pressure field was transient, we speculated that the transient high-pressure foci induced inception of bubbles by IC, while pressure fluctuation allowed the persistence and growth of the generated bubbles by rectified diffusion (Fig. 6). The gray circles in Fig. 6 demonstrated a possible fate of a small cavitation nucleus after exposure to the acoustic field in our system. A small nucleus (position A) experienced transient high pressure and increased in size rapidly by IC. This bubble (at position B) might be subsequently stabilized in a lower pressure field and grew by rectified diffusion (position C). A final equilibrium size could be reached after a few seconds (position D). When a surface membrane was used to suppress the surface agitation, both pressure fluctuation and transient high pressure would be eliminated. Neither bubbles nor shockwaves generated during inertial collapse were detected (negative ACD and PCD results), indicating minimal cavitation activity. These results corroborate the significantly reduced transfection efficiency observed under this experimental condition. By adding a small amount of fluid on top of the surface membrane, surface agitation resumed, so did the bubble generation and gene transfection [see Figs. 4(d) and 3 insert].

At a pressure level of 0.44 MPa, our model calculation suggests that if a free air bubble is smaller than 4.3 μm in radius, the bubble will enter the IC region and break into smaller daughter bubbles, which eventually dissolve in the medium (Fig. 6). However, for bubbles with an initial radius larger than 4.3 μm, they will gradually reach a stabilized equilibrium radius of 6.9 μm by either rectified diffusion or dissolution. These stabilized bubbles will oscillate continuously throughout the whole exposure duration, thus exerting strong and sustained shear stresses up to 250 Pa in the surrounding medium [Fig. 7(b)]. Shear stresses from stable cavitation have been shown to be able to compromise cell membrane temporarily. Using a vibrating Mason horn operating at 21.4 kHz, Wu *et al.* found that the threshold of shear stress to produce reparable sonoporation was about 12 Pa for exposures up to 7 min.[33] At a pressure level lower than the IC threshold, Kim *et al.* found that gene transfection could be produced at 1 MHz but not at 3.5 MHz.[14] Recently, it has also been shown that gentle linear bubble oscillations (acoustic pressure of 0.1 bar and frequency of 180 kHz) are sufficient to achieve rupture of lipid membranes.[34] Shear stresses could also be generated by inertial collapse of cavitation bubbles. A shear stress of less than 50 Pa was found to have the ability to change membrane permeability in a shockwave-induced cavitation system.[35–38] Altogether, both experimental and theoretical evidence accumulated in this study suggest that cavitation, both stable and inertial cavitation, and their associated shear stresses produced by microstreaming are probably the primary mechanism for ultrasound-mediated gene transfection in our system.

The existence and distribution of cavitation nuclei in tissue are quite different from that in *in vitro* cell studies. How to increase ultrasound-mediated gene transfection efficiency *in vivo*, where large free air bubbles do not exist, is still a significant challenge in the field. Several possible strategies can be explored in future studies. First of all, the amplitude of the shear stress generated by oscillating lipid-shelled bubbles is size dependent. In Fig. 7(b), a larger Definity microbubble ($>5.3\ \mu$m in radius), if not fragmented, generates comparable shear stress with a free air bubble of similar size. Thus, contrast agents that are specially designed for therapeutic use (e.g., larger size and more resilient to break) may be beneficial for ultrasound-mediated gene transfection *in vivo*. A recent study using two UCAs with similar albumin shells and perfluorocarbon gas contents demonstrates that perfluorocarbon-exposed-sonicated-dextrose-albumin (PESDA) microbubbles has superior gene transfection efficiency than Optison®.[39] A possible explanation of this observation is that PESDA contains more microbubbles with diameters above 8 μm, and PESDA has less soluble gas (C^4F^{10}) than Optison (C^3F^8). A larger UCA microbubble produces higher shear stress before fragmentation, and may break into larger daughter bubbles that again produce higher shear stresses. Higher transfection efficiency is thus reasonable.

Second, free bubbles can be generated from cavitation nuclei by cotreatment of the gas-saturated fluid with a low-frequency acoustic field or a transient high-pressure field[40] during ultrasound-mediated gene transfection. Modalities such as low-frequency ultrasound, high-intensity focused ultrasound, or shockwaves can be used to induce bubbles in the target tissue. Once the treatment ultrasound wave is switched on, the generated bubbles will establish new equilibrium radii and exert sufficient shear stresses necessary for gene transfection. The induction of free bubbles for gene transfection may have important implication for *in vivo* applications since UCA may not be readily available for the target cells in tissue without breaking the vessel barriers.

VI. CONCLUSIONS

In this study, we demonstrated *in vitro* that effective ultrasound-mediated gene transfection could be produced without the use of contrast microbubbles. The *in situ* generation of free bubbles, which was shown to be associated with surface agitation of the culture medium during insonification, was the key factor for producing effective gene transfection. The resultant pressure fluctuation and/or transient high-pressure loci in culture medium induced IC and created sustained bubble activities. Our results exhibit important implications for future studies of ultrasound-mediated gene transfection *in vivo*, in which adequate bubble environment needs to be artificially created. Bubbles necessary for sonoporation effect may be locally generated by intensive but intermittent acoustic waves such as shockwaves or pulsed high-intensity focused ultrasound. Low-frequency ultrasound with a lower IC threshold may also work.

ACKNOWLEDGMENTS

This work was supported in part by NIH Grant Nos. R01-EB02682, R21EB03299, and R21-CA91166. The authors sincerely thank Dr. Zaiquan Li for technical assistance and intuitive discussion, and the Bristol-Myers Squibb Company (North Billerica, MA) for kindly providing the Definity contrast agent.

TABLE I. Constants and variables used in the simulation.

Constants /variables	Description	Value
c_∞	Concentration of dissolved gas in solution (kg/m^3)	
c_0	Saturation concentration of dissolved gas (kg/m^3)	
D	Diffusion coefficient (m^2/s)	2.4×10^{-9}
d_{se}	Shell thickness (m)	Free air bubble: 0 Definity: 4×10^{-9}[a]
f	Frequency (Hz), $\omega=2\pi f$	1×10^6
G	Velocity gradient (s^{-1})	
G_s	Shear modulus of the shell (MPa)	50[a]
L_A	Ostwald coefficient of air	1.71×10^{-2}
L_{ms}	Thickness of the velocity gradient for a pulsating microbubble (m)	
M	Molecular weight of gas (kg/mole)	
P_a	Peak acoustic pressure (Pa)	2×10^5
P_0	Ambient pressure (Pa)	101300
P_v	Vapor pressure (N/m^2)	2330
R	Radius (m)	
R_0	Initial or equilibrium radius (m)	
$R_{\varepsilon0}$	Change in radius for a pulsating bubble, $R-R_0$ (m)	
R_G	Gas constant (J/mole °K)	8.314[b]
S	Shear stress (Pa)	
S_p	Shell elasticity parameter	Calculated from G_s and d_{se}
T	Temperature (°K)	310
t	Time (s)	
η_L	Shear Viscosity of water (K g/m s)	0.001
δ	Total damping factor	
δ_{th}	Thermal damping	
δ_R	Radiation damping	
δ_η	Viscous damping	
δ_f	Friction damping by shell	
ρ	Density of water ($K g/m^3$)	1000
κ	Polytropic exponent of gas	1.0002
σ	Surface tension (N/m)	73×10^{-3}[b]
ω_0	Resonance frequency	[rad/s]

[a]Assume that the shell thickness of Definity contrast bubble is equal to that of the Sonazoid, since both are phospholipids shelled microbubbles.
[b]The CRC Handbook of Chemistry and Physics, 80th ed.

APPENDIX

A. Modified Rayleigh–Plesset equations for free (Eq. A1) and shelled (Eq. A3) bubble oscillation

$$\rho R\ddot{R}+\frac{3}{2}\rho R^2=\left(P_0-P_v+\frac{2\sigma}{R_0}\right)\left(\frac{R_0}{R}\right)^{3\kappa}-P_g-\delta\omega\rho R\dot{R}+P_a\sin(\omega\cdot t), \tag{A1}$$

where

$$P_g=P_0-P_v+\frac{2\sigma}{R}. \tag{A2}$$

The value of κ in use was 1.0002, since the bubble radius was in the order of the thermal diffusion length and the bubble was assumed undergoing isothermal oscillation. The detailed descriptions of the other variables and constants used in the above and the following equations are listed in Table I:

$$\rho R\ddot{R}+\frac{3}{2}\rho R^2=\left(P_0-P_v+\frac{2\sigma}{R_0}\right)\left(\frac{R_0}{R}\right)^{3\kappa}-P_g-2S_p\left(\frac{1}{R_0}-\frac{1}{R}\right)-\delta\omega\rho R\dot{R}+P_a\sin(\omega\cdot t), \tag{A3}$$

where S_p, a shell elasticity parameter, was related to shear modulus (G_s) and shell thickness (d_{se}) by

$$S_p=6G_sd_{se}\left(\frac{R}{R_0}\right)^3. \tag{A4}$$

In addition to thermal (δ_{th}), radiation (δ_R), and viscous (δ_η) damping factors for an oscillating bubble in water, the total damping factor (δ) for a shelled contrast bubble includes an additional friction damping term (δ_f):

$$\delta=\delta_{th}+\delta_R+\delta_\eta+\delta_f, \tag{A5}$$

The details of damping factors (δ_{th}, δ_{th}, δ_R, δ_η) are described elsewhere.[24,25]

B. Shear stresses generated by bubble oscillation

The shear stress generated was found to be equal to

$$S=\eta_LG=\frac{\eta_L\omega R_{\varepsilon0}^2}{R_0L_{ms}}, \tag{B1}$$

where G was the velocity gradient, $R_{\varepsilon0}=R-R_0$ was the change in bubble radius during pulsation, and the thickness of the boundary layer (L_{ms}) can be described by

$$L_{ms}=\sqrt{\frac{2\eta_L}{\rho\omega}}. \tag{B2}$$

C. Rectified diffusion

$$\frac{dR_0}{dt}=\frac{Dd}{R_0}\left[\langle R/R_0\rangle+R_0\left(\frac{\langle(R/R_0)^4\rangle}{\pi Dt}\right)^{1/2}\right]\times\left(1+\frac{4\sigma}{3P_0R_0}\right)^{-1}\left(\frac{c_\infty}{c_0}-\frac{\langle(R/R_0)^4(P_g/P_0)\rangle}{\langle(R/R_0)^4\rangle}\right), \tag{C1}$$

where

$$d=\frac{R_GTc_0}{P_0}. \tag{C2}$$

The detailed expressions for the time averages $\langle R/R_0\rangle$, $\langle(R/R_0)^4\rangle$ and $\langle(R/R_0)^4(P_g/P_0)\rangle$ are given elsewhere.[30]

The threshold pressure of the rectified diffusion can be obtained by setting $dR_0/dt=0$, and the simplified equation in megahertz range becomes

$$P_a=\left(\frac{(\rho R_0^2\omega_0^2)^2\left[\left(1-\frac{\omega^2}{\omega_0^2}\right)^2+\delta_\eta^2\right]\left[1+\frac{2\sigma}{R_0P_0}-\frac{c_\infty}{c_0}\right]}{(3+4K)\left(\frac{c_\infty}{c_0}\right)-K\left(1+\frac{2\sigma}{R_0P_0}\right)}\right)^{1/2}, \tag{C3}$$

where

$$K=\frac{\left(1-\frac{\rho\omega^2R_0^2}{12P_0}\right)+\frac{5\sigma}{3P_0R_0}}{1+\frac{4\sigma}{3P_0R_0}}. \tag{C4}$$

D. Static diffusion

$$\frac{dR}{dt}=-DL_A\frac{1-\frac{c_\infty}{c_0}+\frac{\tau}{R\rho}}{1+\frac{2\tau}{3R\rho}}\left(\frac{1}{R}+\frac{1}{\sqrt{\pi Dt}}\right),\quad \tau=\frac{2M\sigma}{R_GT}. \tag{D1}$$

The initial condition for both the rectified and static diffusion used in the calculation is $R_{(t=0)}=R_0$.

[1] D. J. Rader, "Gene therapy for atherosclerosis," Int. J. Clin. Lab Res. **27**, 35–43 (1997).
[2] A. Fischer, S. Hacein-Bey, and M. Cavazzana-Calvo, "Gene therapy of severe combined immunodeficiencies," Nat. Rev. Immunol. **2**, 615–621 (2002).
[3] J. C. Cusack, Jr. and K. K. Tanabe, "Introduction to cancer gene therapy," Surg. Oncol. Clin. N. Am. **11**, 497–519 (v 2002).
[4] S. Hacein-Bey-Abina, A. Fischer, and M. Cavazzana-Calvo, "Gene therapy of X-linked severe combined immunodeficiency," Int. J. Hematol. **76**, 295–298 (2002).
[5] N. R. Hackett, S. M. Kaminsky, D. Sondhi, and R. G. Crystal, "Antivector and antitransgene host responses in gene therapy," Curr. Opin. Mol. Ther. **2**, 376–382 (2000).
[6] B. D. Brown and D. Lillicrap, "Dangerous liaisons: the role of "danger" signals in the immune response to gene therapy," Blood **100**, 1133–1140 (2002).
[7] E. Check, "Cancer fears cast doubts on future of gene therapy," Nature (London) **421**, 678 (2003).
[8] E. Check, "Second cancer case halts gene-therapy trials," Nature (London) **421**, 305 (2003).
[9] S. Hacein-Bey-Abina *et al.*, "A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency," N. Engl. J. Med. **348**, 255–256 (2003).
[10] T. Niidome and L. Huang, "Gene therapy progress and prospects: nonviral vectors," Gene Ther. **9**, 1647–1652 (2002).
[11] R. V. Shohet, S. Chen, Y. T. Zhou, Z. Wang, R. S. Meidell, R. H. Unger, and P. A. Grayburn, "Echocardiographic destruction of albumin microbubbles directs gene delivery to the myocardium," Circulation **101**, 2554–2556 (2000).
[12] Q. L. Lu, H. D. Liang, T. Partridge, and M. J. Blomley, "Microbubble ultrasound improves the efficiency of gene transduction in skeletal muscle in vivo with reduced tissue damage," Gene Ther. **10**, 396–405 (2003).
[13] D. B. Tata, F. Dunn, and D. J. Tindall, "Selective clinical ultrasound signals mediate differential gene transfer and expression in two human prostate cancer cell lines: LnCap and PC-3," Biochem. Biophys. Res. Commun. **234**, 64–67 (1997).
[14] H. J. Kim, J. F. Greenleaf, R. R. Kinnick, J. T. Bronk, and M. E. Bolander, "Ultrasound-mediated transfection of mammalian cells," Hum. Gene Ther. **7**, 1339–1346 (1996).
[15] W. J. Greenleaf, M. E. Bolander, G. Sarkar, M. B. Goldring, and J. F. Greenleaf, "Artificial cavitation nuclei significantly enhance acoustically induced cell transfection," Ultrasound Med. Biol. **24**, 587–595 (1998).
[16] Y. Taniyama *et al.*, "Local delivery of plasmid DNA into rat carotid artery using ultrasound," Circulation **105**, 1233–1239 (2002).
[17] C. Teupe *et al.*, "Vascular gene transfer of phosphomimetic endothelial nitric oxide synthase (S1177D) using ultrasound-enhanced destruction of plasmid-loaded microbubbles improves vasoreactivity," Circulation **105**, 1104–1109 (2002).
[18] A. Lawrie, A. F. Brisken, S. E. Francis, D. C. Cumberland, D. C. Crossman, and C. M. Newman, "Microbubble-enhanced ultrasound for vascular gene delivery," Gene Ther. **7**, 2023–2027 (2000).
[19] AIUM, "Mechanical bioeffects from diagnostic ultrasound: AIUM consensus statements. American Institute of Ultrasound in Medicine," J. Ultrasound Med. **19**, 69–72 (2000).
[20] Lord Rayleigh, "On the pressure developed in a liquid during the collapse of a spherical cavity," Philos. Mag. **34**, 94–98 (1917).
[21] M. S. Plesset, "The dynamics of cavitation bubbles," J. Appl. Mech. **16**, 277–282 (1949).
[22] P. P. Chang and K. K. Shung, "Interaction of ultrasound and contrast agents," in *Trends in Contrast Media* (Springer, New York, 1999), pp. 311–320.
[23] L. Hoff, P. C. Sontum, and J. M. Hovem, "Oscillations of polymeric microbubbles: Effect of the encapsulating shell," J. Acoust. Soc. Am. **107**, 2272–2280 (2000).
[24] L. Hoff, *Acoustic Characterization of Contrast Agents for Medical Ultrasound Imaging* (Kluwer Academic, Dordrecht, 2001).
[25] C. C. Church, "The effects of an elastic solid surface layer on the radial pulsations of gas bubbles," J. Acoust. Soc. Am. **97**, 1510–1521 (1995).
[26] R. E. Apfel, *Methods in Experimental Physics* (Academic, New York, 1981).
[27] W. L. Nyborg, "Acoustic streaming near a boundary," J. Acoust. Soc. Am. **30**, 329–339 (1958).
[28] T. G. Leighton, "The forced bubble," in *The Acoustic Bubble* (Academic San Diego, 1994), pp. 287–438.
[29] J. Wu, "Theoretical study on shear stress generated by microstreaming surrounding contrast agents attached to living cells," Ultrasound Med. Biol. **28**, 125–129 (2002).
[30] L. A. Crum, "Rectified diffusion," Ultrasonics **22**, 215–223 (1984).
[31] A. Eller and H. G. Flynn, "Rectified diffusion during nonlinear pulsations of cavitation bubbles," J. Acoust. Soc. Am. **37**, 493–503 (1965).
[32] W. S. Chen, T. J. Matula, and L. A. Crum, "The disappearance of ultrasound contrast bubbles: Observations of bubble dissolution and cavitation nucleation," Ultrasound Med. Biol. **28**, 793–803 (2002).
[33] J. Wu, J. P. Ross, and J. F. Chiu, "Reparable sonoporation generated by microstreaming," J. Acoust. Soc. Am. **111**, 1460–1464 (2002).
[34] P. Marmottant and S. Hilgenfeldt, "Controlled vesicle deformation and lysis by single oscillating bubbles," Nature (London) **423**, 153–156 (2003).
[35] M. Delius, P. H. Hofschneider, U. Lauer, and K. Messmer, "Extracorporeal shock waves for gene therapy?," Lancet **345**, 1377 (1995).
[36] D. L. Miller, S. Bao, R. A. Gies, and B. D. Thrall, "Ultrasonic enhancement of gene transfection in murine melanoma tumors," Ultrasound Med. Biol. **25**, 1425–1430 (1999).
[37] L. Junge, C. D. Ohl, B. Wolfrum, M. Arora, and R. Ikink, "Cell detachment method using shock-wave-induced cavitation," Ultrasound Med. Biol. **29**, 1769–1776 (2003).
[38] C. D. Ohl and B. Wolfrum, "Detachment and sonoporation of adherent HeLa-cells by shock wave-induced cavitation," Biochim. Biophys. Acta **1624**, 131–138 (2003).
[39] S. V. Pislaru, C. Pislaru, R. R. Kinnick, R. Singh, R. Gulati, J. F. Greenleaf, and R. D. Simari, "Comparison of two albumin-based echocardiographic contrast agents for ultrasound-induced gene transfer," Proc. 2nd Int. Symp. Therapeutic Ultrasound, 2003, Vol. **1**, pp. 506–509.
[40] D. Melodelima, J. Y. Chapelon, Y. Theillere, and D. Cathignol, "Combination of thermal and cavitation effects to generate deep lesions with an endocavitary applicator using a plane transducer: *Ex vivo* studies," Ultrasound Med. Biol. **30**, 103–111 (2004).

Experimental validation of a tractable numerical model for focused ultrasound heating in flow-through tissue phantoms

Jinlan Huang, R. Glynn Holt,[a] Robin O. Cleveland, and Ronald A. Roy
Boston University, Department of Aerospace and Mechanical Engineering, Boston, Massachusetts 02215

(Received 27 February 2004; revised 6 July 2004; accepted 11 July 2004)

Heating from high intensity focused ultrasound (HIFU) can be used to control bleeding, both from individual blood vessels as well as from gross damage to the capillary bed. The presence of vascularity can limit one's ability to elevate the temperature owing to convective heat transport. In an effort to better understand the heating process in tissues with vascular structure we have developed a numerical simulation that couples models for ultrasound propagation, acoustic streaming, ultrasound heating and blood cooling in a Newtonian viscous medium. The 3-D simulation allows for the study of complicated biological structures and insonation geometries. We have also undertaken a series of *in vitro* experiments employing non-uniform flow-through tissue phantoms and designed to provide verification of the model predictions. We show that blood flow of 2 cm/s (6.4 ml/min through a 2.6 mm 'vessel') can reduce peak temperature in a vessel wall by 25%. We also show that HIFU intensities of 6.5×10^5 W/m^2 can induce acoustic streaming with peak velocities up to 5 cm/s and this can reduce heating near a vessel wall by more than 10%. These results demonstrate that convective cooling is important in HIFU and can be accounted for within simulation models. [DOI: 10.1121/1.1787124]

PACS numbers: 43.80.Gx, 43.80.Sh, 43.25.Nm [FD] Pages: 2451–2458

I. INTRODUCTION

High intensity focused ultrasound (HIFU), often referred to as focused ultrasound surgery (FUS), is a therapeutic medical treatment modality that is enjoying a renewed interest in research and clinical applications.[1–4] The manner in which one applies the acoustic energy determines, to a large extent, the nature and spatial extent of the biological effect. A broad spectrum of therapy is achievable, ranging from gentle heating of tumors to violent tissue ablation, from drug delivery through sonoporation to kidney stone comminution. A barrier to safe and efficacious HIFU therapy involves targeting and treatment planning: is the sound energy going to the right spot and in the correct dosage? To address this, one needs an accurate model that, in the context of HIFU therapy, includes multiple physical effects: nonlinear sound propagation, arbitrary media inhomogeneity, thermal transport phenomena, convective transport phenomena (i.e., blood flow) and other second-order effects as needed (acoustic streaming, acoustic radiation stress, and cavitation).

The years have seen numerous studies devoted to modeling tissue heating from ultrasound exposure, starting with the pioneering works of Fry,[5] Lele,[6] and Parker,[7,8] the latter of whom derived approximate analytical expressions for the temperature rise in a 2-D conducting medium. Most published methods are limited to simple situations for which analytical solutions exist and the use of cylindrical geometries suffice. Pond[9] broke down the heating volume into a series of cylinders that he used as elemental heat sources. Robinson and Lele[10] assumed a cylindrical heat deposition pattern and produced analytical expressions for the temperature distributions as a function of time. Muir and Carstensen[11] studied the physics of both focused and unfocused ultrasound heat deposition and produced analytical expressions that incorporated nonlinear effects. Several investigators[12–15] considered how to accurately compute heat generation during ultrasound exposure. Lizzi *et al.*[16,17] developed both analytical and numerical solutions to the Pennes bioheat transfer equation[18] in the context of glaucoma treatment. Hill *et al.*[19] and later Wu and Du[20] developed general analytical models based on a Gaussian approximation to the beam shape.

Owing to the complexity of the propagation medium, accurate predictions of HIFU-induced temperature rise are generally facilitated by numerical modeling. Kolios *et al.*[21] employed a finite difference method to model perfused tissues in 2-D cylindrical coordinates. Curra *et al.*[22] used a 2-D finite difference implementation to investigate the importance of nonlinear effects on wave propagation and heat generation in perfused liver models. Fan and Hynynen[23] investigated FUS by phased arrays using a 3-D finite difference model. Hoffelner *et al.*[24] developed a finite element method to solve the KZK equation. Wan *et al.*[25] used a matrix relaxation method to investigate critical parameters governing the performance of their phased-array system and Meaney *et al.*[26] computed 3-D heat deposition patterns assuming a linear propagation model. Krasovitski and Kimmel[27] presented a 3-D simulation of the temperature field in and around a blood vessel for a simplified geometry. The literature on modeling of ultrasound-induced tissue heating is extensive. Readers are directed to the review by Bailey *et al.*[28] for a comprehensive survey.

This paper describes a finite-difference-time-domain (FDTD) implementation of the basic governing equations for sound propagation and heat transfer in the presence of blood

[a]Author to whom correspondence should be addressed. Electronic mail: rgholt@bu.edu

flow and acoustic streaming. We use an implementation of the coupled nonlinear acoustic propagation and absorption (resulting in spatio-temporally resolved thermal source terms which are then incorporated into a numerical implementation of the heat conduction equation) originally introduced by Hallaj and Cleveland.[29,30] For the current work we have modified the implementation to include a convective term in the heat conduction equation. The simulation is 2-D (axisymmetric) in pressure and 3-D in all thermal and flow quantities. Numerical predictions are assessed through a quantitative comparison with experiments run in uniform and vascularized gel phantoms. The acoustic and thermal properties of these phantoms were independently (experimentally) characterized, and the phantoms were instrumented to facilitate measurement of acoustic pressure and temperature in space and time during the heating experiments. The comparison with model predictions proceeds without reliance on fitting parameters. Described below are the theoretical model, numerical implementation, and subsequent experimental validation. Additional details describing all facets of the work are presented by Huang.[31]

II. MODEL

We model nonlinear sound propagation in a thermoviscous medium with a modified Westervelt equation, accounting for the effects of diffraction, absorption, and nonlinearity:[32]

$$\left(\nabla^2-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)p+\frac{\delta}{c^4}\frac{\partial^3 p}{\partial t^3}+\frac{\beta}{\rho c^4}\frac{\partial^2 p^2}{\partial t^2}=0. \tag{1}$$

Here c is the sound speed, ρ is the density, δ is the acoustic diffusivity, and $\beta=1+B/2A$ is the coefficient of nonlinearity with B/A being the nonlinearity parameter of the medium. The simulation is cast in the time domain, which facilitates the investigation of the spatial and temporal characteristics of the energy deposition and heating in an arbitrary medium using either pulsed or continuous ultrasound. The model assumes a classical thermoviscous medium in which the absorption increases as the frequency squared, though for tissues the power law for absorption is closer to $f^{1.1}$. This will introduce error in the case of strongly nonlinear waves. Using a plane wave analysis of nonlinear propagation we estimated that the error in the heating term is less than 5% for focal peak pressures up to 2 MPa (the range used in this study). Moreover, in our experiments the FUS fields generated acoustic cavitation before significant propagation nonlinearity set in, and, once bubbles are present, the model reported herein is no longer valid.[33–36]

We assume that the acoustic properties of our *in vitro* setup are such that the acoustic field can be calculated assuming a uniform medium. Therefore the sound field generated by our axisymmetric transducer can be adequately described by a 2-D cylindrical computation, a fact that *greatly* reduces the computation time required for each simulation. This ''engineering approximation'' is motivated purely by practical considerations and recognizes the fact that the acoustic contrast of various structures in tissue is small (bone and lung being notable exceptions).

Finally, in order to couple the pressure field model to the temperature field model (described below) we need to quantify the thermal energy deposition associated with the absorption of the ultrasonic wave. The following expression, adapted from Ref. 37, gives the spatially dependent ultrasonic power deposition per unit volume in a nonrelaxing medium:[38]

$$q=2\alpha_{ABS}I=\frac{2\alpha_{ABS}}{\omega^2\rho c}\left\langle\left(\frac{\partial p}{\partial t}\right)^2\right\rangle. \tag{2}$$

Here, α_{ABS} refers to the local absorption coefficient of the medium, I is the local acoustic intensity, and the brackets denote time average over one acoustic cycle.

To model thermal transport in perfused tissue containing one or more larger blood vessels, we adapt the widely used Pennes bioheat transfer equation[18] (BHTE), a form of the heat conduction equation with biologically relevant source and sink terms. The temperature field is split into two domains, perfused tissue and flowing blood, and can be expressed as

$$\rho_t C_t\frac{\partial T}{\partial t}=K_t\nabla^2 T-w_b C_b(T-T_\infty)+q \quad \text{(tissue domain)}, \tag{3}$$

$$\rho_b C_b\frac{\partial T}{\partial t}=K_b\nabla^2 T-\rho_b C_b(\vec{u}\cdot\nabla T)+q \quad \text{(blood domain)}, \tag{4}$$

where ρ, C, and K are the density, specific heat, and thermal conductivity with the subscripts t and b referring to tissue and blood domain; T_∞ refers to the temperature at large distances from the focus. w_b is the ''perfusion rate'' which is an average mass flow rate accounting in an approximate way for blood flow in capillary beds. The perfusion rate in our tissue phantoms was zero. Finally $\vec{u}$ is the blood flow velocity. The blood flow field has two components: a fully developed parabolic flow (the Reynolds number $\mathrm{Re}=2\rho r_0 U_0/\mu$ is less than 100 for all flow conditions studied here) plus an acoustic streaming flow. Thus, the total blood flow velocity field is written as

$$\vec{u}=\vec{u}_{ext}+\vec{u}_{str}=2U_0\left[1-\left(\frac{r}{r_0}\right)^2\right]+\vec{u}_{str}, \tag{5}$$

where $\vec{u}_{ext}$ represents parabolic Pouiseille flow and $\vec{u}_{str}$ is acoustic streaming, U_0 is the average velocity of the Pouiseille flow, μ is the shear viscosity, r is the radial distance from the flow axis, and r_0 is the radius of the vessel.

There is a considerable body of literature that develops the theoretical basis of acoustic streaming. The fluid motion is described by the continuity and Navier–Stokes (NS) equations where the driving force derives from the acoustic field and is manifested as a spatially dependent momentum transfer from sound waves to fluid motion.[39] Much of the early work (Eckart[40] and others) assumed continuous plane waves and a second-order approximation, which is unsuitable for HIFU beams. We employ a model that follows from the work of Kamakura *et al.*[41,42] in which we begin with the continuity and NS equations for a viscous incompressible fluid. The representation for the acoustic stress is accurate to second

order and the formulation includes full hydrodynamic nonlinearity. We consider only the axial component of the acoustic particle velocity field, which we obtain from the pressure solution [Eq. (1)] using a linear impedance relationship, which is described in Appendix A. The reader is referred to Ref. 31 for more information.

Unlike the pressure field, 3-D solutions for the BHTE and streaming equations are readily obtained, for the spatial grid and time steps required for the latter are considerably larger than the former. Thus coding for the heat conduction and streaming problems is fully 3-D, a feature that allows us to consider a host of insonation geometries in media with arbitrary variability

A. Numerical implementation

Determining the temperature rise due to acoustic absorption proceeds as a three-step process:

(1) Solve for the 2-D axisymmetric, steady-state pressure in the medium [Eq. (1)] based on the known parameters for the acoustic source, the propagation geometry, and material properties. The rate of ultrasonic energy deposition per unit volume q [Eq. (2)] and the driving force for acoustic streaming $\vec{F}$ (see Appendix A) are calculated.
(2) Incorporate the driving force $\vec{F}$ in the flow equations and solve for $\vec{u}_{str}$, the 3-D time-dependent acoustic streaming field in blood domain.
(3) The specified blood flow characteristics, the energy deposition term, and the streaming velocity field are then fed into the BHTE model, Eqs. (3)–(5), yielding the 3-D time-dependent temperature field computation. The model supports a time-varying blood flow velocity.

A finite-difference-time-domain (FDTD) simulation is used to calculate the acoustic pressure, the acoustic streaming, and the temperature as in Ref. 30. The FDTD method relies on discrete differences in place of partial derivatives in the model equations by dividing the spatial and time domains into discrete spatial grid points and discrete time steps.[43] We employ an explicit method where only known values from past time steps are required. The initial condition is the pressure at the surface of the focused bowl sound source and absorbing boundary conditions are employed at the edges of the computational domain. The time step for the acoustic and the BHTE/streaming calculations are 10^{-8} s or 10^{-5} ms and (typically) 0.5 ms, respectively. (The time step in the BHTE calculations was adaptive.) Grid point spacing for all three calculations (acoustic, thermal, and streaming) was maintained at typically 0.1 mm. Additional details regarding the coding and implementation of the model are provided in Ref. 31.

III. VALIDATION BY COMPARISON WITH EXPERIMENTS

A schematic of the measurement apparatus is shown in Fig. 1. The acoustic source and tissue phantom are immersed in filtered, deionized, and degassed water contained in a 58-cm-long, 43-cm-wide, and 46-cm-high acrylic tank, which is open to the atmosphere. A three-dimensional computer-controlled positioning system (not shown) is used to move the transducer along the beam axis and in both orthogonal directions. We employ an axially symmetric tissue phantom (10.72-cm diam, and 8-cm length) in which fluid flow is sustained in a cylindrical, wall-less flow channel (a cylindrical hole in the phantom formed by removing a rod which is originally part of the phantom mold) that is aligned parallel to the acoustic axis. ''Blood'' flow is created by gravity feed, and a flow control and monitor system is used to vary and stabilize the flow rate.

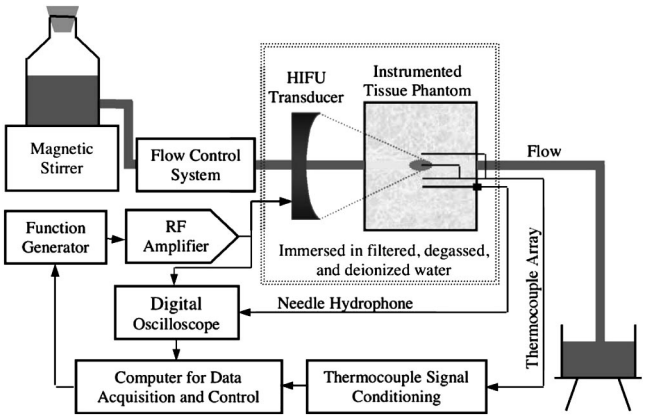


FIG. 1. Schematic diagram of the apparatus. The HIFU transducer has a hole in the center through which we feed the simulated vascular flow.

A PZT needle hydrophone (Dapco, 1.0 mm active diameter) is embedded in the phantom for *in situ* pressure measurements and position calibration. The hydrophone voltage is sampled at 100 MHz using a digital oscilloscope (LeCroy). The hydrophone was calibrated in water using a PVDF membrane hydrophone (Model 804, 0.6 mm active element, Perceptron, Hatboro, PA, mfg. calibration provided).

Embedded thermocouples (type E, bare junction, 125 μm diam, response time less than 40 ms, Omega Engineering Inc., Stamford, CT) monitor the temperature in the flow (central focus, upstream and downstream, and near wall) and in the outer ''tissue'' (near wall and further away). The thermocouple outputs are connected to an electronically compensated isothermal terminal block (TBX-1328, National Instruments, Austin, TX) and the terminal block outputs are connected to a signal-conditioning module (SCXI-1120, National Instruments, Austin, TX). The thermocouple voltages are low-pass filtered (BW=10 kHz), amplified, and multiplexed by the SCXI module, then sampled at 1 kHz by a data acquisition board (AT-MIO-16E-1, 12-bit resolution, 1.25 MS/s maximum sampling rate, National Instruments, Austin, TX), after which they are smoothed with a 20-pt moving

TABLE I. Measured material properties for the tissue phantom.

Physical property	Agar tissue phantom	Human tissue
Density (kg/m^3)	1045±20	1000–1100
Sound speed (m/s)	1550±15	1450–1640
Attenuation (Np/m·MHz)	10.2±0.30	4.03–17.27
Specific heat (J/kg·°C)	3700±200	3600–3890
Thermal conductivity (W/m·°C)	0.59±0.02	0.45–0.56

TABLE II. Measured material properties for the blood simulants.

Physical property	BMF1	BMF2	Human whole blood
Density (kg/m^3)	1108 ± 1	1036 ± 1	1052–1064
Sound Speed (m/s)	1704 ± 15	1549 ± 15	1540–1590
Attenuation (Np/m·MHz)	1.32 ± 0.20	0.46 ± 0.20	1.32–1.84
Specific Heat (J/kg·°C)	3450 ± 200	3930 ± 200	3600–3840
Thermal Conductivity (W/m·°C)	0.45 ± 0.02	0.57 ± 0.02	0.48–0.53
Viscosity (kg/s·m)	0.0037 ± 0.0001	0.0042 ± 0.0001	0.0035–0.0045

average. The function generator, the oscilloscope, and the processed thermocouple output are coupled to a computer so that we can control the source level, capture the *in situ* pressure, and monitor the temperature field.

The sound source is a single-element, spherically focused, piezoceramic transducer (Models H-102 and H-101, Sonic Concepts, Woodinville, WA), with a 20-mm-diam hole in the center (to allow tubing containing the blood-mimicking fluid flow to pass thru to facilitate coaxial flow and acoustics as indicated in Fig. 1), a focal length of 62.64 mm, an aperture of 70.0 mm, and a center frequency of 1.1 MHz. It was calibrated in water using the PVDF membrane. The agar-based phantom material is made according to a recipe adapted from Burlew *et al.*,[44] and is a mixture of water, agar, graphite powder (acts as a scatterer), methyl paraben (acts as a preservative), and 1-propanol (acts as sound speed tuning). The most appealing feature of this recipe is that the attenuation and sound speed can be varied by altering the concentrations of graphite and 1-propanol, respectively; there exists a nearly linear dependence on the weight percentage of each component. The measured[31] acoustical and thermal properties of the phantom and blood simulants are provided in Tables I and II. The quantities corresponding to ''typical'' literature values for human tissue and blood are also listed for reference.[45]

We employ two different suspensions to mimic human blood in the experiments. The first blood mimicking fluid (BMF1) employs a suspension of cellulose powder (MN301, Matherey and Nagel, Duren, Germany) in a glycerin/water mixture. This cellulose powder has particle sizes of 2–20 μm. A suspension containing 3.5% (by mass) cellulose in a glycerin/water mixture of ratio 9:10 at room temperature has the same dynamic viscosity at room temperature (4 cP) as whole blood at body temperature.[46] The advantage of this suspension is that cellulose powder is hydrophilic, and no surfactant is required to promote wetting; this makes the suspension easy to prepare and degas. Moreover, the attenuation coefficient of BMF1 (roughly 1.3 Np/m/Mhz)[31] is close to that of human whole blood. However, because cellulose powder is denser than the glycerin/water mixture, the particles settle very quickly. Consequently, the suspension needs to be continuously mixed using a magnetic stirrer—stirring is impractical in the flow system employed, particularly in the test volume. Therefore, BMF1 is not suitable for extended measurement periods, especially when the original vessel size is small and the flow rate is slow.

To mitigate the problem with particle settling, we employ a second blood mimicking fluid (BMF2) using a formulation originally developed for flow Doppler test phantoms.[47] Orgasol particles (Orgasol 2001 UD NAT, Atofina Chemicals Inc., Philadelphia, PA) have very fine size (5 μm) and a nominal density close to that of water (1030 kg/m^3). The particles are suspended in the glycerol/water mixture and the concentration of glycerol is adjusted to match the density of the fluid to that of the particles. Unlike cellulose powder, Orgasol particles are hydrophobic and a surfactant is employed to ensure wetability (Synperonic N, Trademark of ICI, PLC, BDH Laboratory Supplies, Poole, England). Finally, sodium azide (Sigma S-8032) serves as a preservative and dextran (Sigma D-4876) is added to adjust the viscosity. Though advantageous because of its neutral buoyancy, BMF2 possessed a lower absorption coefficient (roughly 0.5 Np/m/Mhz).

A. Computed and measured pressure fields in water and phantom material

Figure 2 shows pressure profiles along acoustic axis, and as a function of the radial distance from the acoustic axis (in

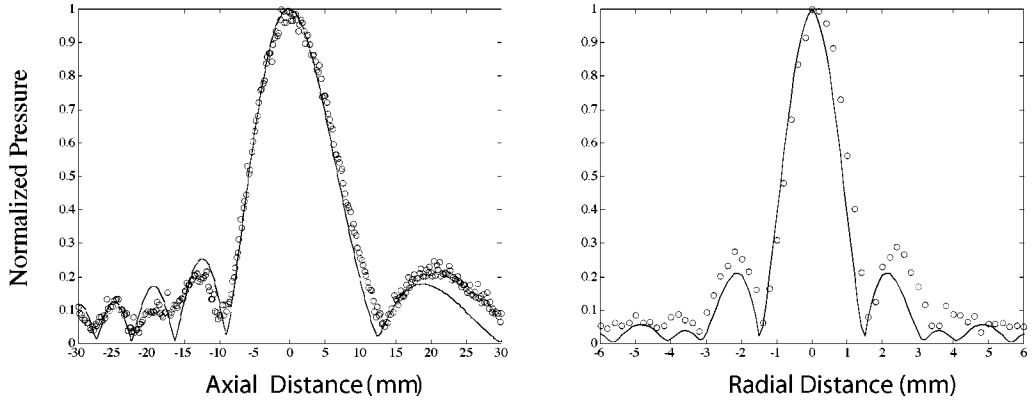


FIG. 2. Measured and computed pressure profiles in water at 1.0 MHz and 0.4 MPa peak negative focal pressure.

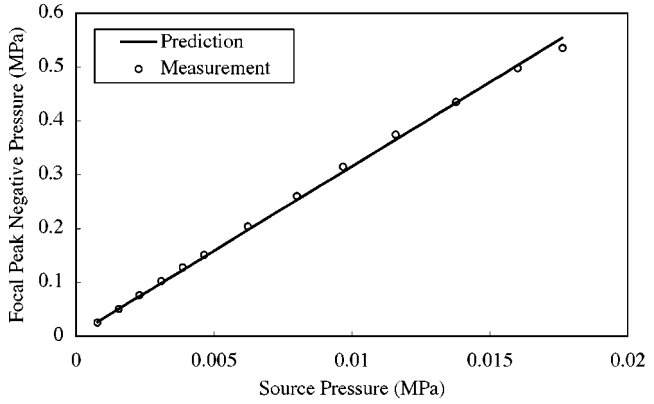


FIG. 3. A comparison of the measured (circles) and predicted (solid line) peak negative pressure at the focus in phantom as a function of source pressure. The measurement precision is estimated to be better than 1.5% and is too small to display on this scale.

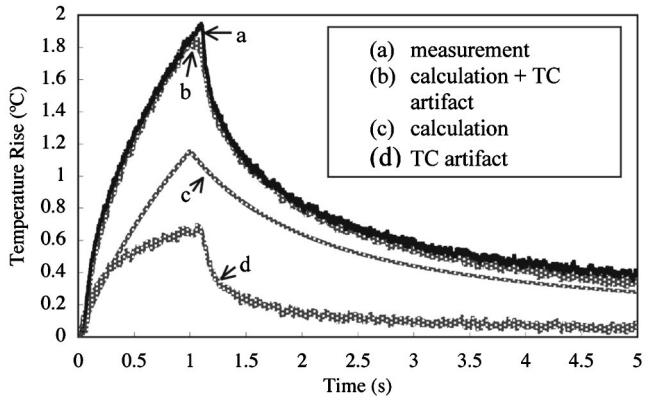


FIG. 4. A comparison of the measured and simulated focal temperature in a uniform phantom. (a) Output of thermocouple (TC) embedded at the acoustic focus in our usual tissue-mimicking Agar-graphite phantom. (b) Simulation result including the thermocouple artifact heating. (c) Simulation result using the measured acoustic and thermal properties of the agar-graphite phantom with no thermocouple artifact included. (d) The TC artifact: output of a TC independently measured in a separate experiment with the TC in the acoustic focus embedded in an low-absorbing Agar-only gel. The acoustic frequency, peak negative focal pressure, and exposure time was 1.0 MHz, 1.11 MPa, and 1 s for all data.

the focal plane); solid lines and open circles correspond to predictions and measurements, respectively. These results were obtained in water at 30 °C using the PVDF membrane hydrophone. The pressures shown are all peak negative quantities normalized to a focal pressure of 0.40 MPa. Good agreement is found in the focal region and along the side lobes in the focal plane. Since power deposition is quadratic in the pressure, deviations between measurement and model in the low-amplitude regions are not deemed significant.

Although a full-field field pressure measurement cannot be done in the phantom, we can nevertheless compare predicted and measured (*in situ*) focal pressures. Figure 3 shows the measured (circles) and computed (solid line) peak negative pressure at the focus as a function of source pressure (peak negative pressure at the face of the transducer). The source pressure resulted from the acoustic pressure calibration method described by Huang[31] and the measurements were obtained with the calibrated needle hydrophone. We show good agreement between the measured and predicted values; all deviations are less than 5% (0.4 dB).

B. Computed and measured temperature fields in a uniform phantom

As an initial test of the BHTE model, we constructed a solid agar-graphite phantom with no vascularity and measured the focal temperature as a function of time. Results shown in Fig. 4 exhibit the characteristic heating and cooling curves; however, there is a significant difference between the measured (a) and calculated (c) responses, owing to the well-known ''thermocouple artifact'' effect, which is caused by enhanced heating in the viscous boundary layer adjacent to the thermocouple surface.[48] We are able to independently measure this effect by embedding the thermocouple in a low-absorption material (either agar alone when the thermocouple is in the phantom region, or particle-free BMF when the thermocouple is in the flow region) so that the measured temperature rise is solely the result of the boundary-layer viscous heating at the thermocouple—line (d) in Fig. 4. We then corrected the calculation to account for this artifactual heating (b), and the resulting agreement between model and measurement is quite good. We stress that the artifactual heating was measured independently—we do not employ any fitting parameters

C. Computed and measured temperature fields in a flow phantom

We next equip the tissue phantom with a 6.35-mm-diam wall-less ''vessel'' positioned parallel to the acoustic axis (Fig. 1). The vessel is constructed by molding the phantom around a solid brass rod, which was subsequently removed. The blood simulant used is BMF1. A thermocouple is imbedded in the flow center and 35 mm from the surface of the phantom. The acoustic focus is positioned directly on the thermocouple. There is no externally applied flow. The ob-

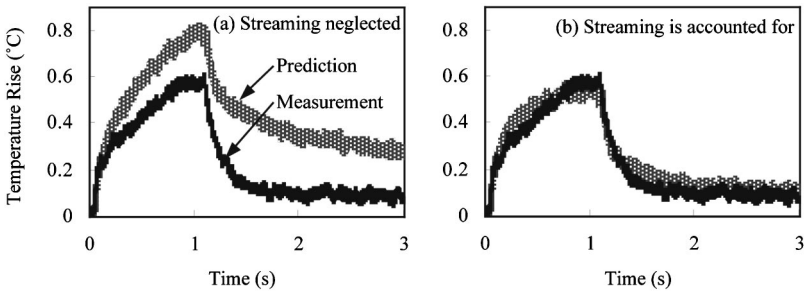


FIG. 5. The computed and measured temperature response for vessel 6.5 mm in diameter and a 1-s insonation [BMF1, 1 MHz frequency and peak negative focal pressure of 1.52 MPa (Spatial peak temporal peak Intensity $=6.1\times10^5$ W/m^2]. (a) Shows the predicted response without acoustic streaming which is about 30% above the measured response; (b) shows the prediction when streaming is accounted for.

jective is to assess the contributions of acoustic streaming as well as to validate the code when used to model axisymmetric configurations.

Figure 5 shows the computed and measured temperature response for 1-s insonation when the acoustic streaming is neglected in the calculation (a) and when the acoustic streaming is accounted for (b). The frequency and peak negative focal pressure are 1.0 MHz and 1.52 MPa, respectively. Note the excellent agreement between the measured and computed temperature profiles when the effect of acoustic streaming is included. (The aforementioned thermocouple artifact correction was applied to the calculation.) The result shows that acoustic streaming present during HIFU insonation leads to significant cooling at the focus; an overestimation (33% in this case) of the heating of the ''blood'' ensues when acoustic streaming is neglected. Although there is no externally applied flow, the maximum flow velocity along the transducer axis can still reach as high as 4.9 cm/s (predicted value) for the pressure amplitude of 1.52 MPa at the end of 1 s. The calculated streaming profile around the focal region is given in Fig. 6. For this axisymmetric case, streaming is illustrated only in the half-space. The direction of the arrow indicates the direction of the velocity, and the size of the arrow denotes the magnitude of the velocity. The plot exhibits a local circulation around the focus, with the maximum velocity in postfocal region.

We then equip another tissue phantom with a 2.6-mm-diam vessel positioned parallel to the acoustic axis. A thermocouple is embedded in the phantom material, 0.4 mm from the vessel wall and in the focal plane of the transducer. The acoustic focus is positioned on the thermocouple. The blood simulant is BMF2. The objective is to assess the contributions of convective cooling as well as to validate the code when used to model non-axisymmetric configurations. Figure 7 shows the computed and measured temperature response for insonation times of 1, 3, and 5 s and for imposed mean flow velocities of 0 cm/sec and 1.87 cm/sec. The frequency and peak negative focal pressure are 1.0 MHz and 1.45 MPa, respectively. For the case of no flow the peak temperature rise increased from 3.5° to 7° with increasing

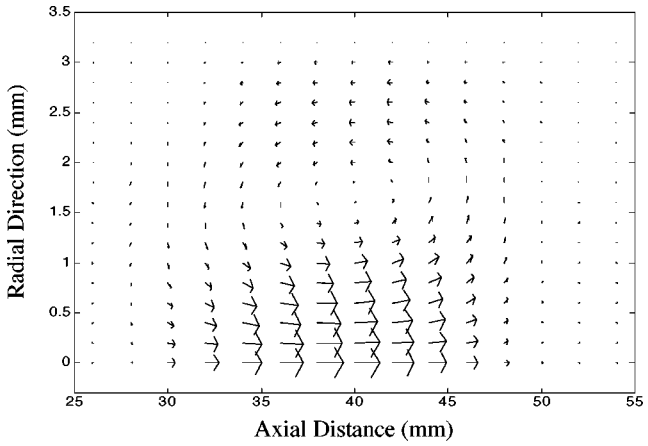


FIG. 6. Simulated streaming profile around the focal region inside the 6.5 mm vessel. Axial distances are measured from the surface of the tissue phantom. Focal plane is at 35 mm. The length of the arrow denotes flow speed (the maximum speed indicated is 4.93 cm/s) and the acoustic frequency and peak negative focal pressure were 1.0 MHz and 1.52 MPa, respectively. BMF1 properties were used.

insonation time. When flow was present the same trend with increasing insonation time was observed, but the peak values at the end of each insonation were slightly lower than the no-flow counterparts. Longer insonation times exacerbated the difference between the flow and no-flow peak temperatures at all pressures. It appears that convective heat transfer owing to externally imposed flow alone has a small, but measurable, impact on the temperature elevation in the phantom material for these time scales, which are short with respect to thermal conduction owing to diffusion alone. For much longer insonation times we expect the variation to become larger. Note also the excellent agreement between the measured and computed temperature profiles.

The comparison of measured and predicted peak temperature rise at the focus as a function of pressure with varying flow speed for a 5-s insonation time is given in Fig. 8. The focus is again positioned just outside the vessel as in the case of Fig. 7. Each data point is the average of five sequential measurements, minus the thermocouple artifact. The

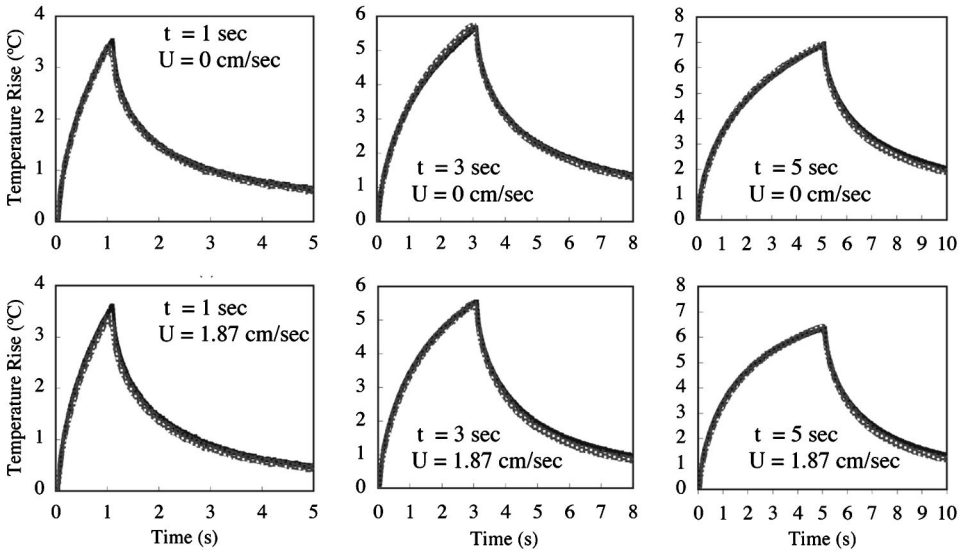


FIG. 7. Measured and predicted temperature rise versus time at the focus (positioned in phantom material and 0.4 mm from the vessel wall) for a 2.6 cm diameter vessel, for differing insonation times and imposed flow velocities in BMF2. The acoustic frequency and peak focal pressure was 1.0 MHz and 1.45 MPa (spatial peak temporal peak Intensity=6.5 $\times 10^5$ W/m^2), respectively. Dark line—data; light line—prediction.

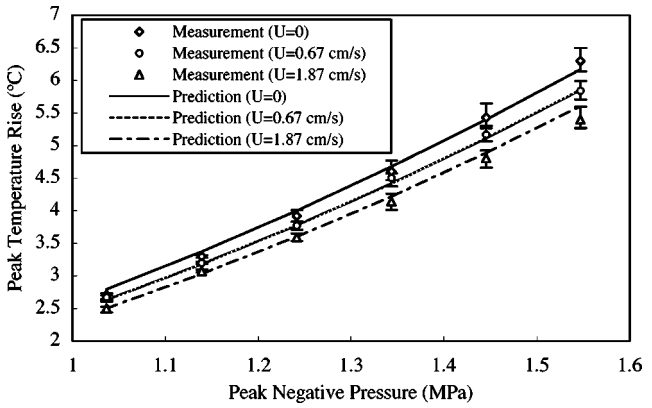


FIG. 8. Comparison of the measured and predicted peak temperature rise at the focus (positioned in phantom material and 0.4 mm from the vessel wall) as a function of pressure for varying external applied flow speeds in BMF2. The vessel diameter is 2.6 mm. The insonation time is 5 s. Each point is an average of five measurements and the error bars are the maximum deviation from the mean.

phantom is allowed to cool completely between measurements and the error bars indicate the maximum deviation from the mean. Good agreement is found for mean flow velocity of 0, 0.67, and 1.87 cm/s within the estimated uncertainties in temperature measurement and model prediction.[31] These results offer strong evidence of both the accuracy and precision of the model, even when applied to nonaxisymmetric insonation arrangements.

IV. SUMMARY AND CONCLUSIONS

A finite difference time domain code was developed to predict HIFU-induced pressure and temperature fields in absorbing media with flow. The pressure model is 2-D axisymmetric and the temperature field model is 3-D. Three-dimensional acoustic streaming was included in the simulation. Simulation performance was demonstrated by a quantitative comparison with experiment results obtained using an instrumented agar and graphite phantom equipped with a wall-less flow channel designed to simulate the convective cooling effects of a large blood vessel. Good agreement was demonstrated between simulations and measurements once the thermocouple artifact was accounted for. The model supports relatively fast 3-D temperature simulations for complex insonation geometries.

The results show that convective cooling, due to both blood flow and streaming induced by HIFU, reduces the temperature rise in tissue near a large vessel. This demonstrated reduction in temperature rise associated with both phenomena illustrates the importance of including vascular convective effects in any model of HIFU heating in real tissue. The present model will yield valid predictions in real tissue as long as the assumptions in the model are not violated. Thus, the model should apply in all soft tissues (bone and cartilage would violate the weak scattering assumption) for which the acoustic and thermal properties are well known at the frequencies of interest, and for which the propagation nonlinearity is minimal. The range of the model's usefulness extends until pressures above which the cavitation nucleation threshold is exceeded, at which point bubble-enhanced heating must be included as a source term in Eqs. (3) and (4), and bubbly liquid propagation effects must also be considered.[33,35,36]

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Dr. Ibrahim Hallaj in setting up the numerical model, help from Dr. Patrick Edson in the initial stages of experimentation, and helpful comments from Professor Lawrence Crum and Professor Kullervo Hynynen. We also acknowledge the financial support of DARPA (subcontract through the University of Washington) and the U.S. Army (subcontract from the National Center for Physical Acoustics).

APPENDIX A: ACOUSTIC STREAMING MODEL

After Kamakura *et al.*,[36] we begin with the continuity equation and the Navier–Stokes equation in a viscous incompressible fluid:

$$\nabla \cdot \vec{U} = 0, \tag{A1}$$

$$\frac{\partial \vec{U}}{\partial t} + (\vec{U} \cdot \nabla)\vec{U} - \frac{\mu}{\rho_0}\nabla^2 \vec{U} = -\frac{1}{\rho_0}\nabla P + \frac{1}{\rho_0}\vec{F}, \tag{A2}$$

where $\vec{U}$ is the streaming velocity, P is the static pressure, μ is the shear viscosity of the medium, and ρ_0 is the ambient density. The first two terms in Eq. (A2) describe the acceleration of the fluid, the third term accounts for viscous stress, and the fourth term captures the stress due to pressure variation in the fluid. $\vec{F}$ in the last term is the force acting on the fluid by the presence of sound. The forcing function due to sound can be written as[49]

$$\vec{F} = -\rho_0 \langle (\vec{u}_1 \cdot \nabla)\vec{u}_1 + \vec{u}_1 (\nabla \cdot \vec{u}_1) \rangle, \tag{A3}$$

where $\vec{u}_1$ is the first-order approximation for the acoustic particle velocity. The brackets indicate an average over time, whose interval is much shorter than the transient time of streaming and much longer than the acoustic period.

Since our acoustic model gives the space- and time-dependent pressure field, we will cast Eq. (A3) in terms of acoustic pressure rather than particle velocity. The axial component of the acoustic particle velocity in an ultrasonic beam is generally much larger than the radial component. This fact also holds true for the driving force of acoustic streaming. If we neglect the radial component of acoustic particle motion, then the force in the acoustic axis direction can be written as follows (taking z as the axis of the beam),

$$F_z = -\rho_0 \left\langle 2u_{1z}\frac{\partial u_{1z}}{\partial z} \right\rangle = -\rho_0 \left\langle \frac{\partial u_{1z}^2}{\partial z} \right\rangle = -\rho_0 \frac{\partial}{\partial z}\langle u_{1z}^2 \rangle. \tag{A4}$$

Using the plane wave impedance relation between the sound pressure and the particle velocity, $p_1 = \rho_0 c_0 u_1$, where p_1 is the first-order approximation for the acoustic pressure, Eq. (A4) can be rewritten as

$$F_z = -\frac{1}{\rho_0 c_0^2}\frac{\partial}{\partial z}\langle p_1^2 \rangle = -\frac{1}{c_0}\frac{\partial I}{\partial z} = \frac{2\alpha_{ABS}}{c_0} I, \tag{A5}$$

where I is the intensity in the direction of propagation, and $\partial I/\partial z = -2\alpha_{ABS} I$ for a linear plane wave approximation, where α_{ABS} is the pressure absorption coefficient of the medium. For a time-harmonic wave the intensity is given by

$$I = \frac{1}{\rho_0 c_0}\langle p_1^2 \rangle = \frac{1}{\omega^2 \rho_0 c_0}\left\langle \left(\frac{\partial p_1}{\partial t} \right)^2 \right\rangle. \tag{A6}$$

Equation (A5) may thus be solved simultaneously with the Westervelt equation (1). The resulting streaming velocity used in Eq. (5) is numerically obtained from combining Eq. (A5) with Eqs. (A1) and (A2).

[1] G. ter Haar, "Ultrasound focal beam surgery," Ultrasound Med. Biol. **21**(9), 1089–1100 (1995).
[2] S. Vaezy, R. Martin, P. Mourad, and L. Crum, "Review: Hemostasis using high intensity focused ultrasound," Eur. J. Ultrasound **9**, 79–87 (1999).
[3] N. T. Sanghvi, K. Hynynen, and F. L. Lizzi, "New developments in therapeutic ultrasound," IEEE Eng. Med. Biol. Mag. **15**(6), 83–92 (1996).
[4] G. ter Haar, "Acoustic surgery," Phys. Today **54**(12), 29–34 (2001).
[5] W. J. Fry and R. B. Fry, "Determination of absolute sound levels and acoustic absorption coefficients by thermocouple probes—theory," J. Acoust. Soc. Am. **26**, 311–317 (1950).
[6] P. P. Lele, "A simple method for the production of trackless focal lesions with focused ultrasound: Physical factors," J. Physiol. **160**, 494–512 (1962).
[7] K. J. Parker, "The thermal pulse decay technique for measuring ultrasonic absorption coefficients," J. Acoust. Soc. Am. **74**, 1356–1361 (1983).
[8] K. J. Parker, "Effects of heat conduction and sample size on ultrasonic absorption measurements," J. Acoust. Soc. Am. **77**, 719–725 (1985).
[9] J. B. Pond, "The role of heat in the production of ultrasonic focal lesions," J. Acoust. Soc. Am. **47**, 1607–1611 (1970).
[10] T. C. Robinson and P. P. Lele, "An analysis of lesion development in the brain and in plastics by high-intensity focused ultrasound at low-megahertz frequencies," J. Acoust. Soc. Am. **51**, 1333–1351 (1972).
[11] T. G. Muir and E. L. Carstensen, "Prediction of nonlinear acoustic effects at biomedical frequencies and intensities," Ultrasound Med. Biol. **6**(4), 345–357 (1980).
[12] W. L. Nyborg, "Heat generation by ultrasound in a relaxing medium," J. Acoust. Soc. Am. **70**, 310–312 (1981).
[13] K. Beissner, "On the plane-wave approximation of acoustic intensity," J. Acoust. Soc. Am. **71**, 1406–1411 (1982).
[14] T. J. Cavicchi and W. D. O'Brien, Jr., "Heat generated by ultrasound in an absorbing medium," J. Acoust. Soc. Am. **76**, 1244–1245 (1984).
[15] H. D. Mair, D. A. Hutchins, and P. A. Puhach, "Intensity fields of continuous-wave axisymmetric transducers," J. Acoust. Soc. Am. **81**, 328–334 (1987).
[16] F. L. Lizzi, J. Driller, and M. Ostromogilsky, "Thermal model for ultrasound treatment of glaucoma," Ultrasound Med. Biol. **10**(3), 289–298 (1984).
[17] F. L. Lizzi, J. Driller, B. Lunzer, A. Kalisz, and D. J. Coleman, "Computer model of ultrasound hyperthermia and ablation of tissue volumes using high intensity focused ultrasound," Ultrasound Med. Biol. **18**(1), 59–73 (1992).
[18] H. H. Pennes, "Analysis of tissue and arterial blood temperatures in the resting human forearm," J. Appl. Physiol. **2**, 93–122 (1948).
[19] C. R. Hill, I. H. Rivens, M. G. Vaughan, and G. ter Haar, "Lesion development in focused ultrasound surgery: a general model," Ultrasound Med. Biol. **20**(3), 259–269 (1994).
[20] J. Wu and G. Du, "Temperature elevation generated by a focused Gaussian beam of ultrasound," Ultrasound Med. Biol. **16**(5), 489–498 (1990).
[21] M. C. Kolios, M. D. Sherar, and J. W. Hunt, "Blood flow cooling and ultrasonic lesion formation," Med. Phys. **23**, 1287–1298 (1996).
[22] F. P. Curra, P. D. Mourad, V. A. Khokhlova, R. O. Cleveland, and L. A. Crum, "Numerical simulations of heating patterns and tissue temperature response due to high-intensity focused ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**(4), 1077–1088 (2000).
[23] X. Fan and K. Hynynen, "A study of various parameters of spherically curved phased arrays for noninvasive ultrasound surgery," Phys. Med. Biol. **41**(4), 591–608 (1996).
[24] J. Hoffelner, H. Landes, M. Kaltenbacher, and R. Lerch, "Finite element simulation of nonlinear wave propagation in thermoviscous fluids including dissipation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**(3), 779–786 (2001).
[25] H. Wan, P. VanBaren, E. S. Ebbini, and C. A. Cain, "Ultrasound surgery: comparison of strategies using phased array systems," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**(6), 1085–1098 (1996).
[26] P. M. Meaney, R. L. Clarke, G. R. ter Haar, and I. H. Rivens, "A 3-D finite-element model for computation of temperature profiles and regions of thermal damage during focused ultrasound surgery exposures," Ultrasound Med. Biol. **24**(9), 1489–1499 (1998).
[27] B. Krasovitski and E. Kimmel, "A blood vessel exposed to ultrasound: A mathematical simulation of the temperature field," J. Acoust. Soc. Am. **111**, 1454–1459 (2002).
[28] M. R. Bailey, V. A. Khokhlova, O. A. Sapozhnikov, S. G. Kargl, and L. A. Crum, "Physical mechanisms of the therapeutic effect of ultrasound (A review)," Acoust. Phys. **49**(4), 369–388 (2003).
[29] I. M. Hallaj, "Nonlinear Acoustics in Underwater and Biomedical Applications: Array Performance Degradation and Time Reversal Invariance," Ph.D. dissertation, University of Washington, 1999.
[30] I. M. Hallaj and R. O. Cleveland, "FDTD simulation of finite-amplitude pressure and temperature fields for biomedical ultrasound," J. Acoust. Soc. Am. **105**, L7–L12 (1999).
[31] J. Huang, "Heating in Vascular and Flow-Through Tissue Phantoms Induced by Focused Ultrasound," Ph.D. dissertation, Boston University, 2002.
[32] M. F. Hamilton and C. L. Morfey, "Model equations," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York, 1998), Chap. 3.
[33] F. Chavrier, J. Y. Chapelon, A. Gelet, and D. Cathignol, "Modeling of high-intensity focused ultrasound-induced lesions in the presence of cavitation bubbles," J. Acoust. Soc. Am. **108**, 432–440 (2000).
[34] P. L. Edson, "The Role of Acoustic Cavitation in Enhanced Ultrasound-Induced Heating in a Tissue-Mimicking Phantom," Ph.D. dissertation, Boston University, 2001.
[35] R. G. Holt and R. A. Roy, "Measurements of bubble-enhanced heating from focused, MHz-frequency ultrasound in a tissue-mimicking material," Ultrasound Med. Biol. **27**, 1399–1412 (2001).
[36] X. Yang, "Investigation of Bubble Dynamics and Heating During Focused Ultrasound Insonation in Tissue-mimicking Materials," Ph.D. thesis, Boston University, 2003.
[37] A. D. Pierce, *Acoustics, An introduction to its physical principles and applications* (McGraw-Hill, New York, 1981), Chap. 10.
[38] W. L. Nyborg, "Sonically produced heat in a fluid with bulk viscosity and shear viscosity," J. Acoust. Soc. Am. **80**, 1133–1139 (1986).
[39] O. V. Rudenko and S. I. Soluyan, *Theoretical Foundations of Nonlinear Acoustics* (Plenum, New York, 1977), Chap. 8.
[40] C. Eckart, "Vortices and streams caused by sound waves," Phys. Rev. **73**, 68–76 (1948).
[41] T. Kamakura, M. Matsuda, Y. Kumamoto, and M. A. Breazeale, "Acoustic streaming induced in focused Gaussian beams," J. Acoust. Soc. Am. **97**, 2740–2746 (1995).
[42] K. Matsuda, T. Kamakura, and Y. Kumamoto, "Buildup of acoustic streaming in focused beams," Ultrasonics **34**, 763–765 (1996).
[43] C. A. J. Fletcher, *Computational Techniques for Fluid Dynamics*, 2nd ed., Vol. 1 of Springer series in computational physics (Spring-Verlag, Berlin, 1991).
[44] M. M. Burlew, E. L. Madsen, J. A. Zagzebski, R. A. Banjavic, and S. W. Sum, "A new ultrasound tissue-equivalent material," Rad. Phys. **134**, 517–520 (1980).
[45] F. A. Duck, *Physical Properties of Tissue* (Academic, New York, 1990).
[46] J. Petrick, M. Zomack, and R. Schlief, "An investigation of the relationship between ultrasound echo enhancement and Doppler frequency shift using a pulsatile arterial flow phantom," Invest. Radiol. **32**(4), 225–235 (1997).
[47] J. Lubbers, "Application of a new blood-mimicking fluid in a flow Doppler test objects," Eur. J. Radiol. **9**, 267–276 (1999).
[48] K. Hynynen, C. J. Martin, D. J. Watmough, and J. R. Mallard, "Errors in temperature measurement by thermocouple probes during ultrasound induced hyperthermia," Br. J. Radiol. **56**, 968 (1983).
[49] W. L. Nyborg, "Acoustic Streaming," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York, 1998), Chap. 7.

Geographic variation and acoustic structure of the underwater vocalization of harbor seal (*Phoca vitulina*) in Norway, Sweden and Scotland

Anders Bjørgesæter and Karl Inne Ugland
Department of Marine Biology and Limnology, University of Oslo, P.O. Box 1064, Blindern, 0316 Oslo, Norway

Arne Bjørge
Institute of Marine Research, P.O. Box 1870, Nordnes, 5817 Bergen, Norway

(Received 21 November 2003; revised 23 June 2004; accepted 24 June 2004)

The male harbor seal (*Phoca vitulina*) produces broadband nonharmonic vocalizations underwater during the breeding season. In total, 120 vocalizations from six colonies were analyzed to provide a description of the acoustic structure and for the presence of geographic variation. The complex harbor seal vocalizations may be described by how the frequency bandwidth varies over time. An algorithm that identifies the boundaries between noise and signal from digital spectrograms was developed in order to extract a frequency bandwidth contour. The contours were used as inputs for multivariate analysis. The vocalizations' sound types (e.g., pulsed sound, whistle, and broadband nonharmonic sound) were determined by comparing the vocalizations' spectrographic representations with sound waves produced by known sound sources. Comparison between colonies revealed differences in the frequency contours, as well as some geographical variation in use of sound types. The vocal differences may reflect a limited exchange of individuals between the six colonies due to long distances and strong site fidelity. Geographically different vocal repertoires have potential for identifying discrete breeding colonies of harbor seals, but more information is needed on the nature and extent of early movements of young, the degree of learning, and the stability of the vocal repertoire. A characteristic feature of many vocalizations in this study was the presence of tonal-like introductory phrases that fit into the categories pulsed sound and whistles. The functions of these phrases are unknown but may be important in distance perception and localization of the sound source. The potential behavioral consequences of the observed variability may be indicative of adaptations to different environmental properties influencing determination of distance and direction and plausible different male mating tactics.
[DOI: 10.1121/1.1782933]

PACS numbers: 43.80.Ka, 43.80.Lb, 43.80.Nd [WWL] Pages: 2459–2468

I. INTRODUCTION

The harbor seal (*Phoca vitulina*) is an aquatic-mating phocid with a mating system generally considered to include serial monogamy and lek-type system (Fisher, 1954; Bigg, 1981; Boness, Coltman *et al*. 1997; 1991; Hanggi and Schusterman, 1994; Thompson *et al.*, 1994; Van Parijs *et al.*, 1997, 1999, 2000b). During the mating season, from July to late August, male harbor seals restrict their home range and start spending much of their time in the water at particular sites where they perform short stereotypic dives described as display activity (Bjørge, 1995; Van Parijs *et al.*, 1997, 1999, 2000a, 2000b; Hayes *et al.*, 2004). Favorable display sites are generally located in areas where female encounter rate is particular high (e.g., close to female haul-out and pupping sites), but males have been shown to display also over a wider area covering the whole of the female distribution (Van Parijs *et al.*, 1997, 1999, 2000a). The short, stereotypic dives may be repeated for periods up to 7 hours and males appear to show strong site fidelity towards display sites (Bjørge *et al.*, 1995; 2002; Van Parijs *et al.*, 2000b). During each dive, the male emits between one and five loud vocalizations (Hanggi and Schusterman, 1994; Bjørge *et al.*, 1995; Van Parijs *et al.*, 1997, 1999, 2000a; Hayes *et al.*, 2004; see Fig. 1). The best-known harbor seal vocalization is a broadband, nonharmonic roar with energy in the frequency range between 50 and 4000 Hz. The roar is the only harbor seal vocalization that is reported from all studied areas, i.e., USA (Hanggi and Schusterman, 1994; Hayes *et al.*, 2004), Norway (Bjørge *et al.*, 1995), Sweden (Wahlberg *et al.*, 2002), Scotland, U.K. (Van Parijs *et al.*, 1997, 2000a), and Canada (Van Parijs *et al.*, 2002; 2003). The function of the roar vocalization appears primarily to be advertising the presence of a male in breeding condition, and has therefore been suggested to be used in male–male competition and/or as reproductive advertisement display to attract females (Hanggi and Schusterman, 1994; Van Parijs *et al.*, 1997, 1999, 2000a, 2000b; Hayes *et al.*, 2004). Similar sexual display behavior is observed in other male pinnipeds such as walruses (*Odobenus rosmarus*) (Ray and Watkins, 1975; Stirling *et al.*, 1983, 1987), bearded seals (*Erignathus barbatus*) (Ray *et al.*, 1969; Cleator *et al.*, 1989), and Weddell seals (*Leptonychotes weddelli*) (Thomas and Kuechle, 1982; Thomas and Stirling, 1983) seals.

Recent research has shown that vocal variation in harbor

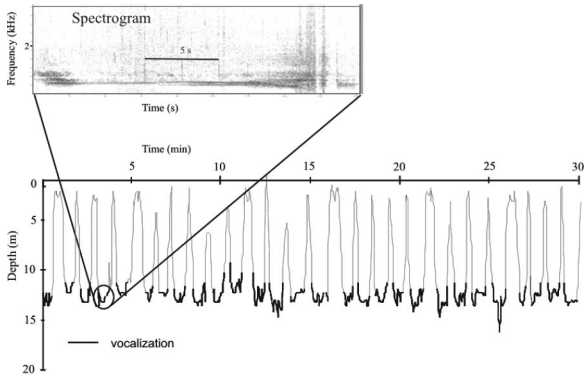


FIG. 1. Dive profiles and a spectrogram of an adult male harbor seal displaying and vocalizing, 23 July 1995 at Eynhallow, the Orkney Islands, Scotland, U.K.

seal roars occurs at the oceanic, regional, population, and subpopulation level (Van Parijs *et al.*, 1999, 2003), as well as individual variation in temporal (Van Parijs *et al.*, 2000a) and spectral features (Hanggi and Schusterman, 1994; Van Parijs *et al.*, 2000a). Factors that appear to have been important in the development of vocal geographic differences in harbor seals and other pinnipeds (Thomas and Stirling, 1983; Thomas *et al.*, 1988; Cleator *et al.*, 1989; Terhune, 1994; Van Parijs *et al.*, 2000a; 2003) includes: (1) long distances between recording sites; (2) a strong fidelity to specific breeding sites; (3) vocal learning; (4) a polygynous mating system; (5) different acoustic transmission properties; and (6) adoptions to various environmental challenges that influence male mating strategy.

To investigate individual and geographical differences, Hanggi and Schusterman (1994) and Van Parijs *et al.* (2000a) identified specific frequency boundaries within the frequency range of the roars. The boundaries (e.g., max and min values) were read off directly from spectrograms, by evaluating the darker and lighter shades that represent the intensity levels. Although the roar's frequency bandwidth may be reliable identified by individual researchers using this method, the technique is subjective and will probably generalize poorly across researchers. Moreover, computer-based spectrographic screen measurements are still manual and do not take advantage of the possibilities that digital spectrograms offer. Using a computer not only reduces human judgments but makes it less cumbersome to extract a large number of variables from each vocalization. Analytic techniques that utilize the underlying numerical intensity levels have recently been explored (Buck and Tyack, 1993; McCowan, 1995; Murray *et al.*, 1998). Buck and Tyack (1993) developed an algorithm that extracted the frequency component with the most energy to obtain a time-varying pitch contour of bottlenose (*Tursiops truncatus*) whistle. Similarly, McCowan (1995) and Murray *et al.* (1998) characterized bottlenose and false killer whale (*Pseudorca crassidens*) vocalizations by 20 and 30 measurements of the peak frequency, as well as changes in the duty cycle (Murray *et al.*, 1998). These techniques ensure an objective description of the vocalizations and make it easier to compare results from different studies. However, in nonperiodic signals, such as the harbor seal roar, dominant frequency measurements are not suitable because the peak frequency will vary randomly across the time axis (in accordance with the random nature of such signals). Clark *et al.* (1987) generated an algorithm that used the entire spectrogram of swamp sparrow (*Melospiza georgiana*) songs, rather than an extracted contour. Similarity between two songs was computed by cross correlating the numerical intensity levels along the time axis, and the resulting peak value of the correlation expressed the extent of how well the intensity levels in two songs overlap each other. This algorithm cannot describe the vocalizations themselves, only the differences between particular pairs of vocalizations.

A. Study objectives

The first aim of the present study was to investigate patterns of vocal geographical variation in six colonies in the Northeast Atlantic by means of an algorithm for extracting frequency bandwidth contours of harbor seal vocalizations. The second aim of this study was to provide a qualitative description of the detailed acoustic structure of the roar produced by harbor seals, and classify it into traditional sound-types categories such as nonharmonic sounds, pulsed sounds, and whistles.

II. DATA COLLECTION

Vocalizations were recorded at six harbor seal colonies in the Northeast Atlantic (Koster in Sweden; Eynhallow in Scotland; Sandøy, Froan, and Kongsfjord in Norway (Fig. 2) during the mating season of 1995 and 1996. Each study site holds the major seal colony of the respective area, and comprises coastal archipelagos with numerous intertidal rocks, small islets, and islands serving as haul-out sites during the breeding season. The diel and tidal cycle varies inversely from south to north. Koster (site 1, Fig. 2) has the smallest tidal amplitude (30 cm), and the sun is down for almost 5 h, while Kongsfjord (site 5, Fig. 2) has the largest tidal amplitude (200 cm) and 24-h daylight during summer. At Sandøy, and partly Nordmjelde (site 3 and 4, Fig. 2), haul-out sites are distributed over large areas, whereas in Eynhallow, and especially Kongsfjord (site 5 and 6; Fig. 2), the seals appear to prefer one or two haul-out sites, and are therefore less dispersed. Koster and Froan (site 1 and 3) may be considered as intermediate regarding the distribution of suitable haul-out sites (for more details on localities see Roen *et al.*, 1994 [Koster, Froan, and Kongsfjord], Bjørge *et al.*, 2002b [Sandøy], Bjørge *et al.*, 1995 [Froan], Wiig, 1988 [Nordmjelde], Henriksen and Haug, 1994 [Kongsfjord], and Van Parijs *et al.*, 1997 [Eynhallow]).

Recordings were made from small boats, either anchored or drifting close to calling animals. The hydrophone was fixed beneath a spar boy to reduce vertical motion, and lowered 1 to 3 m below the surface. A recording session started when we heard vocalizations of good quality (based on our subjective impression) and lasted from a few minutes to 4 h. In 1995, a custom-built hydrophone (frequency response 0.02 to 70.00 kHz) and amplifier (Sea Mammal Research Unit, University of St-Andrews, Scotland, UK) was used, and in 1996 a Brüel & Kjær 8104 hydrophone (frequency response $0.01-75.00\,\text{kHz}\pm3.0\,\text{dB}$) was used. A

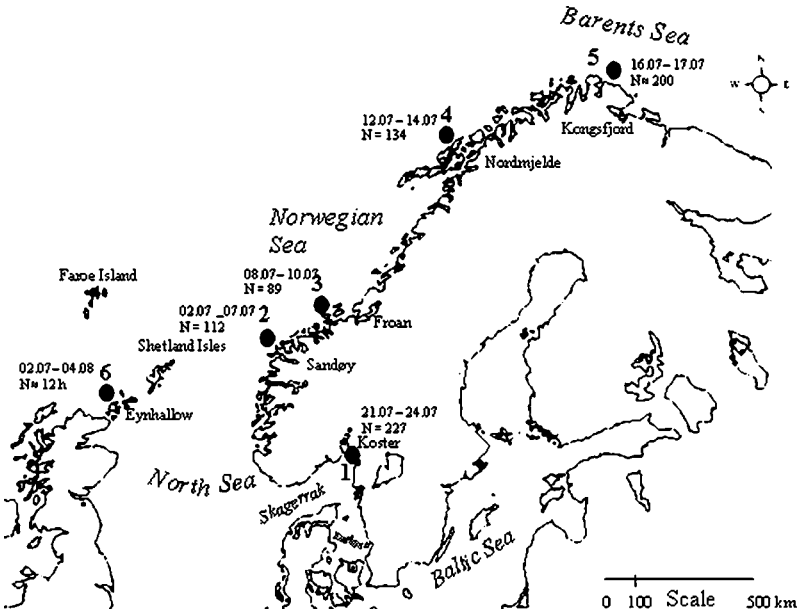


FIG. 2. Map showing the recording sites at six areas of the Northeast Atlantic: (1) Koster in Bohuslän County on the Swedish west coast; (2) Sandøy in Møre County on the Norwegian west coast; (3) Froan in Sør-Trøndelag County on the Norwegian west coast; (4) Nordmjelde in Nordland County in Northern Norway; (5) Kongsfjord in Finnmark County in Northern Norway; (6) Eynhallow in northwest Orkneys in northern Scotland, U.K. Numbers beside sites represent recording dates.

Sony digital audio tape recorder (DAT), TCD-D7, was used for all recordings. The frequency response of the recording equipment was limited by the tape recorder, which uses a sampling rate of 44.1 kHz, for a frequency bandwidth to 22 kHz. This bandwidth is well within the frequency range of harbor seal vocalizations (<5 kHz) reported by Hanggi and Schusterman (1994).

From each colony, 20 vocalizations with a high signal-to-noise ratio were chosen for analysis, yielding a data set of $6\times20=120$ vocalizations. To maximize the breadth of the sample at each site, recordings were made over a radius of at least 20 km at each colony (e.g., recorded 15–60 min and moving the boat to a new position several kilometers away). The recording positions were fixed with a GPS receiver (Garmin GPS 45), and if drifting, start- and endpoint, as well as intermediate readings were taken. The positions were plotted on boat sport charts (1:50.000 or 1:20.000) and clustered in discrete groups considering both space and time. Based on a "signal-to-noise character" (0–6), previously assigned to each vocalization by listening to the tapes, vocalizations with good quality (character 4–6) from at least five different groups were randomly chosen for analysis. Based on the evidence that males seem to display consistently within the same small discrete areas throughout the mating season (Bjørge *et al.*, 1995; Van Parijs *et al.*, 1997), this procedure should ensure that several individuals are analyzed from each colony. In addition, at several recording sites, several seals were obviously calling simultaneously within the audible distance of the hydrophone, increasing the likelihood of recording several individual males.

III. ANALYSIS

A. Frequency contour

In order to describe the harbor seal vocalizations, an algorithm was developed in an attempt to extract a time-varying contour of the vocalizations' frequency bandwidth. The selected 120 vocalizations were first normalized to the same peak amplitude and digitized (sampling rate$=22$ kHz and sampling size$=8$ bit) onto a Macintosh computer using the CANARY software package (Cornell Laboratory of Ornithology; Charif *et al.*, 1995). The digital spectrogram was computed using Hanning windows (Oppenheim and Schafer, 1998), 16 384-point FFTs, and overlap factor of 8. This yields a frequency-by-time matrix in which each column represents a sound spectrum derived over 743 ms. The frequency resolution of the spectrum is 5.4 Hz and the dynamic range is 48 dB. The start- and endpoint of the vocalizations were identified from the digitized waveform, and if obscure, combined with the spectrogram display [Figs. 3(a) and (b)].

The vocalization length (L) varied from 5.8 to 23.8 s. To ensure that corresponding variables in vocalizations with different durations reflect the same relative time sequence (e.g., variable 20 reflects the middle part in all calls), the bandwidth for each vocalization was estimated at regular intervals of $L/40$. With an overlap factor of 8, this procedure had a precision of minimum 46 ms. The use of 40 variables was chosen as a trade-off between adequately describing the shape of long vocalizations and avoiding too much redundant information in short vocalizations. Preliminary analysis showed that no vocalization contained energy above 2500 Hz. Thus, to increase computer speed, the overall bandwidth of the matrix was reduced from 11 500 to 2500 Hz so that before presentation to the "frequency bandwidth contour" algorithm each vocalization was represented with a 41×9800 matrix. The next step was to obtain a contour from the normalized matrix, where the contour was the value of the lower and upper frequency boundary of the bandwidth in each column (Fig. 3). We selected the average intensity of the matrix as reference value, and defined the upper and lower frequency boundaries of the bandwidth as the frequency coordinates where the intensity in ten vertical neighboring cells exceeded the overall average intensity level of the matrix by 12 dB [Fig. 3(c)]. This threshold value was arbitrary, but turned out to be sufficient to capture the general shape of the vocalizations and eliminate background noise. Repeating this approach for each column (spectrum) yielded

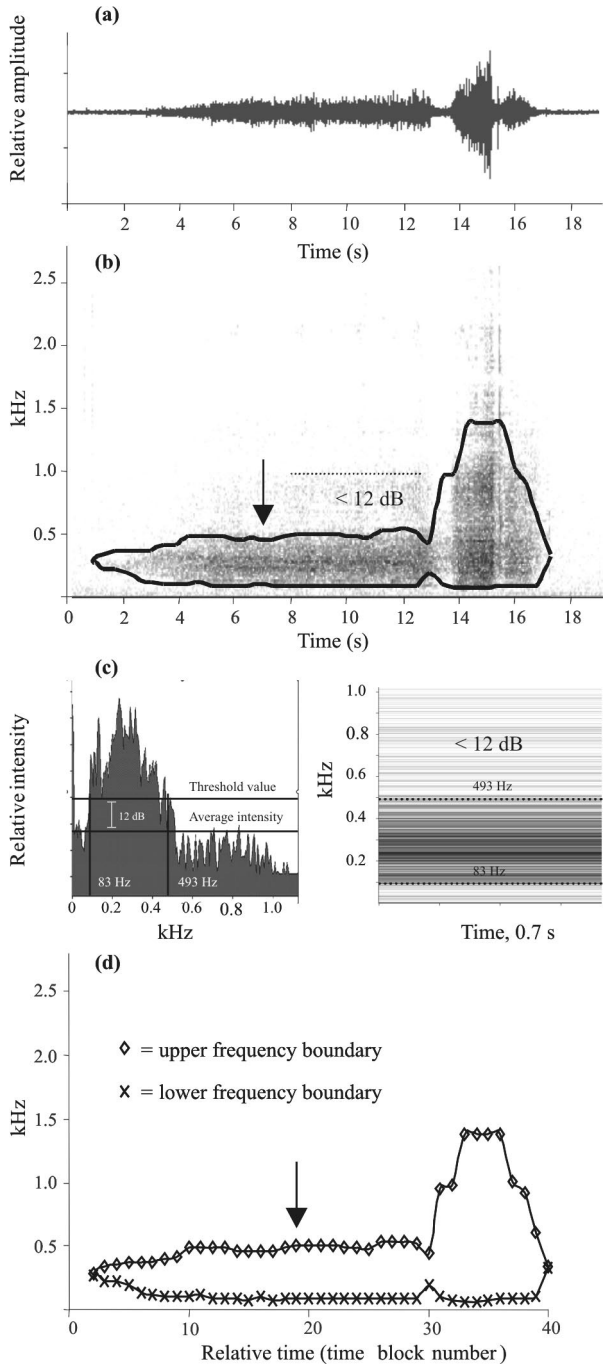


FIG. 3. Graphic illustration of how the frequency bandwidth was extracted from the harbor seal roars: (a) The waveform; (b) The spectrogram with superimposed contour. Analysis resolution = 43 Hz and 93 ms; (c) Spectrum and spectrogram illustrating schematically how the algorithm identifies and extracts the frequency coordinates. Analysis resolution = 5.4 Hz and 743 ms; (d) The frequency contour.

an estimate of the frequency bandwidth of each roar at 41 points in time [Fig. 3(d)]. Some vocalizations included noise (mainly from waves hitting the boat or vertical movements by the boat) in the analysis window with equal intensity as the signal of interest. To avoid that, spurious peaks such as these were registered as part of the vocalizations; the algorithm checked if the average intensity (for the same frequency interval) in either of the two next columns also exceeded the specific frequency boundary. If not, the algorithm assumed that the first value was an artifact and replaced the value of the cell with the average intensity in the matrix.

B. Geographic variation

PRIMER (Clarke and Warwick, 1994) was used as a statistical tool for exploring potential patterns within the data set. The frequency contours were converted to a triangular matrix of similarity between each pair of calls using a Bray–Curtis similarity coefficient (Bray and Curtis, 1957) and grouped by means of hierarchical agglomerative cluster analysis and finally ordinated with nonmetric multidimensional scaling (MDS; Kruskal and Wish, 1978). The results are presented as a dendrogram that clusters the contours in discrete groups, and as a two-dimensional ordination plot that visualizes the relationship between contours. One important feature of the multivariate analyses is that they in no way utilize any known structure among the contours of the colonies (Clarke and Warwick, 1994). The dendrogram and ordination were constructed only from the pairwise similarities among the 120 contours.

C. Traditional analysis

To determine the vocalizations' sound types, traditional analyses based on visual inspections of waveforms, spectra and spectrograms (Davies, 1964; Watkins, 1967) were performed on all vocalizations. Analysis resolution (filter bandwidth/frame length) was selected to emphasis acoustic structure of interest in either the time- or the frequency domain, i.e., both ''wide- and narrow-band'' spectrograms were produced. The time- and frequency range may be scaled to emphasize fine details of interest. Power spectra and waveform were available for any events in the signal.

IV. RESULTS

A. Patterns of geographic variation in the six harbor seals colonies

Based on the frequency contours, the six colonies were divided clearly into two major acoustic groups [Fig. 4(a)]. Group A consisted of Koster, Froan, Nordmjelde, and Kongsfjord, and group B contained Sandøy and Eynhallow [Fig. 4(a)]. Only 6 of the 120 vocalizations did not fit into this pattern (four from Kongsfjord and two from Sandøy). The two major groups could further be subdivided into several subgroups. At approximately 84% similarity threshold, 87% of the contours were grouped together with vocalizations from the same geographical area [Fig. 4(a)]. Thus, all subgroups are dominated by vocalizations from a single colony. The MDS plot (stress = 0.13) is in agreement with the dendrogram, although there is more disturbance between subgroup A5 and A6 [Fig. 4(b)]. The agreement between the two analysis methods suggests that the subgroups, and thus the frequency patterns, varied geographically. The analyses also indicated that seals in Koster and Nordmjelde used two vocalizations types. These types had different frequency patterns, e.g., the powerful roar ending the vocalizations constituted the greater part of the vocalizations in one type from Koster [Fig. 4(c)]. A nonparametric test, ANOSIM (Clarke

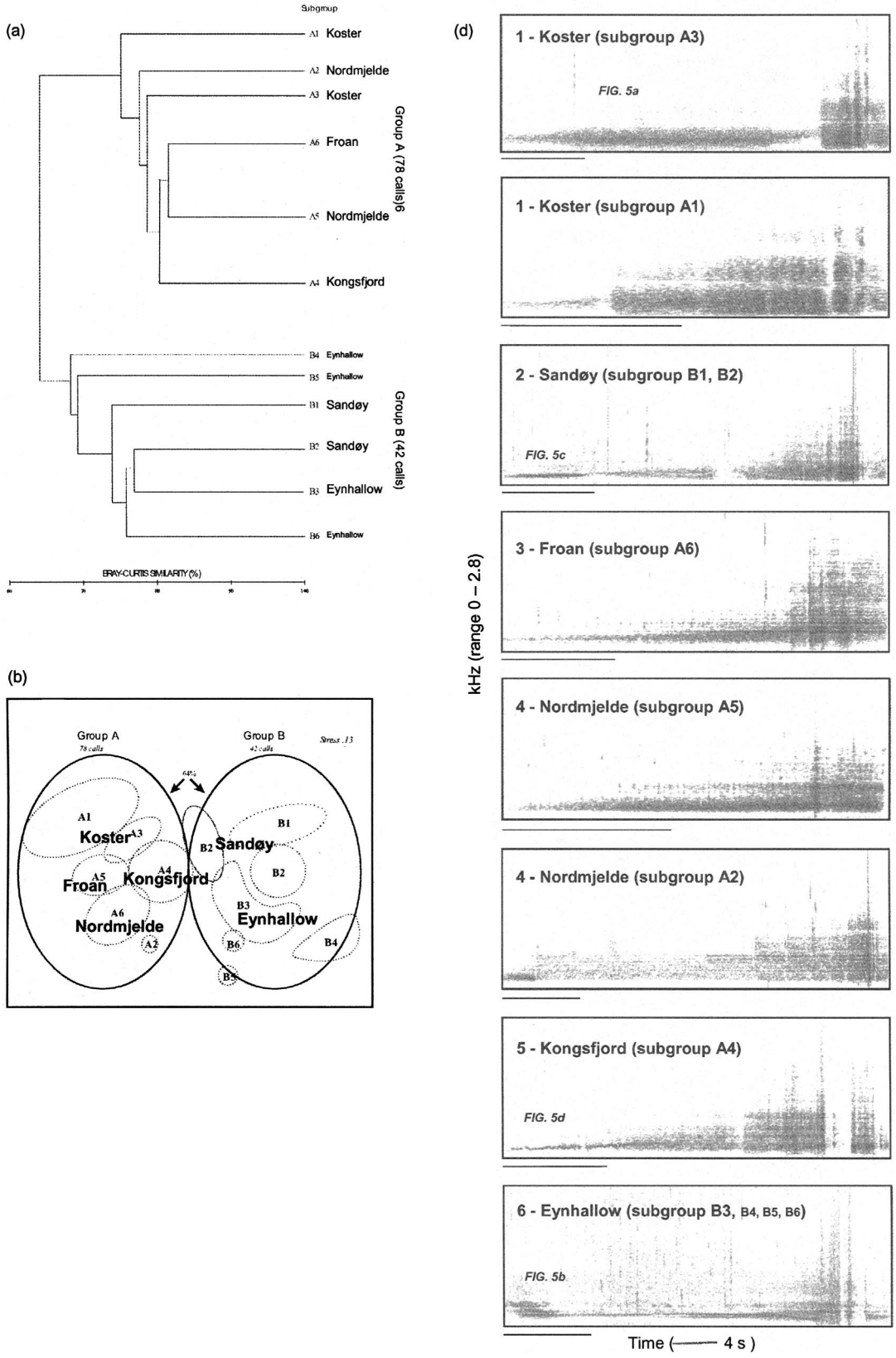


FIG. 4. Summary of acoustic relationship of six harbor seal colonies in the Northeast Atlantic: (a) The degree of acoustic similarity expressed as Bray–Curtis similarities, and displayed as a simplified dendrogram and a two-dimensional MDS plot with superimposed groups from the dendrogram; (b) Sample spectrogram of vocalizations from the colonies and subgroup in the similarity analysis. Note the two different vocalization types at Koster and Nordmjelde. Analysis resolution = 43 Hz and 93 ms. Segment shown in Fig. 5 is marked.

TABLE I. ANOSIM pairwise comparison of the frequency contours to the six seal colonies. The R values indicates the degree of separation, i.e., Sandøy and Froan/Nordmjele are best separated, while Froan and Nordmjele are least separated. Note: All colony pairs are significant different ($P<0.05$).

	Koster	Sandøy	Froan	Nordmjele	Kongsfjord
Sandøy	0.93	⋯	⋯	⋯	⋯
Froan	0.61	0.94	⋯	⋯	⋯
Nordmjele	0.37	0.94	0.29	⋯	⋯
Kongsfjord	0.48	0.67	0.57	0.48	⋯
Eynhallow	0.87	0.35	0.87	0.91	0.64

and Warwick, 1994), showed that differences between all colonies were significant ($P<0.005$; $R_{\text{global}}=0.63$; Table I).

B. Acoustical pattern

The harbor seals' vocalizations recorded in this study were broadband roars with most energy concentrated around 280 Hz (± 74 Hz) but with intense bands (more than 12 dB above the background level) at intervals up to 2000 Hz (average$=1111$ Hz, SD 346 Hz). Frequency range typically increased as the roar progressed, usually with a very abrupt increase in intensity and frequency bandwidth towards the end of the vocalization (highest frequency was located on average $82\% \pm 7\%$ out in the call). Vocalization length varied between 5.8 and 23.9 s, with an average duration of 15.0 s (± 4.0 s).

The majority of the vocalizations from Sandøy, Kongsfjord, and Eynhallow started with a characteristic short, tonal-like introductory phrase that was amplitude- and/or frequency modulated with high relative intensity (Table II). In contrast, the vocalizations from Koster, Froan, and Nordmjelde showed a relative intensity that was building up gradually during the call and was only composed of broadband, nonharmonic sound.

1. Broadband nonharmonic sound

a. "Spectrally structured sound." This was the dominating sound type in all vocalizations (Table II). The sound type is characterized by a broad frequency bandwidth with relatively intense spectral peaks [Fig. 5(a) and Fig. 6(a)] giving an aural impression of a continuous roar [Fig. 5(a)]. In 75% of the vocalizations from Nordmjelde [roars clustered in subgroup A5 in Fig. 4(a)] the roar had a rumbling quality due to much more marked time-varying spectral peaks, i.e., distinct pulses exhibiting broadband energy.

2. Tonal sound

b. "Warble." This sound type showed a combination of pulsed- and frequency-modulated structure [Fig. 5(b)]. The most striking feature was a rhythmic frequency-modulated carrier wave that varied between 240 and 300 Hz at a rate of 18 Hz. The pulsed component, indicated by the third peak in the 4096-point (21-Hz) spectrum [Fig. 6(b)], had a carrier frequency around 450 Hz and a pulse rate of 18 Hz. The large accompanying amplitude modulation was also visible by a close examination of the signal. This distinct sound type was only found in vocalizations from Eynhallow (90%, Table II), and was produced at the beginning of the vocalizations (2.7 ± 0.7 s). To the human ear, the signal had a rumbling quality with characteristic variations in the pitch.

c. "Tonal pulsed." This sound type was composed of a pure tone around 200 Hz with regular pulse length and interpulse intervals [Fig. 5(c)]. Short (3.2 ± 0.6 s) tonal pulsed introductory phrases were typical for vocalization from Sandøy (Table II). The signal had a growling quality, similar to a series of guttural (throaty) R's. Although characterisic for Sandøy, two vocalizations from Eynhallow also started with a similar but longer pulse train (6.5 ± 1.6 s). Moreover, less intense and marked pulses were identified after the ''warble and whistle buzz'' phrases in vocalizations from Eynhallow and Kongsfjord (Table II).

d. "Whistle-buzz." The whistle was similar to a simple, continuous sine wave but differed by having a small, irregular frequency modulation [Fig. 5(d)]. The carrier frequency (center frequency) was around 180 Hz and the maximum frequency magnitude variation was approximately 30 Hz. This sound type was only found in the start of vocalizations from Kongsfjord (Table II). The whistle was usually followed by a very short tonal-pulse train with interpulse intervals close to the lower limit of human perception, making it difficult to separate the pulses in time. To the human ear, the

TABLE II. Frequency of occurrence of sound-types in all six seal colonies.

		Occurrence of frequency %						
		Colony						
Category	Location in call	Koster	Sandøy	Froan	Nordmjelde	Kongsfjord	Eynhallow	Overall
Spectral structured	Whole call ex. the introductory phrase	16.7	16.7	16.7	16.7	16.7	16.7	100
Warble (FM+AM)	Introductory phrase	0.0	0.0	0.0	0.0	0.0	15.0	15.0
Tonal pulsed (AM)	Introductory phrase and part 1	0.0	16.7	0.0	4.2	15.0	16.7	52.5
Whistle (FM)	Introductory phrase	0.0	0.0	0.0	0.0	15.0	0.0	15.0

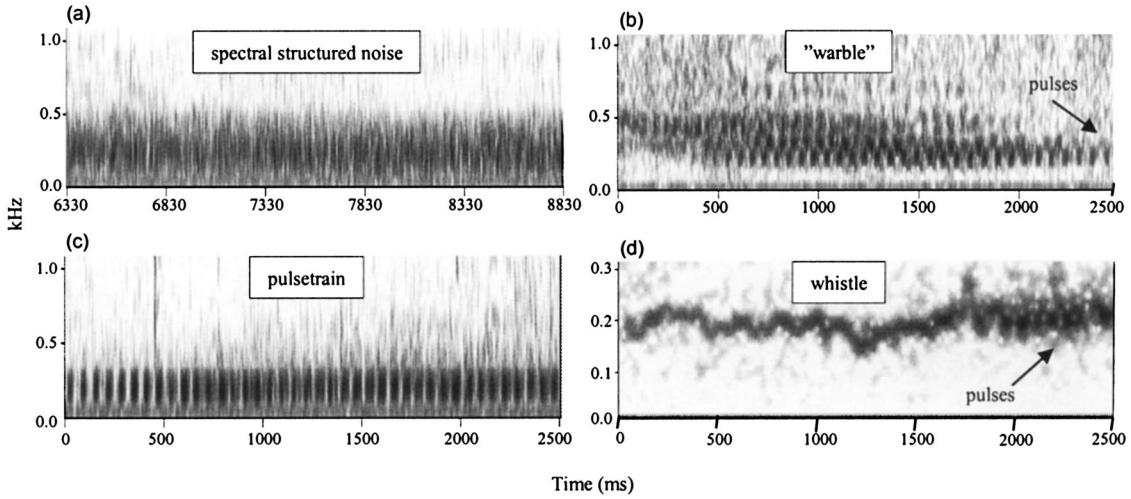


FIG. 5. Examples of four different sound types marked in Fig.4. (a) Spectral-structured sound; (b) Warble; (c) Pulse train; (d) Whistle. Analysis resolution = 342 Hz and 12 ms for (a), (b), (c), and analysis resolution= 171 Hz and 24 ms for (d).

whole phrase appeared virtually continuous and bore a strong resemblance to a buzzing bumblebee.

V. DISCUSSION

A. Methods

Since the harbor seal vocalization has a complex acoustic structure, it was important to develop a method that could objectively recognize the boundaries of the frequency bandwidth. The technique used in this study analyzes numerical values from digital spectrograms to extract a frequency bandwidth contour. There may be reason to question the reliability of bandwidth measurements in general. However, comparing the contour of the vocalizations (i.e., its ''shape'') rather than a few discrete measurements will probably reduce some of the effects of different recording conditions, such as different recording distances. As vocalizations from more colonies are compared, useful variables for exploring variability in harbor seal vocalizations may change. However, a time-varying contour will probably be more universal (but see Van Parijs *et al.*, 2003).

The technique used to ''align'' the contours in this study assumes a high level of temporal consistency, e.g., if vocalization length increases, the length of each part must increase correspondingly. This was the case for most of the harbor seal vocalizations. However, in two vocalizations from Kongsfjord, the powerful end constituted a considerably larger portion of the total vocalization. Other methods of aligning contours exist, e.g., ''time warping,'' which was developed for the problems of speech recognition. This method has been used successfully to align the fundamental frequency contour of bottlenose dolphin whistle (Buck and Tyack, 1993). This is a more complex method, and one loses information about the percentage proportions that each part constitutes.

The threshold value of 12 dB was determined through experimentation and observation. The threshold value was set relative to the average intensity value of the normalized matrix, rather than the standard approach of measuring the distance (in Hz) at a predetermined point down (in dB) from the peak frequency. This made the algorithm vulnerable to intense background noise such as waves hitting the side of the boat. However, the alternative, using the peak frequency for each time block, means that the random peak frequency must be used as reference point. The critical ratio of harbor seal is around 19–27 dB (Turnbull and Terhune, 1990). Since the matrix was reduced to 2500 Hz, the roar constitutes a considerable part of the energy in the analysis window. Thus, it is likely that the contour portrays energy that the seals are able to hear.

B. Patterns of geographic variation

Comparison of harbor seals' vocalizations from six colonies in the Northeast Atlantic revealed geographic variation in the frequency contours and use of sound types. The distance between the six harbor seal colonies ranges from 200 km to more than 2000 km (Fig. 2). The lack of correlation between acoustic relationship (Fig. 4) and distance (Fig. 2) suggests that the main causal factor responsible for the ob-

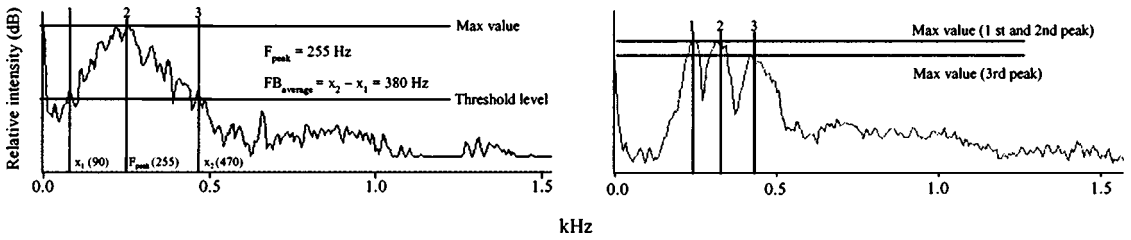


FIG. 6. Spectra of the (a) spectral-structured sound and (b) warble from Fig. 5. Analysis resolution= 21 Hz.

served geographical differences in the repertoire may be geographic isolation. The harbor seal is regarded as stationary (Thompson and Miller, 1990), with only limited movement throughout the year (Bigg, 1981). Strong site fidelity probably results in sufficient isolation for the vocalizations to evolve independently in the colonies, leading to vocal divergence with time, and maintaining acoustic integrity of the colonies. Large distances, combined with a strong fidelity to specific breeding sites, were also thought to be the principal reason for geographic variation in the repertoire of Weddell seals (Thomas and Stirling, 1983; Thomas *et al.*, 1988), bearded seals (Cleator *et al.*, 1989), harp seals (Terhune, 1994), and harbor seals (Van Parijs *et al.*, 2003).

The large vocal difference between Sandøy and Froan (area 2 and 3; Fig. 2) shows that disjunct vocal variations also can occur over relatively small distances. Although the two colonies are only separated by only 200 km, the seals produced very different roars and were actually divided into different major acoustic groups (Fig. 4). Moreover, the two colonies had the largest R value of all colony pairs (ANOSIM=94%; Table I) and all Sandøy roars started with the characteristic pulse train which were not present in roars from Froan [Fig 5(c); Table II]. Local variations in repertoire have been described in a few pinnipeds. Cleator *et al.* (1989) and Van Parijs *et al.* (2000a, 2003) found distinct differences in bearded and harbor seal vocalizations between two colonies separated by only 150 km. If variation in underwater vocalizations reflects the degree of isolation between the colonies, these data suggest that the adult harbor seal is sedentary at a scale of less than 200 km. However, there are several other possible explanations for the observed vocal variation between the colonies. The harbor seal vocalizations are likely acquired through imitation and learning (Ralls *et al.*, 1985; Van Parijs *et al.*, 2003; see the review in Janik and Slater, 1997), and thus are primarily passed from generation to generation by cultural, rather than genetic, transmission (Ford, 1991). Once a transient seal is recruited to a colony, it may show enough social plasticity to learn the local roar. A motivation for changing the vocalization type may be a preference among resident females for site-specific roars and/or an increased aggression among resident males for unfamiliar roars. Another reason for changing or modifying the vocalizations may be different environmental acoustical transmission properties, ambient noise sources, predators, depth, or topographical differences, making it advantageous to communicate over various distances at the different colonies. Since it is not known when juvenile harbor seals learn the roar, another possible explanation of the observed differences between Sandøy and Froan is that the seal pups partly learn the roar in the breeding areas, but the vocalization is developed and improved further after the juvenile seals have been recruited to the new colony. Therefore, Froan and Sandøy may maintain their acoustic integrity even if a limited exchange of animals takes place between the two colonies.

In conclusion, the observed vocal differences between all six harbor seal colonies may be explained by functionless vocal divergence between groups that is isolated due to large distances and strong site fidelity to specific breeding areas. A possible limited exchange of animals may be present but may be masked by the harbor seals ability to learn new sounds by imitation, and by immigration of juvenile seal's that have not yet learned the adult repertoire (the roar vocalization).

C. Acoustical pattern

A particular feature of our dataset were the short introductory phrases that were amplitude- and frequency modulated (Fig. 5 and Fig. 6). These sound types have, to our knowledge, not been documented in other harbor seal colonies that have been fairly well studied (Hanggi and Schusterman 1994; Van Parijs *et al.*, 1999, 2000a, 2003; Hayes *et al.*, 2004). Tonal introductory phrases were characteristic for seals from Sandøy in Møre, Kongsfjord in Finnmark and Eynhallow at the Orkney Islands (Table II). Similar patterns and sound types are well known in other marine mammals, e.g., the ''warble'' [Fig. 5(b)] and ''whistle buzz'' [Fig. 5(d)] are found in harp seals (calls 4, 5, 8 and call 2 in Møhl *et al.*, 1975 and Terhune, 1994).

An important and crucial feature of an advertisement and territorial call is the possibility for receivers to (1) determine the location (i.e., direction and distance); (2) identity; and (3) the reproductive status and quality of the sender.

Tonal pulsed sound was identified in half of the vocalizations and may have functions in sound localization and distance perception. A possible mechanism for indicating the distance between a receiver and a source could be the number of pulses detected, because the number of detected pulses will increase with the improvement of signal-to-noise ratio as the distance between source and receiver decreases. Moreover, an investigation of the capability of harbor seals to localize a sound source showed that pulsed sounds yielded better results than continues tones (Terhune, 1974, 1988). Cleator *et al.*, (1989) suggested that the bearded seal might be able to judge the distance to the singer because various parts of the song travel differently through the water. Also, the humpback whale song may have characteristic spectral structures that may contain information of how far away the singer is located. (Mercardo and Frazer, 1999). Thus, producing vocalization with different sound types and especially pulsed tones may provide harbor seals a possible code for proximity.

Hanggi and Schusterman (1994) and Van Parijs *et al.* (2002) found individual variation within measured frequency bandwidth variables, as well as in temporal variables (Van Parijs *et al.*, 2002). As mentioned before, one of our concerns of using bandwidth measurements is the risk of degradation and/or masking of such signal over distance. A better candidate for such a call structure is a tonal vocalization with an emphasized and modulated carrier frequency (Dabelsteen *et al.*, 1993). If seals live in colonies with high ambient noise, complicated social structure (e.g., display area differences; see Van Parijs *et al.*, 2000a) the frequency bandwidth could become insufficient to assure individual recognition, and more specialized signals such as the introductory phrases may be necessary.

A lek mating system refers to two or more males displaying to females and to each other at traditional sites in

which resources are not defended and the females visit only to mate (Höglund and Alatalo, 1995). Individual recognition and assessment of quality is prerequisite for the presence of lek and lek behavior in harbor seals. We consider that both these conditions are possible in the vocalizations we have recorded from harbor seals.

ACKNOWLEDGMENTS

This work was performed with the support of The Research Council of Norway. Thanks to D. Thompson and E. Bryant at Sea Mammal Research Unit, University of St. Andrews, UK, for cooperation and help during field work. Nomenclature used is from D. W. Rice (1998) Marine Mammals of the World Systematic and Distribution, and systematic and distribution after The Society for Marine Mammalogy, Special Publication Number 4.

Bigg, M. A. (**1981**). ''Harbour seal, *Phoca vitulina and Phoca largha,*'' in *Handbook of Marine Mammals*, edited by S. Ridgway and R. Harrison (Academic, New York), pp. 1–28.

Bjørge, A., Bekkby, T., Bakkestuen, V., and Framstad, E. (**2002a**). ''Interactions between harbour seals, *Phoca vitulina*, and fisheries in complex coastal waters explored by combined Geographic Information System (GIS) and energetics modeling,'' ICES J. Mar. Sci. **59**, 29–42.

Bjørge, A., Thompson, D., Hammond, P., Fedak, M., Bryant, E., Aarefjord, H., Roen, R., and Olsen, M. (**1995**). ''Habitat use and diving behaviour of harbour seals in a coastal archipelago in Norway,'' in *Whales, Seals, Fish and Man*, edited by A. S. Blix, L. Walløe, and Ø. Ulltang (Elsevier Science, Amsterdam), pp. 211–223.

Bjørge, A., Øien, N., Hartvedt, S., and Bothun, G. (**2002b**). ''Dispersal and bycatch mortality in gray, *Halichoerus grypus*, and harbor, *Phoca vitulina*, seals tagged at the Norwegian coast,'' Marine Mammal Sci. **18**, 936–976.

Boness, D. J. (**1991**). ''Determinants of mating system in the Otariidae (Pinnipedia),'' in *Behavior of Pinnipeds*, edited by S. H. Rigdway and R. J. Harrison (Elsevier, Amsterdam), pp. 211–223.

Bray, J. R., and Curtis, J. T. (**1957**). ''An ordination of the upland forest communities of southern Wisconsin,'' Ecol. Monogr. **27**, 325–349.

Buck, J. R., and Tyack, P. L. (**1993**). ''A quantitative measure of similarity for *Tursiops truncatus* signature whistles,'' J. Acoust. Soc. Am. **94**, 2497–2506.

Charif, R. A., Mitchell, S., and Clark C. W. (**1995**). CANARY 1.2 *Users Manual* (Cornell Laboratory of Ornithology, Ithaca, NY).

Clark, C. W., Marler, P., and Beeman, K. (**1987**). ''Quantitative analysis of animal vocal phonology: An application to swamp sparrow song,'' Ethnology **76**(2), 101–115.

Clarke, K., R., and Warwick, R., M. (**1994**). ''Changes in marine communities: An Approach to Statistical Analysis and Interpretation,'' Plymouth Marine Laboratory, Natural Environment Research Council.

Cleator, H. J., Stirling, I., and Smith, T. G. (**1989**). ''Underwater vocalizations of the bearded seal (*Erignathus barbatus*),'' Can. J. Zool. **67**, 1900–1910.

Coltman, D. W., Bowen, W. D., Boness, D. J., and Iverson, S. J. (**1997**). ''Balancing foraging and reproduction in the male harbour seal, an aquatically mating pinniped,'' Anim. Behav. **54**, 663–678.

Dabelsteen, T., Larsen, O. N., and Pedersen, S. B. (**1993**). ''Habitat-induced degradation of sound signals: Quantifying the effects of communications sounds and birds location on blur ratio, excess attenuation, and signal-to-noise ratio in blackbird song,'' J. Acoust. Soc. Am. **93**, 2206–2220.

Davies, L. I. (**1964**). ''Biological acoustic and the use of the sound spectrograph,'' The Southwestern Naturalist **9**, 118–145.

Fisher, H. D. (**1954**). ''Delayed implantation in the harbour seal (*Phoca vitulina*),'' Nature (London) **173**, 879–880.

Ford, J. K. (**1991**). ''Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia,'' Can. J. Zool. **69**, 1454–1483.

Hanggi, E. B., and Schusterman, R. J. (**1994**). ''Underwater acoustic displays and individual variation in male harbour seals (*Phoca vitulina.*),'' Anim. Behav. **48**, 1275–1283.

Hayes, S. A., Kumar, A., Daniel, P. C., Mellinger, D. K., Harvey, J. T., Southall, B. L., and LeBoeuf, B. J. (**2004**). ''Evaluating the function of the male harbour seal, *Phoca vitulina*, roar through playback experiments,'' Anim. Behav. **67**, 1133–1139.

Henriksen, G., and Haug, T. (**1994**). ''Status of the harbour seal *Phoca vitulina* in Finnmark, North Norway,'' Fauna. Norv. Ser. A **15**, 19–24.

Höglund, J., and Alatalo, R. V. (**1995**). Leks (Princeton University Press, Princeton, NJ).

Janik, V. M., and Slater, P. J. B. (**1997**). ''Vocal learning in mammals,'' Adv. Study Anim. Behav. **26**, 59–99.

Kruskal, J. B., and Wish, M. (**1978**). *Multidimensional Scaling* (Sage, Beverly Hills).

McCowan, B. (**1995**). ''A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins (Delphinidae, *Tursiops truncatus*),'' Ethnology **100**, 177–193.

Mercado, E., and Frazer, N. L. (**1999**). ''Environmental constraints on sound transmission by humpback whales,'' J. Acoust. Soc. Am. **106**, 3004–3016.

Møhl, B., Terhune, J. M., and Ronald, K. (**1975**). ''Underwater calls of the harp seal, *Pagophilus groenlandicus*,'' Rapp. P.-V. Reun.-Cons. Int. Explor. Mer **169**, 533–543.

Murray, S. O., Mercado, E., and Roitblat, H. L. (**1998**). ''Characterizing the graded structure of false killer whale (*Pseudorca crassidens*) vocalizations,'' J. Acoust. Soc. Am. **104**, 1679–1688.

Oppenheim, A. V., and Schafer, R. W. (**1998**). *Discrete-time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Ralls, K., Fiorelli, P., and Gish, S. (**1985**). ''Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*,'' Can. J. Zool. **63**, 1050–1056.

Ray, G. C., and Watkins, W. A. (**1975**). ''Social function of underwater sounds in the walrus *Odobenus rosmarus*,'' Rapp. P.-V. Reun.-Cons. Int. Explor. Mer **169**, 524–526.

Ray, G. C., William, A., Watkins, W. A., and Burns, J. J. (**1969**). ''The underwater song of *Erignathus* (bearded seal),'' Zoologica (N.Y.) **54**, 79–83.

Roen, R., and Bjørge, A. (**1994**). ''Haul-out behaviour of the Norwegian harbour seal during summer,'' in *Whales, Seals, Fish and Man. Proceedings of the International Symposium on the Biology of Marine Mammals in the North Atlantic*, edited by A. S. Blix, L. Walloe, and Ø. Ulltang (Elsevier, Amsterdam), pp. 61–67.

Stirling, I., Calvert, W., and Cleator, H. (**1983**). ''Underwater vocalizations as a tool for studying the distribution and relative abundance of wintering pinnipeds in the high Arctic,'' Arctic **36**, 262–274.

Stirling, I., Calvert, W., and Spencer, C. (**1987**). ''Evidence of stereotyped underwater vocalizations of male Atlantic walruses (*Obdobenus rosmarus rosmarus*),'' Can. J. Zool. **65**, 2311–2321.

Terhune, J. M. (**1974**). ''Directional hearing of a harbor seal in air and water,'' J. Acoust. Soc. Am. **56**, 1862–1865.

Terhune, J. M. (**1988**). ''Detection thresholds of a harbour seal to repeated underwater high-frequency, short duration sinusoidal pulses,'' Can. J. Zool. **66**, 1578–1582.

Terhune, J. M. (**1994**). ''Geographic variation of harp seal underwater vocalizations,'' Can. J. Zool. **61**, 2203–2212.

Thomas, J. A., and Kuechle, V. B. (**1982**). ''Quantitative analysis of Weddell seal (*Leptonychotes weddelli*) underwater vocalizations at McMurdo Sound Antarctica,'' J. Acoust. Soc. Am. **72**, 1730–1738.

Thomas, J. A., and Stirling, I. (**1983**). ''Geographic variation in underwater vocalizations of Weddell seals (*Leptonychotes weddelli*) from Palmer Peninsula and McMurdo Sound,'' Can. J. Zool. **61**, 2203–2212.

Thomas, J. A., Puddicombe, R. A., George, M., and Lewis, D. (**1988**). ''Variations in underwater vocalizations of Weddell seals (*Leptonychotes weddelli*) at the Vestfold Hills as a measure of breeding population discreteness,'' Hydrobiologia **165**, 279–284.

Thompson, P. M., and Miller, D. (**1990**). ''Summer foraging activity and movements of radiotagged common seals (*Phoca vitulina* L.) in the Moray Firth, Scotland,'' J. Appl. Ecol. **27**, 492–501.

Thompson, P. M., Miller, D., Cooper, R., and Hammond, P. S. (**1994**). ''Changes in the distribution and activity of female harbour seals during the breeding season: Implication for their lactation strategy and mating patterns,'' J. Anim. Ecol. **63**, 24–30.

Turnbull, S. D., and Terhune, J. M. (**1990**). ''White noise and pure tone marking of pure tone thresholds of a harbour seal listening in air and underwater,'' Can. J. Zool. **68**(10), 2090–2097.

Van Parijs, S. M., and Kovacs, K. M. (**2002**). ''In-air and underwater vocalization of eastern Canadian harbour seals, *Phoca vitulina*,'' Can. J. Zool. **80**(7), 1173–1179.

Van Parijs, S. M., Hastie, G. D., and Thompson, P. M. (**1999**). ''Geographic variation in temporal and spatial vocalization patterns of male harbour seals in the mating season,'' Anim. Behav. **58**,(6) 1231–1239.
Van Parijs, S. M., Hastie, G. D., and Thompson, P. M. (**2000a**). ''Individual and geographical variation in display behaviour of male harbour seals in Scotland,'' Anim. Behav. **59**, 559–568.
Van Parijs, S. M., Janik, V. M., and Thompson, P. M. (**2000b**). ''Display-area size, tenure length, and site fidelity in the aquatically mating male harbour seal, *Phoca vitulina*,'' Can. J. Zool. **78**(12), 2209–2217.
Van Parijs, S. M., Thompson, P. M., Tollit, D. J., and Mackay, A. (**1997**). ''Distribution and activity of male harbour seals during the mating season,'' Anim. Behav. **54**, 35–43.
Van Parijs, S. M., Corkeron, P. J., Harvey, J., Hayes, S. A., Mellinger, D. K., Rouget, P. A., Thompson, P. M., Wahlberg, M., and Kovacs, K. M. (**2003**). ''Patterns in the vocalizations of male harbor seals,'' J. Acoust. Soc. Am. **113**(6), 3403–3410.
Whalberg, M., Lunneryd, S. G., and Westerberg, H. (**2002**). ''The source level of harbour seal flipper slaps,'' Aquatic Mammals **28**(1), 90–92.
Watkins, W. A. (**1967**). ''The harmonic interval: Fact or artifact in spectral analysis of pulse trains,'' in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 15–42.
Wiig, Ø., and Øien, N. (**1988**). ''Recoveries of common seals *Phoca vitulina* L. tagged along the Norwegian coast,'' Fauna Norv. Ser. A **9**, 51–52.

PROGRAM OF

The 148th Meeting of the Acoustical Society of America

San Diego, California • Town and Country Hotel • 15–19 November 2004

NOTE: All Journal articles and Letters to the Editor are peer reviewed before publication. Program abstracts, however, are not reviewed before publication, since we are prohibited by time and schedule.

MONDAY MORNING, 15 NOVEMBER 2004 PACIFIC SALON 3, 8:25 A.M. TO 12:00 NOON

Session 1aAA

Architectural Acoustics and Signal Processing in Acoustics: Topical Meeting on Spatial and Binaural Evaluation of Performing Arts Spaces I: Measurement Techniques and Binaural and Interaural Modeling

Ning Xiang, Cochair
Rensselaer Polytechnic Institute, Architecture, 110 8th Street, Troy, New York 12180

Rendell R. Torres, Cochair
Rennsselaer Polytechnic Institute, School of Architecture, 110 8th Street, Troy, New York 12180-3590

Chair's Introduction—8:25

Invited Papers

8:30

1aAA1. From manikin to microphone arrays development and application of binaural measurement devices. Volker Mellert (Oldenburg Univ., Inst. f. Phys., 26111 Oldenburg, Germany, volker.mellert@uni-oldenburg.de)

Ever since stereophonic systems were designed it was obvious to make the recording by two microphone channels which mimic the aural investigation of the acoustic environment by our two ears. Head-related stereophony aims at a subjectively optimal reproduction of a sound field, in particular with headphones, whereas the multi-channel synthesis of a sound field provides a listening condition which is more independent of the individual listeners acoustic properties (e.g., head size, near-field refraction pattern). Binaural measurement devices are comparatively less complex and costly and therefore in use for about 30 years for sound field investigations, as in concert hall acoustics or in the assessment of environmental and technical sounds. A review is given on the development of head-related stereophony for investigating (mainly) room acoustics. Concepts of future devices for the assessment of technical sound are presented. The classical head-shaped recording system (dummy head) is substituted by beam-forming microphone arrays.

8:50

1aAA2. Acoustic measurements through analysis of binaural recordings of speech and music. David Griesinger (Harman Specialty Group, 3 Oak Park Dr., Bedford, MA 01730)

This paper will present and demonstrate some recent work on the measurement of acoustic properties from binaural recordings of live performances. It is found that models of the process of stream formation can be used to measure intelligibility, and, when combined with band-limited running cross-correlation, can be used to measure spaciousness and envelopment. Analysis of the running cross correlation during sound onsets can be used to measure the accuracy of azimuth perception. It is additionally found that the ease of detecting fundamental pitch from the upper partials of speech and music can be used as a measure of sound quality, particularly for solo instruments and singers.

9:10

1aAA3. Some comparisons of binaural measurements made with different dummy heads and stereo microphone techniques. Peter A. Mapp (Peter Mapp Assoc., 5 Worthington Way, Copford, Colchester, Essex, UK)

Binaural measurements have been made in a number of acoustic environments, and the results from different binaural heads and stereo microphones are compared. The object of the study was not only to establish what practical differences occurred between the various head formats, but also to see if a stereo microphone or pseudohead could be used for making auditorium binaural measurements. Five measurement platforms were employed. These included two binaural dummy heads, binaural in-ear probe microphones, an SAAS pseudohead stereo microphone and a M-S (midside) stereo microphone. In the latter case, three different midside ratios were employed and compared. The measurements were made in a reverberant recital hall (2.5-s RT) and small acoustically treated listening room (RT 0.2 s). Whereas relatively minor differences were found to occur between the heads, significant differences were found to occur with the stereo microphones. It is concluded that while useful information can be obtained from a stereo microphone, it is far from being the same as binaural.

9:30

1aAA4. Visualization of sound fields by three-dimensional representations. Yasuhiro Tokita (RISE, Waseda Univ., Okubo 3-4-1, Shinjuku, 169–8555, Tokyo, Japan, tokita@acoust.rise.waseda.ac.jp) and Yoshio Yamasaki (Waseda Univ.)

In this study, three-dimensional sound fields can be visualized by measured data in an addition to transient numerical solutions. Concretely, the results of measurements by the four-point microphone method are used in the visualization of three-dimensional sound fields. By the results, distribution of virtual sources and hedgehog pattern are visualized as the receiving point is located on the center point in three-dimensional sound fields. The analyses of the transient sound fields are performed by the discretization of the wave equation. As the technique of numerical calculation, the finite difference method (FDM) is used. Transient sound fields are visualized by numerical solutions of sound pressure, instantaneous sound intensity, sound energy and particle displacement in three-dimensional space in addition to two-dimensional space. It has been verified that the numerical solutions have high accuracy by comparison with the corresponding analytical solutions. VRML97 is used as the viewer of three-dimensional sound fields. In this study, several trials of three-dimensional animation are going to be shown by VRML97 in order to visualize the transient sound fields.

9:50

1aAA5. Directional measurements in sound fields using a spherical microphone array. Bradford N. Gover (Inst. for Res. in Construction, Natl. Res. Council, Ottawa, ON K1A 0R6, Canada, brad.gover@nrc-cnrc.gc.ca)

A 32-element spherical array of microphones has been used to perform directional analysis of sound at a point. Optimized filter-and-sum beamforming is used to generate a highly directional beam over a broad frequency range. The beam can be steered in any three-dimensional direction, enabling angular discrimination of sound at a listening point. The directional responses can also be analyzed temporally, giving the directionality of the sound field at any time instant, or over any desired time range. By establishing a sound field with a test signal such as MLS or a sine sweep, the omnidirectional impulse response at each array microphone can be determined. Subsequent beamforming results in a set of directional impulse responses, which enables isolation of individual arrivals, or analysis in terms of early/late sound, for example. Various measurements made with the spherical array system will be presented and discussed.

10:10–10:20 Break

10:20

1aAA6. Front-to-rear movement of reverberation in a concert hall. Jonah Sacks and Mendel Kleiner (Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180)

In some concert halls one can perceive a front-to-rear movement of late reverberation as noted by T. J. Schultz [J. Sound Vib. **43**, 555–561 (1975)]. Listening tests show that the effect is not likely to be due to timbral change but rather to result from the spatial properties of the reverberant field. The effect can be noted as a change in late center time as a function of lateral angle of incidence of reverberant sound. The effect is likely to be a result of the scattering characteristics of hall surfaces. Measurements using a simplified scale model support this hypothesis. Measurements in Boston's Symphony Hall yield similar results. [Work supported by RPI.]

10:40

1aAA7. Classification of binaural-activity patterns. Wolfgang Hess and Jens Blauert (Inst. of Commun. Acoust., Ruhr-Universitaet, 44780 Bochum, Germany, wghess@gmx.de)

The three-dimensional, time-variant output of a computational model of the human auditory system was investigated with the aim of extracting perceptually patterns as relevant to spatial impression. To this end, the same stimuli with which the model was excited were also presented to human listeners which evaluated (i) the auditory source width (ASW), i.e., the horizontal extent of the auditory event, and (ii) the listener envelopment (LEV), i.e., the amount of envelopment by the auditory event. The listeners had to be trained beforehand as the discrimination of source- and room-related parameters of auditory events is not a common task to them. To allow for natural localization and to enable interactive listening, the stimuli were presented through head-tracked headphones. Trained listeners appeared to produce consistent results when depicting auditory objects on a compass-rose-shaped plane, where they had to mark the position and broadening of the auditory event(s) by the position(s) and length(s) of straight lines, and the envelopment by the position and size of an ellipse.

11:00

1aAA8. Binaural signal analysis of diffuse sound fields. Pedro A. Novo (Inst. of Commun. Acoust., IC1 132, Ruhr-Univ. of Bochum, 44780 Bochum, Germany, pedro.novo@ruhr-uni-bochum.de)

The simulation and the binaural recording of sounds produced by large crowds and rain impact sounds has produced an unexpected result. The listeners have reported that the auditory events were mainly concentrated to the left and to right, although the sound sources were equally distributed around the listeners. A similar result was reported with binaural recordings of applause. The results of a binaural signal analysis suggest that the key aspect regarding the lateral position dominance of the auditory events is connected to the sections where the cross-correlation coefficient assumes negative values. A comparison between normalized and non-normalized cross-correlation function predictions indicates that the latter is a better predictor for the cases studied. An adaptation period of several seconds was reported by several listeners. It is suggested that this adaptation period is related to the variations of the cross-correlation, which only average out after, approximately 2 s. The binaural impulse responses of a concert hall at two different listener positions have been analyzed in the light of the previous findings. In particular the diffuse sound field buildup will be analyzed in detail. Implications for the auditory source width and listener envelopment will be discussed.

11:20

1aAA9. Auditory and visual spatial impression: Recent studies of three auditoria. Andy Nguyen and Densil Cabrera (School of Architecture, Design Sci. and Planning, Wilkinson Bldg. G04, Univ. of Sydney, Sydney, NSW 2006, Australia)

Auditory spatial impression is widely studied for its contribution to auditorium acoustical quality. By contrast, visual spatial impression in auditoria has received relatively little attention in formal studies. This paper reports results from a series of experiments investigating the auditory and visual spatial impression of concert auditoria. For auditory stimuli, a fragment of an anechoic recording of orchestral music was convolved with calibrated binaural impulse responses, which had been made with the dummy head microphone at a wide range of positions in three auditoria and the sound source on the stage. For visual stimuli, greyscale photographs were used, taken at the same positions in the three auditoria, with a visual target on the stage. Subjective experiments were conducted with auditory stimuli alone, visual stimuli alone, and visual and auditory stimuli combined. In these experiments, subjects rated apparent source width, listener envelopment, intimacy and source distance (auditory stimuli), and spaciousness, envelopment, stage dominance, intimacy and target distance (visual stimuli). Results show target distance to be of primary importance in auditory and visual spatial impression—thereby providing a basis for covariance between some attributes of auditory and visual spatial impression. Nevertheless, some attributes of spatial impression diverge between the senses.

11:40

1aAA10. Evaluation of a model of auditory source width based on the interaural cross-correlation coefficient. Russell Mason, Tim Brookes, and Francis Rumsey (Inst. of Sound Recording, Univ. of Surrey, Guildford, Surrey, GU2 7XH, UK)

A model of human perception of auditory source width has been developed, based on the interaural cross-correlation coefficient (IACC). The following factors differentiate it from more commonly used IACC-based models: the use of a running analysis to quantify variations in width over time; half-wave rectification and low pass filtering of the input signal to mimic the breakdown of phase locking in the ear; compensation for the frequency and loudness dependency of perceived width; combination of a model of perceived location with a model of perceived width; and conversion of the results to an intuitive scale. Objective and subjective methods have been used to evaluate the accuracy and limitations of the resulting model.

Session 1aBB

Biomedical Ultrasound/Bioresponse to Vibration: Topical Meeting on Ultrasound Characterization of Cancellous and Cortical Bone I: Propagation and Scattering Properties

Kendall R. Waters, Cochair
National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305-3328

Brent K. Hoffmeister, Cochair
Department of Psychology, Rhodes College, 2000 North Parkway, Memphis, Tennessee 38112

Chair's Introduction—7:55

Invited Papers

8:00

1aBB1. Ultrasonic wave propagation and scattering in cancellous bone. Keith A. Wear (FDA CDRH HFZ-142, 12720 Twinbrook Pky., Rockville, MD 20852)

Theoretical models and experimental data describing the interaction between ultrasound and cancellous bone will be discussed. Ultrasonic attenuation in cancellous bone is much greater than that for soft tissues and varies approximately linearly with frequency between 400 kHz and 1.7 MHz. Speed of sound in cancellous bone is slightly higher than that for soft tissues and decreases gradually with frequency at diagnostic frequencies (between 300 and 700 kHz). The dependence of phase velocity on porosity may be predicted from theory of acoustic propagation in fluid-filled porous solids. The negative dispersion can be explained using a stratified two-component model. At diagnostic frequencies, scattering varies approximately as frequency to the nth power where $3<n<3.5$. This may be explained either by a continuous or discrete scattering model. The latter represents trabeculae as small finite-length cylindrical scatterers. The frequency dependence of scattering, along with the anisotropy of scattering and attenuation, implies that absorption is a greater source of attenuation than scattering. [Funding from the FDA Office of Women's Health is gratefully acknowledged.]

8:25

1aBB2. Ultrasonic wave propagation in human cancellous bone: Application of Biot theory. Zine El Abiddine Fellah, Jean Yves Chapelon (INSERM, unite 556, 151 cours Albert Thomas, 69424 Lyon), Walter Lauriks (Laboratorium voor Akoestiek en Thermische Fysica, Katholieke, France), and Claude Depollier (Laboratoire d'Acoustique de l'Universite du Maine, UMR-CNRS, France)

Ultrasonic wave propagation in human cancellous bone is considered. Reflection and transmission coefficients are derived for a slab of cancellous bone having an elastic frame using Biot's theory modified by the Johnson *et al.* model for viscous exchange between fluid and structure. Numerical simulations of transmitted waves in the time domain are worked out by varying the modified Biot parameters. The sensitivity of each physical parameter used in the theory has been studied in transmission. Some parameters play an important role in slow wave waveform, such as the viscous characteristic length and pore fluid bulk modulus. However, other parameters play an important role in the fast wave waveform, such as solid density and shear modulus. We also note from these simulations that some parameters, such as porosity, tortuosity, thickness, solid bulk modulus and skeletal compressibility frame, play an important role simultaneously in both fast and slow waveforms compared to other parameters which act on the waveform of just one of the two waves. Experimental results for slow and fast waves transmitted through human cancellous bone samples are given and compared with theoretical predictions.

8:50

1aBB3. Simulation of ultrasound propagation in bone. Jonathan J. Kaufman, Gangming Luo (CyberLogic, Inc., 611 Broadway, Ste. 707, New York, NY 10012, jjkaufman@cyberlogic.org), and Robert S. Siffert (Mount Sinai School of Medicine, New York, NY 10012)

Ultrasound has been proposed as a means to noninvasively assess bone and, particularly, bone strength and fracture risk, as for example in osteoporosis. Because strength is a function of both mineral density and architecture, ultrasound has the potential to provide more accurate measurement of bone integrity than, for example, with x-ray absorptiometric methods. Although some of this potential has already been realized—a number of clinical devices are presently available—there is still much that is unknown regarding the interaction of ultrasound with bone. Because of the inherent complexity of the propagation medium, few analytic solutions exist with practical application. For this reason, ultrasound simulation techniques have been developed and applied to a number of different problems of interest in ultrasonic bone assessment. Both 2D and 3D simulation results will be presented, including the effects of architecture and density on the received waveform, propagation effects of both cortical and trabecular bone, and the

relative contributions of scattering and absorption to attenuation in trabecular bone. The results of these simulation studies should lead to improved understanding and ultimately to more effective clinical devices for ultrasound bone assessment. [This work was supported by The Carroll and Milton Petrie Foundation and by SBIR Grant No. 1R43RR16750 from the National Center for Research Resources of the NIH.]

Contributed Papers

9:15

1aBB4. Apparent integrated backscatter from cancellous bone in the frequency range 2.5–7.5 MHz. Brent K. Hoffmeister, Charles I. Jones III (Dept. of Phys., Rhodes College, 2000 North Pkwy., Memphis, TN 38112, hoffmeister@rhodes.edu), and Sue C. Kaste (St. Jude Children's Res. Hospital, Memphis, TN 38105)

Ultrasonic backscatter may offer useful new techniques for ultrasonic bone assessment. For this study, backscatter measurements were performed on 18 specimens of bovine cancellous bone using a 5-MHz broadband ultrasonic system operating between 2.5–7.5 MHz. Specimens were obtained from the proximal tibia, and prepared in the shape of cubes (15-mm side length) with faces oriented along principal anatomic directions (anterior, posterior, medial, lateral, superior, and inferior). A mechanical scanning system was used to acquire ultrasonic backscatter signals from 144 sites perpendicular to each face of each cube. The signals were analyzed to extract values of apparent integrated backscatter (AIB), a parameter that represents the frequency-averaged backscattered power uncompensated for the effects of attenuation and diffraction. AIB demonstrated highly significant linear correlations with bone mineral density (BMD) for both the transverse (anterior, posterior, medial, and lateral) and longitudinal (superior and inferior) directions. In all cases AIB decreased with increasing BMD. No significant anisotropy was detected. We conclude that AIB measurements performed on bovine cancellous bone in this range of frequencies correlate well with BMD.

9:30

1aBB5. Some aspects of the atypical acoustic behavior of cancellous bone. Patrick H. Nicholson (Dept. of Health Sci., Univ. of Jyvaskyla, P.O. Box 35, 40014, Finland, patrick_nicholson@btopenworld.com)

Cancellous bone, consisting of a porous framework of solid trabeculae saturated in marrow fluid, is a challenging material in terms of ultrasonic characterization. This study considers some unusual aspects of the ultrasonic behavior of cancellous bone. These anomalies highlight the boundaries of our knowledge and, arguably, point us towards promising directions for future progress. Profoundly different wave propagation regimes can exist along different anatomical axes. Along the main axis of trabecular orientation, two longitudinal waves are observed, but only one wave is observed for propagation transversely across the trabeculae. In the latter case, a positive attenuation slope and negative velocity dispersion are seen, in apparent violation of the Kramers–Kronig relations. The properties of the saturating fluid have major effects on acoustic properties, and temperature studies also point towards the importance of the fluid. In some clinical studies, velocity and attenuation in the heel show opposite changes, which is difficult to explain without considering the possibility of changes in the marrow as well as the bone. Shear wave propagation in fluid-filled cancellous bone has received little attention. Measurements in whale bone suggest that shear waves propagate in dry bone, but are rapidly attenuated in the presence of saturating fluid.

9:45

1aBB6. Dependence of dispersive properties of cancellous bone on bone mineral and collagen content. K. R. Waters (NIST, Mater. Reliability Div., 325 Broadway, Boulder, CO 80305), B. K. Hoffmeister, and S. A. Whitten (Rhodes College, Memphis, TN 38112)

Cancellous bone is known to exhibit dispersion over clinically-relevant ultrasonic frequencies. However, the dependence of the dispersive properties of cancellous bone on bone mineral and collagen content are currently not well understood. Cancellous bone specimens ($n=16$) obtained from the proximal end of bovine tibia were cut into cubes (15 mm sides) and prepared for measurement along the anterior-posterior (AP), medial-lateral (ML), and superior-inferior (SI) directions. Bone mineral densities (BMD) ranged from 100 to 250 kg/m^3. Subsets of the bone specimens were chemically treated to remove bone mineral ($n_M=10$) and collagen content ($n_C=6$). Ultrasonic measurements were performed before and after treatment in a water bath (23 °C) using a pair of 2.25-MHz planar transducers aligned coaxially. Phase velocity was determined from Fourier analysis of the ultrasonic signals, and dispersion was calculated as the change in phase velocity across the usable bandwidth. Demineralization produced highly significant changes in dispersion along the AP ($p=0.004$) and ML ($p=0.002$) directions, but not the SI direction ($p=0.6$). Decollagenization produced no significant change for any direction ($p>0.3$). These results indicate that bone mineral plays a much larger role than collagen in the dispersive behavior of cancellous bone.

10:00–10:15 Break

10:15

1aBB7. Ultrasound attenuation as a quantitative measure of fracture healing. Sabina Gheduzzi (Orthopaedic Surgery, Univ. of Bristol, AOC Lower Level, Southmead Hospital, Bristol BS10 5NB, UK, s.gheduzzi@bristol.ac.uk), Victor F. Humphrey (Univ. of Southampton, UK), Simon P. Dodd, James L. Cunningham, and Anthony W. Miles (Univ. of Bath, UK)

The monitoring of fracture healing still relies upon the judgment of callus formation and on the manual assessment of the stiffness of the fracture. A diagnostic tool capable of quantitatively measuring healing progression of a fracture would allow the fine-tuning of the treatment regime. Ultrasound attenuation measurements were adopted as a possible method of assessing the healing process in human long bones. The method involves exciting ultrasonic waves at 200 kHz in the bone and measuring the reradiation along the bone and across the fracture zone. Seven cadaveric femora were tested *in vitro* in intact form and after creating a transverse fracture by sawing through the cortex. The effects of five different fracture types were investigated. A partial fracture, corresponding to a 50% cut through the cortex, a closed fracture, and fractures of widths varying between 1, 2, and 4 mm were investigated. The introduction of a fracture was found to produce a dramatic effect on the amplitude of the signal. Ultrasound attenuation was found to be sensitive to the presence of a fracture, even when the fracture was well reduced. It would therefore appear feasible to adopt attenuation across a fracture as a quantitative measurement of fracture healing.

10:30

1aBB8. Experimental evaluation of bone quality using speed of sound measurement in cadaver mandibles. Iyad Al Haffar, Frederic Padilla, Pascal Laugier (Laboratoire d'Imagerie Paramtrique, CNRS UMR 7623, Universit Paris 6, 15 rue de l'Ecole de Medecine, 75006 Paris, France), Raphael Nefussi (Laboratoire de Biologie, 75006 Paris, France), Sami Kolta (Universit Ren Descartes, Paris, France), and Jean-Michel Foucart (Service dImagerie Medicale Hotel-Dieu de Paris)

This study is the first attempt to use speed of sound (SOS) as a new ultrasonic diagnostic tool for bone quality assessment before oral implant treatment. The objective is to demonstrate the *in vitro* feasibility of local SOS measurement at the mandible, and to investigate the relationships between mandibular SOS and local bone mineral density (BMD) and the ratio between the trabecular and cortical thickness (Tb.Th/Cort.Th). Fourteen excised human mandible were measured in transmission with a pair of flat 1.6-MHz central frequency transducers. Three regions of interest (ROIs) were selected in the specimens: incisor, premolar and molar regions. Ten measurements with repositioning were performed on each ROI

to determine the short-term precision. Dual x-ray absorptiometry scans were performed on the samples for local BMD measurements. Computed tomography (CT) was used to determine mandibular cross-sectional morphological measurements. SOS measurements at different sites were significantly different, reflecting the heterogeneity between the different sites. A strong linear relationship was found between SOS and BMD ($r^2 = 0.68$, $p < 0{,}0001$) while a nonlinear relationship was found between SOS and Tb.Th/Cort.Th ($r^2 = 0.48$, $p < 0{,}0001$). This study demonstrates *in vitro* the feasibility of SOS measurement at the mandible. *In vitro* mandibular SOS reflects local BMD and Tb.Th/Cort.Th before implant.

10:45

1aBB9. Investigation of ultrasonic properties of human hip bone: Relationship to bone mineral density (BMD) and micro-architecture. Frederic Jenson, Frederic Padilla, and Pascal Laugier (Laboratoire d'Imagerie Paramtrique, CNRS UMR 7623 Universit Paris 6, 15 rue de l'Ecole de Medecine, 75006 Paris, France)

The goal of this study is to investigate the relationships between QUS measurements (transmission and reflection) and BMD and micro-architecture measured on human proximal femur. Thirty-eight 1-cm thick slices of trabecular bones were removed from fresh human proximal femurs. Two-dimensional scans were performed using 1-MHz focused transducers to measure QUS parameters: normalized BUA (nBUA), SOS and broadband ultrasonic backscatter (BUB). BMD was determined using QCT. Six hundred sixty non-overlapping ROIs were selected for quantitative analysis on both QUS and QCT images. Finally, 37 cylindrical cores were extracted from 31 specimens (8 mm diameter). Their micro-architecture was derived using micro-tomography. All QUS parameters were significantly correlated to BMD (nBUA: $r^2 = 0.7$; SOS: $r^2 = 0.75$; BUB: $r^2 = 0.44$). Multiple regression models were tested with BMD and micro-architecture. SOS only was best explained by a combination of bone quantity (BV/TV) and micro-architecture parameters (SOS with BMD and structure model index: $r^2 = 0.93$). No additional variability of nBUA and BUB could be explained by micro-architectural parameters in addition to BMD or BV/TV. These results suggest that QUS measurement at the hip may be relevant for fracture risk prediction.

11:00

1aBB10. Numerical simulation of wave propagation in cancellous bone. Emmanuel Bossy (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215), Frederic Padilla, Frederic Jenson, Guillaume Haiat, and Pascal Laugier (Laboratoire d'Imagerie Paramtrique, 75006 Paris, France)

Physical mechanisms implied in the propagation of ultrasound waves in trabecular bones are not yet clearly understood. The relative role of scattering and absorption in the attenuation process is unknown. To dissociate these two effects, a simulation of wave propagation in 3D volumes of trabecular bone is performed, and compared to experimental results. The simulation algorithm accounts for scattering and volumetric absorption into both the saturating fluid and bone. The simulation software uses a finite difference approach based on the Virieux numerical scheme. An incident plane wave is propagated on a volume of bone of approximately $5 \times 5 \times 8$ mm^3. These volumes were reconstructed from high resolution micro-CT experiments. Several types of simulations are computed: bone described as a fluid matrix saturated by fluid (i.e., neglecting the shear wave propagation into the matrix) and bone described as a solid matrix saturated by fluid, with or without absorption. From the transmitted signals, attenuation and speed of sound through the specimens are calculated and compared to experimental values obtained on the same specimens. Finally, the contribution of scattering to the total attenuation is estimated, and the importance of mode conversions as well as absorption.

11:15

1aBB11. Ultrasonic wave propagation in cortical bone mimics. Simon P. Dodd, James L. Cunningham, Anthony W. Miles (Ctr. for Orthopaedic Biomechanics, Univ. of Bath, Bath, BA2 7AY, UK), Victor F. Humphrey (Univ. of Southampton, Southampton, SO17 1BJ, UK), and Sabina Gheduzzi (Univ. of Bristol, Bristol, BS10 5NB, UK)

Understanding the velocity and attenuation of ultrasonic waves in cortical bone is important for studies of osteoporosis and fractures. In particular, propagation in free- and water-loaded acrylic plates, with a thickness range of around 1–6 mm, has been widely used to mimic cortical bone behavior. A theoretical investigation of Lamb mode propagation at 200 kHz in free- and water-loaded acrylic plates revealed a marked difference in the form of their velocity and attenuation dispersion curves as a function of frequency thickness product. In experimental studies, this difference between free and loaded plates is not seen. Over short measurement distances, the results for both free and loaded plates are consistent with previous modeling and experimental studies: for thicker plates (above 3–4 mm), the velocity calculated using the first arrival signal is a lateral wave comparable with the longitudinal velocity. As the plate thickness decreases, the velocity approaches the S_0 Lamb mode value. WAVE2000 modeling of the experimental setup agrees with experimental data. The data are also used to test a hypothesis that for thin plates the velocity approaches the corresponding S_0 Lamb mode velocity at large measurement distances or when different arrival time criteria are used. [Work supported by Action Medical Research.]

11:30

1aBB12. Structural and elastic determinants of axial transmission ultrasonic velocity in the human radius. Kay Raum (Q-BAM Group, Dept. of Orthopedics, Martin Luther Univ., 06097 Halle, Germany; kay.raum@medizin.uni-halle.de), Ingrid Leguerney, Florent Chandelier, Maryline Talmant, Amena Saied, Pascal Laugier (Laboratoire d'Imagerie Paramtrique, 75006 Paris, France), and Françoise Peyrin (CREATIS and European Synchrotron Radiation Facility, 38043 Grenoble, France)

Accurate clinical interpretation of the sound velocity derived from axial transmission devices requires a detailed understanding of the propagation phenomena involved and of the bone factors that have an impact on measurements. In the low-megahertz range, ultrasonic propagation in cortical bone depends on anisotropic elastic tissue properties, porosity, and the spatial dimensions, e.g., cortical thickness. A subset of ten human radius samples from a previous biaxial transmission investigation was inspected using 50-MHz scanning acoustic microscopy (SAM) and synchrotron radiation computed tomography (SR-CT). Low-frequency axial transmission sound speed at 1 and 2 MHz was related to structural properties (cortical thickness C.Th, porosity POR, Haversian cavity density CDH) and tissue parameters (acoustic impedance Z, mineral density MD) on site-matched cross sections. Significant linear multivariate regression models (1 MHz: $R = 0.84$, $p < 1\text{E}-4$, 2 MHz: $R = 0.65$, $p < 1\text{E}-4$) were found for the combination of C.Th with POR and Z (measured in the external cortical quarter). A modified model accounting for the nonlinear dispersion relation with C.Th was also highly significant ($R = 0.75$, $p < 1\text{E}-4$, rmse = 49.22 m/s) and explained (after adjustment for dispersion) 55.6% of the variance of the sound velocity by variations of porosity (15.6%) and impedance (40%).

11:45

1aBB13. Simulated axial transmission propagation on cortical bone. M. Talmant, E. Bossy, and P. Laugier (Laboratoire d'Imagerie Paramétrique, UMR CNRS 7623, 15 rue de l'école de Médecine, 75006 Paris, France)

The ultrasonic axial transmission technique, used to assess cortical shells of long bones, is investigated using numerical simulations. The 3-D finite difference code generates synthetic signals recorded at different dis-

tance emitter–receiver for a linear arrangement of transducers placed along the bone axis. Academic modeling of bone based on cylindrical tubular shape made of anisotropic and porous material has been reported [J. Acoust. Soc. Am. **115** (2004)]. The aim of this paper is to refine the model by taking into account more realistic structural and material bone properties. Finite difference modeling was applied to 50 human radius specimens which were examined both by x-ray tomography at different resolutions and by ultrasonic axial transmission technique (1 MHz). The x-ray macroscopic geometry (pixel around 100 μm) of the 50 samples was imported in the code. Material properties are assigned to each bone according to its own microarchitecture examined at a 10 μm scale. Different assumptions of the relationship between bone structural properties and material properties were tested. Simulations were validated by comparison with experimental results. Numerical simulations of transient propagation in bone is a powerful tool to enlighten interaction between ultrasound and bone and consequently to improve ultrasound based devices for clinical use.

MONDAY MORNING, 15 NOVEMBER 2004 CALIFORNIA ROOM, 9:00 A.M. TO 12:00 NOON

Session 1aSC

Speech Communication and Signal Processing in Acoustics: Current Developments in Speech Communication: Techniques and Technology (Poster Session)

Shrikanth Narayanan, Chair
Signal and Image Processing Institute, University of Southern California, 3740 McClintock Avenue, Los Angeles, California 90089-2564

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon.

1aSC1. Supplementary features for improving phone recognition. Mridul Balaraman, Sorin Dusan, and James L. Flanagan (Ctr. for Adv. Information Processing, 96 Frelinghuysen Rd., Piscataway, NJ 08855, mridulb@caip.rutgers.edu)

Traditional speech recognition systems use mel-frequency cepstral coefficients (MFCCs) as acoustic features. The present research aims to study the classification characteristics and the performance of some supplementary features (SFs) such as periodicity, zero crossing rate, log energy and ratio of low frequency energy to total energy, in a phone recognition system, built using the Hidden Markov Model Tool Kit. To demonstrate the performance of the SFs, training is done on a subset of the TIMIT data base (DR1 data set) on context independent phones using a single mixture. When only the SFs and their first derivatives (feature set of dimension 8) are used the recognition accuracy is found to be 42.96% as compared to 54.65% when 12 MFCCs and their corresponding derivatives are used. The performance of the system improves to 56.49%, when the SFs and their derivatives are used along with the MFCCs. A further improvement to 60.34% is observed when the last 4 MFCCs and their derivatives are replaced by SFs and their derivatives, respectively. These results indicate that the supplementary features contain classification characteristics which can be useful in automatic speech recognition.

1aSC2. Non-monotonic spectral transitions between successive phonemes. Sorin Dusan (Ctr. for Adv. Information Processing, Rutgers Univ., 96 Frelinghuysen Rd., Piscataway, NJ 08854, sdusan@caip.rutgers.edu)

In fluent speech one would expect that the transition between two successive acoustic targets would exhibit either a monotonic increasing or a monotonic decreasing trajectory in the time domain of each spectral parameter. Closer examination reveals that this is not always the case. This study investigates the existence of non-monotonic trajectories found in the acoustic spectral domain at the transition between some successive phonemes. This non-monotonic behavior consists of transitional regions that exhibit either increasing and then decreasing or decreasing and then increasing trajectories for some acoustic spectral parameters. Using the superposition principle, the training part of the TIMIT acoustic-phonetic database is used to build 2471 diphone trajectory models, based on the 61 symbols used in the database for phonetic transcription. Various non-monotonic trajectories are found in some of these models for some spectral representations including the linear prediction coding (LPC) parameters and the mel-frequency cepstral coefficients (MFCC). In some cases these spectral peaks or valleys between phonemes have significant amplitudes reaching approximately the standard deviation at the center, positions of the adjacent phonemes. These non-monotonic trajectories cannot be explained by the Stevens quantal theory of speech in which abrupt changes appear at transitions between two acoustic targets.

1aSC3. Learning segments from waveforms: A statistical model of early phonological acquisition. Ying Lin (Phonet. Lab, Linguist. Dept., UCLA, Los Angeles, CA 90095-1543)

We consider the problem of early phonological acquisition within a statistical learning framework. Given a set of waveforms, our goal is to output a set of models for phonological categories as well as representations of the utterances in terms of these categories. The basic assumption of our model is that the learning of sound categories neither precedes nor follows segmentation of the waveform, but improves together with segmentation in an iterative manner. In statistical learning terms, segmentation is regarded as ''missing data,'' while category models are treated as parameters of interest. Our algorithm iterates over two basic steps: First, given a set of category models, we compute a mixture of segmentations for each utterance; Second, we improve the category models with unlabeled segments using unsupervised learning. Starting from an acoustic segmentation, each iteration produces an initial estimate of the parameters for the next iteration, until the algorithm converges to a set of sublexical models and a set of segmentations for each utterance. The results of run-

ning this algorithm on TIMIT and motherese data suggest that it approximately identifies segment-sized units and their associated categories in a completely unsupervised manner.

1aSC4. Phoneme recognition with kernel learning algorithms. Hassan H. Namarvar and Theodore W. Berger (Dept. of Biomed. Eng., Univ. of Southern California, OHE-500, Los Angeles, CA 90089-1451, heidarin@usc.edu)

An isolated phoneme recognition system is proposed using time-frequency domain analysis and support vector machines (SVMs). The TIMIT corpus which contains a total of 6300 sentences, ten sentences spoken by each of 630 speakers from eight major dialect regions of the United States, was used in this experiment. Provided time-aligned phonetic transcription was used to extract phonemes from speech samples. A 55-output classifier system was designed corresponding to 55 classes of phonemes and trained with the kernel learning algorithms. The training dataset was extracted from clean training samples. A portion of the database, i.e., 65 338 samples of training dataset, was used to train the system. The performance of the system on the training dataset was 76.4%. The whole test dataset of the TIMIT corpus was used to test the generalization of the system. All samples, i.e., 55 655 samples of the test dataset, were used to test the system. The performance of the system on the test dataset was 45.3%. This approach is currently under development to extend the algorithm for continuous phoneme recognition. [Work supported in part by grants from DARPA, NASA, and ONR.]

1aSC5. Spoken name pronunciation evaluation. Joseph Tepperman and Shrikanth Narayanan (Speech Anal. and Interpretation Lab., USC, 3740 McClintock Ave., Ste. 131, Los Angeles, CA 90089, tepperma@usc.edu)

Recognition of spoken names is an important ASR task since many speech applications can be associated with it. However, the task is also among the most difficult ones due to the large number of names, their varying origins, and the multiple valid pronunciations of any given name, largely dependent upon the speaker's mother tongue and familiarity with the name. In order to explore the speaker- and language-dependent pronunciation variability issues present in name pronunciation, a spoken name database was collected from 101 speakers with varying native languages. Each speaker was asked to pronounce 80 polysyllabic names, uniformly chosen from ten language origins. In preliminary experiments, various prosodic features were used to train Gaussian mixture models (GMMs) to identify misplaced syllabic emphasis within the name, at roughly 85% accuracy. Articulatory features (voicing, place, and manner of articulation) derived from MFCCs were also incorporated for that purpose. The combined prosodic and articulatory features were used to automatically grade the quality of name pronunciation. These scores can be used to provide meaningful feedback to foreign language learners. A detailed description of the name database and some preliminary results on the accuracy of detecting misplaced stress patterns will be reported.

1aSC6. Determining the threshold for usable speech within co-channel speech with the SPHINX automated speech recognition system. William T. Hicks and Robert E. Yantorno (Speech Processing Lab., Elec. and Comput. Eng. Dept., College of Eng., Temple Univ., Philadelphia, PA 19122, bill.hicks@ieee.org)

Much research has been and is continuing to be done in the area of separating the original utterances of two speakers from co-channel speech. This is very important in the area of automated speech recognition (ASR), where the current state of technology is not nearly as accurate as human listeners when the speech is co-channel. It is desired to determine what types of speech (voiced, unvoiced, and silence) and at what target to interference ratio (TIR) two speakers can speak at the same time and not reduce speech intelligibility of the target speaker (referred to as usable speech). Knowing which segments of co-channel speech are usable in ASR can be used to improve the reconstruction of single speaker speech. Tests were performed using the SPHINX ASR software and the TIDIGITS database. It was found that interfering voiced speech with a TIR of 6 dB or greater (on a per frame basis) did not significantly reduce the intelligibility of the target speaker in co-channel speech. It was further found that interfering unvoiced speech with a TIR of 18 dB or greater (on a per frame basis) did not significantly reduce the intelligibility of the target speaker in co-channel speech.

1aSC7. Noise-robust speech recognition using multi-band spectral features. Yoshitaka Nishimura, Takahiro Shinozaki, Koji Iwano, and Sadaoki Furui (Tokyo Inst. of Technol., Dept. of Comput. Sci., 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552, Japan)

In most of the state-of-the-art automatic speech recognition (ASR) systems, speech is converted into a time function of the MFCC (Mel Frequency Cepstrum Coefficient) vector. However, the problem with using the MFCC is that noise effects spread over all the coefficients even when the noise is limited within a narrow frequency band. If a spectrum feature is directly used, such a problem can be avoided and thus robustness against noise could be expected to increase. Although various researches on using spectral domain features have been conducted, improvement of recognition performances has been reported only in limited noise conditions. This paper proposes a novel multi-band ASR method using a new log-spectral domain feature. In order to increase the robustness, log-spectrum features are normalized by applying the three processes: subtracting the mean log-energy for each frame, emphasizing spectral peaks, and subtracting the log-spectral mean averaged over an utterance. Spectral component likelihood values in each frame are weighted by normalized spectral level (spectral peaks) or SNR of each component. Experimental results using babble noise-added speech show that recognition performance is improved by the proposed method in comparison with the MFCC-based method. The performance is further improved by spectral-peak weighting and SNR-based frequency-band weighting techniques.

1aSC8. On audio-visual synchronization for viseme-based speech synthesis. Jianxia Xue, Abeer Alwan (Dept. of Elec. Eng., UCLA, 405 Hilgard Ave., Los Angeles, CA 90095), Edward T. Auer, Jr., and Lynne E. Bernstein (House Ear Inst., Los Angeles, CA 90057)

In viseme-based visual speech synthesis [e.g., T. Ezzat and T. Poggio, Int. J. Comput. Vision **38**, 45–57 (2000)], it is frequently assumed that each viseme can be represented by an image which corresponds to the temporal mid-point of the phoneme's acoustic segment. We investigated the validity of this assumption in a study of 3-D motion data extracted from marker positions on the face of one talker. Using principle components analysis (PCA), the first five principle motion tracks comprising the lips and the jaw were studied. The temporal positions of the local extrema in the principle motion tracks were analyzed. Statistics of the distances between these local extrema and the mid-points of the corresponding acoustic segments were compared. Results showed that the local extrema of the motion tracks were, for the most part, not well-aligned to the phoneme's acoustic midpoint with a few exceptions. For example, for /s/ the local extrema were well aligned (small variance), while for /m/ and /f/, alignments pattern could be found for several, but not all, tokens. The results suggest that phonetic context and speaking rate must be taken into account when characterizing facial configurations in visual speech synthesis. [Work supported in part by the NSF.]

1aSC9. Emotion to emotion speech conversion in phoneme level. Murtaza Bulut, Serdar Yildirim, Carlos Busso, Chul Min Lee, Ebrahim Kazemzadeh, Sungbok Lee, and Shrikanth Narayanan (Univ. of Southern California, Los Angeles, CA 90089)

Having an ability to synthesize emotional speech can make human–machine interaction more natural in spoken dialogue management. This study investigates the effectiveness of prosodic and spectral modification in phoneme level on emotion-to-emotion speech conversion. The prosody modification is performed with the TD-PSOLA algorithm (Moulines and Charpentier, 1990). We also transform the spectral envelopes of source phonemes to match those of target phonemes using LPC-based spectral transformation approach (Kain, 2001). Prosodic speech parameters ($F0$, duration, and energy) for target phonemes are estimated from the statistics obtained from the analysis of an emotional speech database of happy, angry, sad, and neutral utterances collected from actors. Listening experiments conducted with native American English speakers indicate that the modification of prosody only or spectrum only is not sufficient to elicit targeted emotions. The simultaneous modification of both prosody and spectrum results in higher acceptance rates of target emotions, suggesting that not only modeling speech prosody but also modeling spectral patterns that reflect underlying speech articulations are equally important to synthesize emotional speech with good quality. We are investigating suprasegmental level modifications for further improvement in speech quality and expressiveness.

1aSC10. Study of acoustic correlates associate with emotional speech. Serdar Yildirim, Sungbok Lee, Chul Min Lee, Murtaza Bulut, Carlos Busso, Ebrahim Kazemzadeh, and Shrikanth Narayanan (Univ. of Southern California, Los Angeles, CA 90089)

This study investigates the acoustic characteristics of four different emotions expressed in speech. The aim is to obtain detailed acoustic knowledge on how a speech signal is modulated by changes from neutral to a certain emotional state. Such knowledge is necessary for automatic emotion recognition and classification and emotional speech synthesis. Speech data obtained from two semi-professional actresses are analyzed and compared. Each subject produces 211 sentences with four different emotions; neutral, sad, angry, happy. We analyze changes in temporal and acoustic parameters such as magnitude and variability of segmental duration, fundamental frequency and the first three formant frequencies as a function of emotion. Acoustic differences among the emotions are also explored with mutual information computation, multidimensional scaling and acoustic likelihood comparison with normal speech. Results indicate that speech associated with anger and happiness is characterized by longer duration, shorter interword silence, higher pitch and rms energy with wider ranges. Sadness is distinguished from other emotions by lower rms energy and longer interword silence. Interestingly, the difference in formant pattern between [happiness/anger] and [neutral/sadness] are better reflected in back vowels such as /a/(/father/) than in front vowels. Detailed results on intra- and interspeaker variability will be reported.

1aSC11. Analysis and synthesis of laughter. Shiva Sundaram and Shrikanth Narayanan (Speech Anal. and Interpretation Lab., Univ. of Southern California, 3740 McClintock Ave., EEB400, Los Angeles, CA 90089, ssundara@usc.edu)

There is much enthusiasm in the text-to-speech community for synthesis of emotional and natural speech. One idea being proposed is to include emotion dependent paralinguistic cues during synthesis to convey emotions effectively. This requires modeling and synthesis techniques of various cues for different emotions. Motivated by this, a technique to synthesize human laughter is proposed. Laughter is a complex mechanism of expression and has high variability in terms of types and usage in human–human communication. People have their own characteristic way of laughing. Laughter can be seen as a controlled/uncontrolled physiological process of a person resulting from an initial excitation in context. A parametric model based on damped simple harmonic motion to effectively capture these diversities and also maintain the individuals characteristics is developed here. Limited laughter/speech data from actual humans and synthesis ease are the constraints imposed on the accuracy of the model. Analysis techniques are also developed to determine the parameters of the model for a given individual or laughter type. Finally, the effectiveness of the model to capture the individual characteristics and naturalness compared to real human laughter has been analyzed. Through this the factors involved in individual human laughter and their importance can be better understood.

1aSC12. Effects of emotion on different phoneme classes. Chul Min Lee, Serdar Yildirim, Murtaza Bulut, Carlos Busso, Abe Kazemzadeh, Sungbok Lee, and Shrikanth Narayanan (Univ. of Southern California, Los Angeles, CA 90089)

This study investigates the effects of emotion on different phoneme classes using short-term spectral features. In the research on emotion in speech, most studies have focused on prosodic features of speech. In this study, based on the hypothesis that different emotions have varying effects on the properties of the different speech sounds, we investigate the usefulness of phoneme-class level acoustic modeling for automatic emotion classification. Hidden Markov models (HMM) based on short-term spectral features for five broad phonetic classes are used for this purpose using data obtained from recordings of two actresses. Each speaker produces 211 sentences with four different emotions (neutral, sad, angry, happy). Using the speech material we trained and compared the performances of two sets of HMM classifiers: a generic set of ''emotional speech'' HMMs (one for each emotion) and a set of broad phonetic-class based HMMs (vowel, glide, nasal, stop, fricative) for each emotion type considered. Comparison of classification results indicates that different phoneme classes were affected differently by emotional change and that the vowel sounds are the most important indicator of emotions in speech. Detailed results and their implications on the underlying speech articulation will be discussed.

1aSC13. Representation for dialect recognition using topographic independent component analysis. Qu wei (Dept. of Information Technol., Univ. of Sci. and Technol. of Beijing in China, 100084, Anhuili, Chaoyang District, Beijing, China)

In dialect speech recognition, the feature of tone in one dialect is subject to changes in pitch frequency as well as the length of tone. It is beneficial for the recognition if a representation can be derived to account for the frequency and length changes of tone in an effective and meaningful way. In this paper, we propose a method for learning such a representation from a set of unlabeled speech sentences containing the features of the dialect changed from various pitch frequencies and time length. Topographic independent component analysis (TICA) is applied for the unsupervised learning to produce an emergent result that is a topographic matrix made up of basis components. The dialect speech is topographic in the following sense: the basis components as the units of the speech are ordered in the feature matrix such that components of one dialect are grouped in one axis and changes in time windows are accounted for in the other axis. This provides a meaningful set of basis vectors that may be used to construct dialect subspaces for dialect speech recognition.

1aSC14. Robust speaker recognition using glottal information-based cepstral mean subtraction. Pu Yang, Yingchun Yang, and Zhaohui Wu (Zheda Rd. 38#, Hangzhou, Zhejiang Province, Peoples' Republic of China)

Channel distortion and background noise often severely degrade the performance of automatic speaker recognition (ASR) system. In this paper, a new compensation method called glottal information-based cepstral mean subtraction (GIBCMS), which improves upon the conventional cepstral mean subtraction (CMS) method, is presented. Besides the cepstral

information, GIBCMS has utilized the speaker's glottal information, which also holds speaker-dependent characteristics, but is less vulnerable to environment than the cepstral one. In order to test its robustness under channel distortion, even with high level of background noise, we applied this method to the SRMC corpus which is added by noise at different SNRs. The experimental results show that GIBCMS gains better performance over other improved CMS methods on it.

1aSC15. Systematic speaker variation and within-speaker center of gravity correlations in the TIMIT database. Hector R. Javkin, Carol Christie, Gaston R. Cangiano, Elaine Drom (Dept. of Linguist. and Lang. Development, San Jose State Univ., San Jose, CA 95192-0093), and Katia McClain (Univ. of California, Santa Barbara, CA)

The systematicity of speaker variation in consonants was examined by measuring the noise component of consonants in the TIMIT database. Fricatives and stops were compared by measuring at the temporal middle of fricatives and at the release of stops. There is a high correlation between many of the center of gravity measures. For example, if a speaker has a particularly high center of gravity in the sound /t/, she or he will also have a high center of gravity in /d/, /s/, /p/, /k/, and other consonants. The full set of correlations will be described in the paper. The correlations appear to stem from individual differences and not from dialect variations. The implications of the results for rapid speaker adaptation in speech recognition will be explored. [Work supported by San Jose State University Faculty Grant.]

1aSC16. Enhanced energy redistribution speech intelligibility algorithm with real-time implementation. Manasa Raghavan, Mark D. Skowronski, and John G. Harris (Computational NeuroEng. Lab, Elec. and Comput. Eng., Univ. of Florida, Gainesville, FL)

Speech intelligibility enhancement is a concern for mobile platforms operating in noisy environments. Current noise-reduction techniques, such as subspace methods and spectral subtraction, operate on speech corrupted by acoustic and transmission noise. Yet preprocessing techniques, which operate on clean speech before noise corruption, have received little attention. Previously, the authors have developed the energy redistribution algorithm [J. Acoust. Soc. Am. **112**, 2305 (2002)], which, based on characteristics of clear speech as well as the Lombard effect, redistributes energy in time from voiced regions to unvoiced regions of speech. The algorithm is designed efficiently for real-time implementation, and in this work the algorithm is demonstrated on a mobile platform, TIs TMS320C6713 DSK board. Furthermore, two enhancements to the algorithm are introduced: (1) a variable unvoiced gain factor and (2) a high pass filter (HPF). The variable unvoiced gain factor adjusts the amount of energy redistributed, and the HPF, a compact algorithm shown previously to enhance clean speech in noisy environments [Niederjohn and Grotelueschen, IEEE Trans. Acoust., Speech, Signal Process. **24**(5), 277–282 (1976)], is tested on sentences from the Hearing in Noise Test (HINT) corrupted by speech-shaped noise. Results show improved speech intelligibility for both enhancements.

1aSC17. User expertise in speech centered multimodal human computer interaction. Rajesh Chandrasekaran, Sorin Dusan, and James L Flanagan (CAIP Ctr., 96, Frelinghuysen Rd., Piscataway, NJ 08855)

Multimodal interfaces aim to permit natural communication by speech and gesture. Typically the speech modality bears the principal information in the interaction with gesture complementing spoken commands. A continuing challenge is how to correlate and interpret the simultaneous inputs to estimate meaning and user intent. User expertise and familiarity figure prominently in the interpretation. The present research studies the effect of user expertise on multimodal human computer interaction. Users are classified into experienced and inexperienced depending on the amount of their exposure and interaction with multimodal systems. Each user is asked to perform simple tasks using a multimodal system. For each task the automatically recognized speech input is time stamped and the lag or lead of the gesture input is computed with respect to this time stamp. The time interval around the time stamp in which all the users' gesture inputs occur is determined. For experienced users this interval averages 56.9% less than that for inexperienced users. The implication is that for experienced users the spoken input are the corresponding gesture input are more closely related in time than for inexperienced users. This behavior can be exploited in multimodal systems to increase efficiency and reduce time of response for the system.

1aSC18. Using Flanagan's phase vocoder to improve cochlear implant performance. Fan-Gang Zeng (364 Med Surge II, Univ. of California, Irvine, CA 92697)

The cochlear implant has restored partial hearing to more than 100 000 deaf people worldwide, allowing the average user to talk on the telephone in quiet environment. However, significant difficulty still remains for speech recognition in noise, music perception, and tonal language understanding. This difficulty may be related to speech processing strategies in current cochlear implants that emphasized the extraction and encoding of the temporal envelope while ignoring the temporal fine structure in speech sounds. A novel strategy was developed based on Flanagan's phase vocoder [Flanagan and Golden, Bell Syst. Tech. **45**, 1493–1509 (1966)], in which frequency modulation was extracted from the temporal fine structure and then added to amplitude modulation in the current cochlear implants. Acoustic simulation results showed that amplitude and frequency modulation contributed complementarily to speech perception with amplitude modulation contributing mainly to intelligibility whereas frequency modulation contributed to speaker identification and auditory grouping. The results also showed that the novel strategy significantly improved cochlear implant performance under realistic listening situations. Overall, the present result demonstrated that Flanagan's classic work on phase vocoder still shed insight on current problems of both theoretical and practical importance. [Work supported by NIH.]

1aSC19. Separation of formants from glottal impulses in high-resolution spectrograms using higher-order short-time Fourier transform (STFT) phase derivatives. Sean A. Fulop (Linguist. Dept., Cal. State Univ., Fresno, CA 93740-8001)

The *channelized instantaneous frequency* of a signal $x(T)$ is $\mathrm{CIF}_x(\omega,T)=(\partial/\partial T)\arg(F_h(\omega,T))$, where F_h is the short-time Fourier transform of $x(T)$ using window function h. The *local group delay* of a signal is $\mathrm{LGD}_x(\omega,T)=-(\partial/\partial\omega)\arg(F_h(\omega,T))$. For each point $F_h(\omega_0,T_0)$ in the STFT, the $f-t$ coordinates $[\mathrm{CIF}(\omega_0,T_0),t-\mathrm{LGD}(\omega_0,T_0)]$ pinpoint the local mean of the Rihacek distribution of complex signal energy, and this reassignment of the STFT magnitude yields a high-resolution spectrogram without blurriness [Auger and Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," IEEE Trans. Signal Process **43**(5), 1068–1089 (1995)]. Nelson ["Instantaneous higher order phase derivatives," Digital Signal Process **12**, 416–428 (2002)] demonstrated that the significant AM/FM components of a signal $x(T)$ have $(\partial/\partial\omega)\mathrm{CIF}_x(\omega,T)\approx 0$. By plotting just those STFT points meeting this condition to within a threshold, a spectrogram showing only speech resonances can be drawn using short analysis frames. By moving to a long-frame analysis, the resonance tracker becomes an effective pitch and harmonic tracker. Nelson further demonstrated that the impulses in a signal $x(T)$ have $(\partial/\partial T)\mathrm{LGD}_x(\omega,T)\approx 1$. By plotting just those points meeting this condition to within a threshold, a spectrogram of speech impulses can alternatively be drawn.

Session 1pAA

Architectural Acoustics and Signal Processing in Acoustics: Topical Meeting on Spatial and Binaural Evaluation of Performing Arts Spaces II: Auralization and Simulation

Mendel Kleiner, Cochair
School of Architecture, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180-3590

Rendell R. Torres, Cochair
Rennsselaer Polytechnic Institute, School of Architecture, 110 8th Street, Troy, New York 12180-3590

Invited Papers

Chair's Introduction—1:20

Invited Papers

1:25

1pAA1. Evaluation of room acoustic qualities and defects by use of auralization. Jens H. Rindel (Acoust. Technol., Oersted DTU, Tech. Univ. of Denmark, Bldg. 352, DK-2800 Kgs. Lyngby, Denmark, jhr@oersted.dtu.dk)

Auralizations generated by room acoustic computer modeling programs may be used as a tool for evaluation of acoustic qualities and defects, some of which are not easily detected by objective measures. Examples include reverberance, flutter echoes, perceived room size and distance, apparent source width, listener envelopment, and sound propagation in coupled rooms. In order to reach a sufficiently high level of realism in auralizations for such room acoustic applications, it is necessary that all parts of the chain are modeled with certain accuracy. In particular, it is found that the modeling of source directivity and the late room reflections (the reverberation tail) need careful consideration in order to achieve reliable and realistic-sounding results. However, when implemented in the software the practical use can be simple and quick.

1:45

1pAA2. Uncertainties in measurement of spatial parameters in room acoustics. Michael Vorlaender and Ingo Witew (Inst. of Tech. Acoust., RWTH Aachen Univ., 52056 Aachen, Germany, mvo@akustik.rwth-aachen.de)

With the goal to assess spatial impression, such as apparent source width and envelopment in concert halls, lateral energy fraction is a fairly simple measure to be determined as a single number parameter. ISO 3382 describes the procedure to conduct the measurements and also defines certain quality requirements imposed on the loudspeakers and the microphones. This case study portrays how changes in the source and the microphone character are reflected in the parameter. This aspect is investigated on the grounds of measurements and simulations. In a measurement scenario, LF is calculated for different sound sources, different source setups and different microphone arrangements. In simulations, the effects of variations of the directional pattern of the microphone on the measurement result are investigated. The results suggest that despite the proven background of the lateral fraction parameter, the measurement results are distorted with noteworthy uncertainties that originate from deficiencies in measurement equipment.

2:05

1pAA3. The value of utilizing binaural dummy head recordings in evaluating physical acoustic changes in concert halls. Christopher Jaffe, Russell Cooper, and Carlos Rivera (Jaffe Holden Acoust., Inc., 114A Washington St., Norwalk, CT 06854)

In some instances, after a concert hall is built there may be a need to modify the physical environment of the space through the application of diffusion or absorptive surfaces, the addition of reflector systems or the repositioning of the orchestra in the space. Prior to moving forward with suggested changes to the physical environment, it has been customary to conduct evaluation rehearsals with physical mock-ups installed to confirm the acousticians recommendations. Questionnaires are given to the musicians, the conductor and the administration staff to document the effect of the changes, and physical measurements are taken before and after the installation of the mock-ups. The questionnaires can be difficult to correlate and the differences in data resulting from the physical measurements may be too small to properly evaluate. More recently, Jaffe Holden Acoustics has added dummy head recordings to the mix. These recordings are extremely representative of what a human hears and one can place these devices in various locations on stage and in the audience chamber. The recordings create a permanent record of the event and the results of subsequent A/B evaluation can be more closely correlated to render judgments.

2:25

1pAA4. High-resolution spatial auralization techniques for room acoustics design. Paul D. Henderson, Rendell R. Torres, and Ioana N. Pieleanu (Program in Architectural Acoust., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, hendep2@rpi.edu)

Historically, consultants have avoided the use of computed auralizations when evaluating critical architectural design components due to uncertainties in the physical validity and perceptual accuracy of simulations. This work demonstrates the use of advanced techniques for increasing the confidence of auralized results, applied to the acoustical design of the Experimental Media and Performing Arts Center (EMPAC), currently under construction at Rensselaer. The 1200-seat performance hall in this facility contains an acoustically transmissive overhead canopy, which creates potential difficulties for conventional simulation techniques, specifically in the modeling of early reflected and transmitted sound paths. To overcome these challenges, a new ray-tracing engine was developed which introduces a ray-spawning method in the search for deterministic sound transmission paths in the early impulse response. In addition, this method allows for surface reflection and transmission characteristics to be defined with both arbitrary frequency resolution and with dependence on incidence angle. The resulting early-part simulations are coupled with statistically computed late reverberation. These auralizations are then rendered using second-order Ambisonics over a dodecahedron loudspeaker array, recreating a full-sphere sound field for the listener. In addition, the *in situ* localization performance of human subjects using this rendering system is reviewed. [Work supported by EMPAC at Rensselaer Polytechnic Institute.]

2:45

1pAA5. Acoustical simulation based on head-tracked auralization and measured high-resolution head related transfer function (HRTF) and impulse responses. Wolfgang Ahnert, Stefan Feistel, and Christoph Moldrzyk (ADA Acoust. Design Ahnert, Arkonastr. 45-49, D13189 Berlin, Germany)

A very desirable feature of modern acoustical simulation programs is the easy, fast and reliable auralization of the prediction results. In this paper we consider a new auralization method, based on a head-tracked headphone system with high spatial resolution and real-time convolution. We discuss the way to measure the directional head-related transfer functions, the calculation of the directional binaural impulse responses and the realization as a real-time convolution software. Furthermore, high-resolution impulse responses have been measured to compare reality, measurement and prediction results of an example room. The measurements were performed with a newly developed software tool EASERA. We conclude that, using this new method, auralization results are obtained equivalent to the human perception in reality.

3:05–3:15 Break

3:15

1pAA6. Capturing the acoustic response of historical spaces for interactive music performance and recording. Wieslaw Woszczyk and William Martens (McGill Univ., 555 Sherbrooke St. W., Montreal, QC H3A 1E3, Canada)

Performers engaged in musical recording while they are located in relatively dry recording studios generally find their musical performance facilitated when they are provided with synthetic reverberation. This well established practice is extended in the project described here to include highly realistic virtual acoustic recreation of original rooms in which Haydn taught his students to play pianoforte. The project has two primary components, the first of which is to capture for posterity the acoustic response of such historical rooms that may no longer be available or functional for performance. The project's second component is to reproduce as accurately as possible the virtual acoustic interactions between a performer and the re-created acoustic space, as performers, during their performance, move relative to their instrument and the boundaries of surrounding enclosure. In the first of two presentations on this ongoing project, the method for measurement of broadband impulse responses for these historical rooms is described. The test signal is radiated by a group of omnidirectional loudspeakers approximating the layout and the complex directional radiation pattern of the pianoforte, and the room response is sampled by a spaced microphone array. The companion presentation will describe the method employed for virtual acoustic reproduction for the performer.

3:35

1pAA7. Virtual acoustic reproduction of historical spaces for interactive music performance and recording. William Martens and Wieslaw Woszczyk (McGill Univ., 555 Sherbrooke St. W., Montreal, QC H3A 1E3, Canada)

For the most authentic and successful musical result, a performer engaged in recording pianoforte pieces of Haydn needs to hear the instrument as it would have sounded in historically typical room reverberation, such as that of the original room's in which Haydn taught his students to play pianoforte. After capturing the acoustic response of such historical room's, as described in the companion presentation, there remains the problem of how best to reproduce the virtual acoustical response of the room as a performer moves relative to the instrument and the rooms boundaries. This can be done with a multichannel loudspeaker array enveloping the performer, interactively presenting simulated indirect sound to generate a sense of presence in the previously captured room. The resulting interaction between live musical instrument performance and the sound of the virtual room can be captured both binaurally for the performer's subsequent evaluation, readjusted to provide the most desirable acoustic feedback to the performer, and finally remixed for distribution via conventional 5.1 channel audio media.

1pAA8. Differences in directional sound source behavior and perception between assorted computer room models. Michelle C. Vigeant, Lily M. Wang (Architectural Eng. Program, Univ. of Nebraska-Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, mvigeant@unlnotes.unl.edu), and Jens Holger Rindel (Tech. Univ. of Denmark, DK-2800 Kgs. Lyngby, Denmark)

Source directivity is an important input variable when using room acoustic computer modeling programs to generate auralizations. Previous research has shown that using a multichannel anechoic recording can produce a more natural sounding auralization, particularly as the number of channels is increased [J. H. Rindel, F. Otondo, and C. L. Christensen, Proceedings of the International Symposium on Room Acoustics: Design and Science 2004, Paper V01 (2004)]. Further studies evaluating the quality of auralizations using one-channel, four-channel, and 13-channel anechoic recordings have been pursued. The effect of changing the room's material properties was studied in relation to turning the source around 180 deg and on the range of acoustic parameters from the four- and 13 beams. As the room becomes increasingly diffuse, the importance of the modeled directivity decreases when considering reverberation time. However, for the three other parameters evaluated (sound-pressure level, clarity index, and lateral fraction), the changing diffusivity of the room does not diminish the importance of the directivity. The study therefore shows the importance of considering source directivity when using computer modeling. [Work supported by the National Science Foundation.]

Contributed Papers

4:15

1pAA9. Mapping the sound field of the Lyric Theater. Michael Ermann, Michael Kidner, Joe McCoy, Dan Mennitt (Virginia Tech; 201 Cowgill Hall (0205), Blacksburg, VA 24061), Adam Tawney, and Bill Yoder (Virginia Tech)

Impulse response measurements were taken in each of the 235 orchestra-level house-left seats of the Lyric Theater in order to explore spatial differences in the sound field of the room. These measurements produced graphic mappings depicting gradients of sound decay time and loudness. Significant interseat variations appeared when mapping the decay times for the 63- and 125-Hz octave bands; these variations largely disappeared when mapping the decay times for octave bands greater than 250 Hz. A scaled physical model and a CATT-Acoustics software model of the space were built to explore the veracity and usefulness of modeling to determine intraroom spatial differences in sound fields.

4:30

1pAA10. Interaural cross-correlation coefficient, apparent source width, and time/frequency characteristics of binaural running cross-correlation. David A. Conant (McKay Conant Brook Inc., 5655 Lindero Cyn. Rd. #325, Westlake Village, CA 91362, dconant@mcbinc.com)

The interaural cross-correlation coefficient (IACC) descriptors attempt to characterize in a single number the binaural cross-correlation function at the expense of losing much detail in a manner similar to, but more problematic than, such descriptors as NC, STC, and NRC because much frequency- and rich time-domain information is lost. Recent computer simulations by Mason illuminated these problems and further stimulated the present study. Old binaural impulse recordings run through an analog computer algorithm (drawn from physiological modeling) were re-examined in an attempt to reveal trends and relationships that the current descriptors cannot. Issues of the time and frequency variance of the measured running coefficient are discussed in terms of apparent source width (ASW) for music in Troy Music Hall and simpler spaces.

4:45

1pAA11. Binaural simulation of virtual stage environments for evaluation by the solo violinist. William Chu (Rensselaer Polytechnic Inst., Troy, NY 12180)

A unique computational and testing model for assessing solo violinists' response to various acoustical conditions on-stage was developed and shown to be an effective study tool. Realtime binaural auralizations of CATT-Acoustic models of several different virtual stage-acoustic designs were examined with a small group of professional violinists. Perceived differences as expressed by the subjects were discussed and explored in light of the specific measured acoustical descriptors available through the model. Unlike acoustical descriptors developed for audience receivers, ease of playing, support, and other soloist concerns do not necessarily show strong correlation between different subjects. It was found that this binaural simulation technique could provide a basis for developing a personal contextual vocabulary, to better understand the desired acoustical response of individual musicians.

Session 1pAOa

Acoustical Oceanography: Acoustic Characterization of the Benthic Habitat

Kenneth G. Foote, Chair
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Chair's Introduction—1:00

Invited Paper

1:05

1pAOa1. Geographic Information Systems (GIS) analysis of high-resolution multibeam bathymetry and remotely operated vehicle data to model rockfish habitat preference. Erica Summers-Morris, Pat Iampietro, and Rikk Kvitek (Seafloor Mapping Lab, CSU Monterey Bay, 100 Campus Ctr., Seaside, CA 93955)

Management of declining rockfish stocks requires effective tools capable of providing accurate stock assessments of near-shore, high-relief habitat. Multibeam bathymetry, when analyzed with GIS landscape analysis tools, can create models which can identify preferred habitat based on species-specific parameters. For this study, high-resolution multibeam data of the Del Monte shale beds in Monterey Bay, California were analyzed in GIS for slope, rugosity, and relative topographic position to assess rockfish (*Sebastes spp.*) habitat preference. Video transects collected by a remotely operated vehicle (ROV) provided habitat ground-truth and fish distribution data. A series of habitat suitability models was created in GIS by combining different suitability factors from multibeam-derived grids: slope, rugosity, topographic position index (TPI), and depth. Distance to preferred categories for each parameter were determined for eight rockfish species. Of these, distance to peak features identified by TPI50 proved the most effective means of modeling fish distribution, successfully predicting an average of 80% of the eight rockfish species. Using fish distribution information, stock estimates were calculated for the study area. By combining GIS landscape analysis tools with multibeam bathymetry and ROV video data, we have created a predictive tool that can locate areas of most suitable habitat given rockfish-specific parameters.

Contributed Papers

1:30

1pAOa2. Quantitative benthic habitat characterization at Heceta Bank, Oregon. Curt E. Whitmire, W. Waldo Wakefield (NOAA Fisheries Northwest Fisheries Sci. Ctr., 2725 Montlake Blvd. E, Seattle, WA 98112-2097), Robert W. Embley (NOAA Pacific Marine Environ. Lab., Newport, OR 97365), Susan G. Merle (Cooperative Inst. for Marine Resource Studies, Newport OR 97365), Brian N. Tissot, and Noelani Puniwai (Washington State Univ. Vancouver, Vancouver, WA 98686)

Recent studies have highlighted the shortcomings of regional trawl surveys for quantifying abundance of demersal fishes in rugged habitats. Many species show strong affinities to areas of high topographic relief and rugosity, therefore precluding sampling by bottom trawl gears. Consequently, advanced survey technologies such as submersibles and camera sleds have been utilized to survey demersal fishes *in situ*. One prerequisite in extrapolating densities of demersal fishes observed *in situ* is the mapping of benthic habitats beyond the observational extent of submersibles. Habitat characteristics observed from submersible and ROV dives in the late 1980s, 2000, and 2001 were used to establish habitat classes that have been statistically shown to correlate with demersal fish distributions. Those habitat characteristics were then extrapolated over the extent of a multibeam sonar survey conducted in 1998 using quantitative parameters derived from high-resolution bathymetric and acoustic backscatter data. The resultant map predicts the coverage of four lithologic habitat classes: mud (unconsolidated), sand (unconsolidated), boulder/cobble (high acoustic reflectivity), and rock outcrop (high vertical relief). Those four habitat classes will facilitate the extrapolation of fish densities to the larger spatial scales at which resident fish populations and physiographic provinces occur.

1:45

1pAOa3. Relationships between seafloor substrates, benthic epifauna, and spatial properties of multibeam echosounder bathymetry and backscatter data. George R. Cutter, Jr., Larry A. Mayer (UNH Ctr. for Coastal and Ocean Mapping, 24 Colovos Rd., Durham, NH 03824, gcutter@cisunix.unh.edu), and Larry G. Ward (UNH, Durham, NH 03824)

The seafloor in the mouth of the Piscataqua River Estuary, near Portsmouth, New Hampshire, contains several geomorphological regions that are evident in bathymetric data from multibeam sonar soundings. Those morphological regions generally are associated with surficial sediment type classes previously identified from sediment samples. We demonstrate that the morphologies are distinguishable using parameters from models of spatial variation for gridded depth soundings and backscatter values. Ground-truth seafloor-video transects data suggest that apparent facies and morphological regions are associated with different surficial sediment classes and rates of facies transitions. We show that the spatial variability of depth and backscatter is associated with the substrate transition rate, the number of microhabitats (microfacies), and distribution of certain epifauna identified using video. We show that, in some cases, a single region defined based upon spatial morphological attributes from depth data and apparent sedimentological attributes from backscatter data represented several benthic habitats. The results suggest that although we can use spatial variability of depth or backscatter to segment the seafloor into apparent facies or habitat class regions, some of the regions will have unpredicted associated microhabitats and organisms. The results have implications for habitat-specific assessment of organisms.

2:00

1pAOa4. Use of a 600-kHz ADCP to characterize submerged aquatic vegetation in a very shallow estuary. Joseph D. Warren and Bradley J. Peterson (Southampton College of Long Island Univ., 239 Montauk Hwy., Southampton, NY 11968, joe.warren@liu.edu)

Submerged aquatic vegetation (SAV) serves as a key habitat for the larval forms of many commercially important marine organisms. In coastal waters, anthropogenic factors can cause severe damage and loss to this habitat. In order to measure changes in the spatial coverage of SAV, aerial photography or diver surveys are required. We investigated an alternative approach using a high-frequency (600 kHz) acoustic Doppler current profiler to make high-resolution (in the vertical) measurements of the scattering in several very shallow bays in Long Island, NY. These data were used to determine the presence and amount of SAV in an area. Diver surveys provided ground-truthing of these data. Field measurements are also compared with an acoustic scattering model for eel grass (a common SAV type in the northeast United States). The purpose of this project was to determine whether acoustic sampling of benthic vegetation could accurately determine species type and coverage. If this is possible, then acoustic sampling could provide maps of spatial and temporal changes in SAV distribution and abundance. [Work was supported by an award from UNH CICEET.]

2:15

1pAOa5. Acoustic scattering by benthic shells: Dominant scattering mechanisms and applications. Timothy K. Stanton and Dezhang Chu (Woods Hole Oceanograph. Inst., Dept. of Appl. Ocean Phys. and Eng., Woods Hole, MA 02543, tstanton@whoi.edu)

When benthic shells occur in sufficiently large numbers, they can dominate acoustic backscattering by the seafloor, especially at angles of incidence away from normal. In order to use sound as a tool to remotely detect and quantify the shells, the scattering properties of the shells need to be understood, both in free-space as well as when placed on the seafloor. Through laboratory experimentation, it has been determined that the edges of certain types of shells (such as bivalves and sand dollars) can dominate the scattering over an important range of grazing angles. The surfaces of these shells and others dominate under other conditions. The dominant scattering effects are discussed in the context of interpreting acoustic backscatter data in terms of meaningful parameters such as numerical density of the shells.

2:30

1pAOa6. Acoustic mapping of squid egg clusters and their bottom habitat in Monterey Bay, California. Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Roger T. Hanlon (Marine Biological Lab., Woods Hole, MA 02543), Pat J. Iampietro, and Rikk G. Kvitek (California State Univ., Seaside, CA 93955)

Clusters of gelatinous egg capsules, known as mops or beds, of the market squid (*Loligo opalescens*) were mapped in a shallow-water, sandy habitat of Monterey Bay, California. The benthic egg clusters were imaged using an EdgeTech 272-TD dual-frequency sidescan sonar towed from R/V MACGINITIE, an 8-m-long survey vessel, with data recorded on a Triton Elics International Isis digital data acquisition system. Verification of target identity was accomplished independently by video photography from a remotely operated vehicle. The survey area included a 4-km stretch of sandy seafloor between Lover's Point and Cannery Row in Monterey at depths of 15–30 m. The study area had previously been mapped using the RESON SeaBat 8101 240-kHz multibeam sonar. Resulting high-resolution bathymetric data, with 1-m resolution, were used during the survey planning and execution. Squid egg clusters were clearly visible in the very-high-resolution, 400-kHz backscatter imagery, with pixel size 10–20 cm, recorded from the towed sidescan sonar. The concentration of egg clusters was greatest along a sloping feature believed to be a submarine fault. Egg mops with diameter as small as 0.5 m were distinguishable. [Support by Sea Grant is acknowledged.]

2:45

1pAOa7. How useful is bathymetric information in the classification of high frequency sonar surveys? Louis Atallah (Inst. of Informatics, The Br. Univ. in Dubai/ The Univ. of Edinburgh, P.O. Box 502216 Dubai, UAE) and Penny Probert Smith (Univ. of Oxford, Oxford, OX1 3PJ, UK)

In several sonar studies, bathymetric information; is used for the correction of amplitude data and the calculation of backscattering strength, which is plotted versus grazing angle and used for seabed classification. Bathymetric data is also used as an easily viewed backdrop to visualize backscattered sonar data in surveys. This work proposes an automatic method that combines amplitude features (describing backscattering strength and sonar texture) with bathymetric features (indicating seafloor variability) for sonar classification. Features are selected per window (of user defined size) and areas around grab samples in a survey are used for training. The importance of bathymetric features is investigated in this study, and highlighted by feature selection algorithms as well as by scatter plots exploring the training areas. Classification rates are significantly improved when both amplitude and bathymetry features are used. The final results show the classified windows plotted versus their exact position in the survey. The method described in this work is applied to a sidescan bathymetric sonar dataset taken in Hopvagen Bay Norway. The methods are also applicable to other sonars which provide bathymetric information; a multibeam sonar is such an example.

Session 1pAOb

Acoustical Oceanography: Fisheries Acoustics

Orest I. Diachok, Chair
Poseidon Sound, 3272 Fox Mill Road, Oakton, Virginia 22124

Contributed Papers

3:15

1pAOb1. Effects of fish with swim bladders on absorption and scintillation. Orest Diachok (Poseidon Sound, Oakton, VA 22124, orestdia@aol.com)

Bioacoustic absorption spectroscopy (BAS) experiments, which were conducted in the Santa Barbara Channel in 2001 and 2002, were designed to investigate the effects of fish with swim bladders on absorption and scintillation. These experiments included a broadband source, which transmitted a sequence of 65-s-long tones between 0.25 and 10 kHz, and a vertical array which spanned most of the water column. The range was fixed. A fisheries echo sounder and trawls provided bio-acoustic parameters. Strongest absorption lines and highest values of the scintillation index were observed at night at about 1.1 kHz, the resonance frequency of 15 cm long sardines, when they were dispersed at an average depth of 13 m. Smaller absorption lines were correlated with other species. During the day sardines occupied a depth of about 50 m, where their extinction cross sections were diminished; some were dispersed and resonated at the frequency of individuals; others formed schools and resonated at collective frequencies. As a result of these phenomena, absorption lines due to sardines were much weaker, and the effect of this species on the scintillation index was not evident. [Work was supported by ONR.]

3:30

1pAOb2. Measurements and modeling of low-frequency scattering from fish on the New Jersey shelf. Raymond J. Soukup (Naval Res. Lab. Code 7140, Washington, DC 20375), Redwood W. Nero (Naval Res. Lab. Code 7180, Stennis Space Ctr., MS 39529), and Timothy M. Shepard (Univ. of Texas, Austin, TX 78712)

Low-frequency (400–2000 Hz) scattering strength measurements were obtained from aggregations of fish on the New Jersey shelf using a colocated transducer and vertical line array receiver. Concurrent measurements with a fisheries echosounder revealed the presence of nighttime layer(s) and daytime schools. The day–night behavior is consistent with several species of fish found on the New Jersey shelf, as are the observed resonance frequencies near 800 Hz. Data-model comparisons are used to characterize both the strength and bandwidth of the frequency response, as well as its dependence on fish type and behavior. From these data, parameters such as fish depth, density and size can be estimated, and, in turn, the scattering strength as a function of frequency and grazing angle can be predicted and then compared with those of other potential scattering mechanisms. The implications of the fish scattering to reverberation and clutter are discussed. [Work supported by ONR.]

3:45

1pAOb3. Expected Doppler shift and spread in long range scattering from fish schools on the continental shelf. Nicholas C. Makris, Yisan Lai, Ioannis Bertsatos, Deanelle Symonds, Sunwoong Lee, and Purnima Ratilal (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

A stochastic model for the velocities of individual fish in a school, consistent with continous wide area observations [Symonds *et al.*, J. Acoust. Soc. Am. **115**, 2618 (2004)], is developed. This is combined with the model for scattering from a moving target submerged in a stratified ocean waveguide [Lai and Makris, J. Acoust. Soc. Am. **113**, 223–244 (2003)] to estimate the Doppler shift and spread expected in long-range scattering from fish schools on the continental shelf. The spectral characteristics expected for randomly swarming and migrating fish schools are then compared with those for stationary seafloor features and other moving targets such as underwater vehicles.

4:00

1pAOb4. Modeling low-frequency acoustic scattering from densely populated fish schools. Thomas R. Hahn (Rosenstiel School of Marine and Atmospheric Sci. AMP, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

Scattering of sound from a school of swimbladder-bearing fish is clearly related to the scattering from a cloud of air bubbles. From an ''effective'' medium perspective, such bubble clouds can be viewed as a single object with acoustic bulk properties determined from the physical properties of the individual bubbles and the effect of multiple scattering between them. The acoustic modeling at low frequencies of bubble clouds based on the effective medium approach has been successfully applied to bubbles in the ocean and laboratory bubble plumes by a variety of authors. Applicability and limitations of an effective medium approach to scattering of sound waves from fish schools are investigated and compared to other acoustic models. It is argued that, particularly in the regime of highly populated fish schools, where other approaches (e.g., born or random-phase approximation, etc.) fail due to strong multiple scattering effects, the effective medium model can be applied for a wide range of frequencies, fish numbers, and densities.

4:15

1pAOb5. Diurnal trends in the mid-water biomass community of the Northwestern Hawaiian Islands observed acoustically. Marc O. Lammers (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734), Russell E. Brainard (NOAA Fisheries, Honolulu, HI 96814), and Whitlow W. L. Au (Hawaii Inst. of Marine Biol.)

The nighttime mid-water biomass occurring near six banks in the Northwestern Hawaiian Islands was investigated using 38- and 120-kHz EK60 echosounders. Locations investigated included: French Frigate Shoals, Maro Reef, Lisianksi Island/Neva Shoals, Pearl and Hermes Atoll, Kure Atoll and Midway Atoll. Surveys were designed to sample transect lines parallel and normal to shore between 10 p.m. and 7 a.m. and once during daylight hours. A diurnal trend was found in the occurrence of mid-water sound scattering organisms at all six locations. Communities of organisms accumulate at night on the edges of each island between the 20- and 100-fathom isobaths. The highest densities of organisms restrict their horizontal movements to depths of 20 fathoms or deeper, but increases in biomass were also observed at shallower depths. The northern islands of Kure, Midway, and Pearl and Hermes Atolls exhibited patchier distributions than the southern islands. The composition of the biomass is presently unclear but resembles the mesopelagic boundary community found near the Main Hawaiian Islands. Simultaneous observations with the

TOAD camera system revealed clouds of zooplankton mixed with small fish and other micronekton. The nightly influx of these organisms is likely a significant, though poorly understood, component of these islands ecosystems.

4:30

1pAOb6. Broadband acoustic backscatter and high-resolution morphology of fish: Measurement and modeling. D. Benjamin Reeder (Naval Postgrad. School, Monterey, CA 93943), J. Michael Jech (NOAA/NMFS Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543), and Timothy K. Stanton (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Remote detection and classification of marine organisms require careful acoustic backscattering measurements and acoustic scattering models that adequately incorporate morphology. A study has been conducted on alewife (*Alosa pseudoharengus*), a swimbladder-bearing fish, which incorporated broadband (40–95 kHz) acoustic measurements, high-resolution imaging of fish morphology, and associated scattering modeling using the images. Live, adult alewife were insonified while being rotated in 1-deg increments of orientation angle in two planes of rotation (lateral and dorsal/ventral). These data provide orientation dependence of the scattering over a continuous band of frequencies, and identification of the dominant scattering features of the fish. X-ray and CT scan images of the swimbladder were digitized and incorporated into the Kirchhoff-Ray mode (KRM) model [Clay and Horne, J. Acoust. Soc. Am. **96**, 1661–1668 (1994)] and the Fourier matching method (FMM), which has recently been extended to finite-length bodies of rotation [Reeder and Stanton, J. Acoust. Soc. Am. (2004), in press]. The CT scan imagery provides rapid, noninvasive, high-resolution morphological measurements in three dimensions. The potential of the new FMM formulation is demonstrated, which is a versatile approach, valid over a wide range of shapes, all frequencies and all angles of orientation. [Work supported by ONR and NOAA.]

4:45

1pAOb7. Optimal volume backscattering thresholds for echo integration. Rebecca E. Thomas, Guy W. Fleischer (NOAA Fisheries, Northwest Fisheries Sci. Ctr., 2725 Montlake Blvd. E., Seattle, WA 98112, rebecca.thomas@noaa.gov), and Patrick H. Ressler (NOAA Fisheries, Newport, OR 97365)

Echo integration is a commonly used technique for assessment of fish stocks. In echo integration, a frequently used method to increase the signal-to-noise ratio is to ignore data below a set volume backscattering threshold. Although this form of thresholding is common, objective and replicable techniques for choosing this threshold are rarely used. Two opposing goals come into play when choosing an optimal threshold for estimating fish biomass using echo integration: maximizing the energy from backscatterers of interest, while simultaneously minimizing the energy from backscatterers not of interest. Both empirical and modeling techniques for choosing optimal thresholds were demonstrated using data from the 2003 integrated acoustic and trawl survey of Pacific hake. Empirical techniques were based on the receiver operating characteristic (ROC) curve. An ROC curve is a graphical plot of the number of true positives versus false positives for a binary classification system as the discrimination threshold is varied. Modeling techniques were based on volume backscattering data generated from modeled TS and density of hake and nonhake scatterers. For the case study using the 2003 Pacific hake survey, the historical threshold used for the southern portion of the survey was shown to be nonoptimal for that year.

5:00

1pAOb8. Identifying stable frequencies for performing animal abundance estimation with a multifrequency system. Paul L. Roberts and Jules S. Jaffe (Marine Physical Lab, Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

The practice of using acoustic backscatter at multiple frequencies for estimating animal abundance has become commonplace. However, the choice of frequencies to use, in order to optimize the data analysis and resulting estimates, is often not considered. The usual procedure is to use an inversion that is based on a forward model. One is generally free to choose the model used, and in some cases the frequencies used for data collection. Unfortunately, the frequencies are often chosen without regard for the stability of the inversion. The stability of a model-based multifrequency inversion depends on the condition number of the associated reflectivity matrix. This condition number in turn depends on the assumed size classes of the animals, the scattering model used, and the frequencies used to collect the data. Results indicate that the choice of frequencies has a significant effect on the condition number with greater than 3 orders of magnitude variation over a practical frequency range for the size classes considered. The analysis highlights the use of the condition number in evaluating the stability of the inversions. In addition, it provides a way to interpret estimates from existing systems, and a tool for designing future systems.

5:15

1pAOb9. Acoustical monitoring of fish behavior in a tank. Stephan G. Conti, Benjamin D. Maurer (Southwest Fisheries Sci. Ctr., 8604 La Jolla Shores Dr., La Jolla, CA 92037), Philippe Roux (Marine Physical Lab., La Jolla, CA 92093-0205), Christian Fauvel (Station Exprimentale d'Aquaculture IFREMER, 34250 Palavas les Flots, France), and David A. Demer (Southwest Fisheries Sci. Ctr., La Jolla, CA 92037)

In recent publications, it has been demonstrated that the total scattering cross section of fish moving in a tank can be estimated from ensembles of reverberation time series. However, the reproducibility of these measurements is influenced by parameters such as the motion or the behavior of the fish. In this work, we propose to observe acoustically the behavior of fish in a tank, and to measure their average speed. The total scattering cross section of live fish (sardines, sea bass and bocaccio) in a tank was measured repeatedly over multiple days. The species used in this study have different behaviors, which are reflected in the acoustical measurements. Depending on the behavior of the fish, such as the average displacement between two acoustic pings or the aggregation type, the total scattering cross section is different. Correlation between the acoustical measurements and the day and night behavior of the fish is demonstrated. Interpretation of such measurements can lead to monitoring acoustically and nonintrusively the behavior of fish in tanks.

Session 1pBB

Biomedical Ultrasound/Bioresponse to Vibration: Topical Meeting on Ultrasound Characterization of Cancellous and Cortical Bone II: Experimental Techniques and Clinical Impact and Comparison with Other Modalities

Kendall R. Waters, Chair
National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305-3328

Invited Papers

1:00

1pBB1. The use of ultrasound to probe structural and material properties of bone: State-of-the art. Pascal Laugier (Universite Pierre et Marie Curie CNRS 7623, Paris, France)

The knowledge of the elastic properties of bone can be used to investigate the effects of aging, disease, and treatments. Elastic properties are also required as input for both computational and analytic models. An alternative approach to direct bone mechanical testing is to use ultrasonic testing. Furthermore, ultrasound is appropriate for multiscale analysis. *In vitro*, ultrasound has been widely used to derive nondestructively the stiffness coefficients or the elastic moduli at the whole specimen level at low frequency (structural elasticity) or at the tissue level at higher frequency (material stiffness). Quantitative ultrasound (QUS) techniques have also been developed for *in vivo* skeletal status assessment based on the assumption that ultrasonic properties reflect skeletal factors of bone fragility. *In vivo* QUS is then used to predict fracture risk. Several different techniques are currently available or under development involving different type of waves (bulk compressional waves, surface waves, guided modes). Measured ultrasonic properties in transmission or in reflection depend on a variety of material properties and macro- or micro structural characteristics. However, the complexity of the interaction mechanisms between the incident ultrasonic field and bone is still not fully elucidated and requires a better understanding.

1:25

1pBB2. Use of guided ultrasonic waves for characterization of cancellous and cortical bone. Armen Sarvazyan and Alexej Tatarinov (Artann Labs., 1753 Linvale-Harbourton Rd., Lambertville, NJ 08530)

The possibility of accessing mechanical, structural, and geometrical parameters of both the cancellous and cortical bone components by different modes of ultrasonic guided waves has been demonstrated in several laboratories in Europe and the USA. Making measurements of acoustic wave propagation parameters in a wide frequency band, e.g., from 100 kHz to several MHz, enables assessment of bone layers at different depths from the bone surface. At lower frequencies, the acoustic wave velocity is found to be sensitive to changes of the cortical thickness in middleshaft areas. In epiphyseal zones, the measured low-frequency (100-kHz) wave, velocity reflects contributions of both the spongy and compact bone components and can be sensitive to changes of the trabecular structure. At higher frequencies in the MHz range, the main acoustic mode manifested in the received signal is related to the longitudinal wave, which characterizes mainly the elastic properties of the compact bone closer to the periosteum. Measurements of the ultrasonic pulse propagation parameters in a broad frequency band using surface transmission scanning mode enables the possibility to obtain the profile of acoustic properties of long bones. [Work supported by NIH and NASA.]

1:50

1pBB3. Scanning acoustic microscopy on bone—status and perspectives. Kay Raum (Q-BAM Group, Dept. of Orthopedics, Martin Luther Univ., 06097 Halle, Germany; Laboratoire d'Imagerie Paramtrique, UMR CNRS 7623, France)

High-frequency ultrasound has become one of the most powerful tools for the elastic characterization of hard materials since its invention in 1974. Many of the imaging and measurement techniques developed for NDE were also applied for the characterization of bone. However, the heterogeneous structure at several levels of organization implies multifold problems, e.g., tissue preparation, validity and applicability of the measurement techniques, resolution limitations, interpretation of results, etc. Furthermore, the technical development of commercially available high-resolution SAM lacks far behind other quantitative imaging modalities. The talk will describe the acoustic microscopes developed in our group, concepts for acoustic impedance mapping with frequencies up to 1 GHz, and techniques for determining longitudinal, shear, and surface wave velocities. The relations of acoustic to elastic parameters, either directly derived from the acoustic measurement or obtained from micromechanical tests, are presented. Combinations with other experimental and clinical techniques (US, SR-CT, nanoindentation, Raman) demonstrate the potential of SAM for an improved diagnosis, both in experimental and in clinical studies. *In vitro* results demonstrate the feasibility for assessing the elastic anisotropy and microstructural parameters of cortical bone, compositional variations within alternating osteon lamellae, structural and elastic changes in arthrotic cartilage and subchondral bone, respectively.

2:15

1pBB4. Influence of the precision of spectral backscatter measurements on the estimation of scatterer size in cancellous bone. Frederic Jenson, Frederic Padilla, and Pascal Laugier (Laboratoire d'Imagerie Paramtrique, CNRS UMR 7623 Universit Paris 6, 15 rue de l'Ecole de Medecine, 75006 Paris, France)

Ultrasonic backscatter measurements can be used to characterize trabecular bone structure and to estimate trabecular thickness (Tb.Th, i.e., the size of the scatterers). Our objective was to evaluate the performance of this estimator as well as others spectral estimators, like the frequency dependence and the mid-band amplitude of the backscatter coefficient. The performance of these estimators is degraded mainly by two factors: interference noise due to random positioning of the scatterers and attenuation. We have simulated rf-lines backscattered from trabecular bone assuming a random positioning of the trabeculae (leading to a fully developed speckle), a Gaussian form factor for the scatterers and a linear-frequency dependent attenuation. It is found that the variance in the estimation of the frequency dependence of the backscatter coefficient is as high as the variance due the biological variability in Tb.Th, in agreement with the results showed by K. Wear [J. Acoust. Soc. Am. SA, **110** (2001)]. In contrast, the variances on the mid-band amplitude and on the estimated trabecular thickness are lower than the variance due the biological variability. We also show that the effect of attenuation may be compensated by using an appropriate attenuation-compensation function. These results suggest that the inverse problem can be appropriately addressed.

2:30

1pBB5. Assessment of cortical bone microstructure and material properties using high resolution scanning acoustic microscopy. Kay Raum (Q-BAM Group, Dept. of Orthopedics, Martin Luther Univ., 06097 Halle, Germany, kay.raum@medizin.uni-halle.de), Florent Chandelier, Ingrid Leguerney, Maryline Talmant, Amena Saied, Pascal Laugier (Laboratoire d'Imagerie Paramétrique, 75006 Paris, France), and Françoise Peyrin (CREATIS and European Synchrotron Radiation Facility, 38043 Grenoble, France)

Combined evaluation of bone microstructural and mechanical information remains a challenging task which is required for bone phenotyping or accurate finite element modeling. Our objective was to assess the value of quantitative scanning acoustic microscopy (SAM) for bone characterization in comparison to synchrotron radiation computed tomography (SR-CT). Ten specimens of human cortical bone (radius) were investigated using SR-CT and SAM (200 MHz) with spatial resolution of 10 and 8 μm, respectively. An image fusion and analysis software was developed to derive site-matched estimates of (1) microstructural parameters, e.g., haversian cavity density and mean diameter and porosity, and (2) tissue properties such as mineral density (MD, SR-CT) and acoustic impedance (Z, SAM) for distinct anatomical regions of interest (osteons, interstitial tissue). Local stiffness c_{33} was derived from the combination of MD and Z. An almost perfect correlation was found for all microstructural indices derived by both techniques. Impedance was correlated to the square of MD ($R^2=0.39$, $p<1e-4$). The derived stiffness c_{33} (35.9 ± 12.8) was highly correlated with Z ($R^2=0.99$, $p<1e-4$). These findings suggest that SAM fulfills the requirement for a simultaneous evaluation of cortical bone microstructure and material properties at the tissue level.

2:45

1pBB6. Correction for soft tissue in cortical bone assessment by ultrasound. E. Bossy, M. Talmant, P. Laugier (Laboratoire d'Imagerie Paramétrique, UMR 7623, 15 rue de l'école de Médecine, 75006 Paris, France), C. Roux, S. Kolta (CEMO, 75014 Paris, France), and D. Haguenauer (Hopital Ste Perrine, 75016 Paris, France)

One of the key points in ultrasound measurements on cortical bone is the correction for soft tissue. We designed a new probe based on bi-directional axial transmission which automatically compensates velocity measurements for the soft tissue effect without preliminary evaluation of soft tissue properties. The probe consists in a linear arrangement of transducers with two sources placed on both sides of a unique group of receivers. The velocity of waves propagating parallel to the bone axis is deduced from a combination of the time delays derived from waves propagating in opposite directions at successive receivers separated by a known distance. This technique efficiently corrects for the major source of error on velocity encountered in clinical measurements which is caused by the variation of soft tissue thickness along the probe. The bi-directional technique was validated on test samples for which the residual precision error on velocity measurements was reduced to 0.2%. *In vivo* measurements yielded a value of 0.5% for the interoperator reproducibility. The clinical range of variation of the velocity measured by bi-directional technique is evaluated using clinical measurements on more than 200 subjects. Bi-directional transmission is a promising technique to minimize the variability of *in vivo* velocity measurements.

3:00–3:15 Break

Invited Papers

3:15

1pBB7. Ultrasound critical angle-reflectometry: Measuring ultrasound velocities in the clinic and in the laboratory. Peter P. Antich, Matthew A. Lewis, Edmond Richer, Billy J. Smith (Adv. Radiological Sci., Univ. of Texas Southwestern Med. Ctr., Dallas, TX 75390-9058, peter.antich@utsouthwestern.edu), and Charles Y. C. Pak (Univ. of Texas Southwestern Med. Ctr., Dallas, TX 75390)

Ultrasound critical angle-reflectometry is a modality developed and tested for measuring US velocity in the laboratory or the clinic. The applicator developed for use in the clinic consists of a transmitter concentric to a receiver, immersed in water and capable of analyzing multiple reflections from soft and mineralized tissues over angles from $-45°$ to $45°$. If a critical angle θ_c is detected in the spectrum, the velocity in bone is simply $V=c/\sin\theta c$. The analysis can be repeated at different orientations; considering only p-wave velocities, as bone has transverse symmetry, there is a unique relationship between velocity and coefficients of stiffness: $\rho V^2(\phi)=C_{11}\cos^4(\phi)+C_{33}\sin^4(\phi)+2\cos^2(\phi)\sin^2(\phi)(C_{13}+2C_{44})$. It can be shown that C_{11} and C_{33} give two orthogonal moduli of elasticity. Experiments show that the modulus of elasticity so derived is equal to that measured by mechanical testing. In addition to these angles we can measure the arrival time of the reflected signal and identify the depth from which the signal originates; for cortical layers of sufficiently small thickness it is possible to identify cancellous bone critical angles. Examples are provided which show that these quantities vary in response to metabolic and physical stimuli.

3:40

1pBB8. Assessment of trabecular bone quality in human cadaver calcaneus using scanning confocal ultrasound and dual x-ray absorptiometry (DEXA) measurements. Yixian Qin (Dept. of Biomed. Eng., SUNY Stony Brook, 350 Psych-A Bldg., Stony Brook, NY 11794, yi-xian.qin@sunysb.edu), Yi Xia, Wei Lin, Clinton Rubin, and Barry Gruber (SUNY Stony Brook, Stony Brook, NY 11794)

Microgravity and aging induced bone loss is a critical skeleton complication, occurring particularly in the weight-supporting skeleton, which leads to osteoporosis and fracture. Advents in quantitative ultrasound (QUS) provide a unique method for evaluating bone strength and density. Using a newly developed scanning confocal acoustic diagnostic (SCAD) system, QUS assessment for bone quality in the real body region was evaluated. A total of 19 human cadaver calcanei, age 66 to 97 years old, were tested by both SCAD and nonscan mode. The scanning region covered an approximate 40×40 mm^2 with 0.5 mm resolution. Broadband ultrasound attenuation (BUA, dB/MHz), energy attenuation (ATT, dB), and ultrasound velocity (UV, m/s) were measured. The QUS properties were then correlated to the bone mineral density (BMD) measured by DEXA. Correlations between BMD and QUS parameters were significantly improved by using SCAD as compared to nonscan mode, yielding correlations between BMD and SCAD QUS parameters as $R=0.82$ (BUA), and $R=0.86$ (est. BMD). It is suggested that SCAD is feasible for *in vivo* bone quality mapping. It can be potentially used for monitoring instant changes of bone strength and density. [Work supported by the National Space Biomedical Research Institute (TD00207), and New York Center for Biotechnology.]

4:05

1pBB9. Ultrasonic assessment of bone: Comparison with other characterization modalities. Patrick H. Nicholson (Dept. of Health Sci., Univ. of Jyvaskyla, P.O. Box 35, 40014, Finland, patrick_nicholson@btopenworld.com)

Compared to other techniques, quantitative ultrasound of bone (QUS) has a unique potential arising from the mechanical nature of the wave phenomena involved, the ability to use wavelengths spanning the dimensions of key structural features of bone, and the absence of ionizing radiation. However, QUS suffers from fundamental problems of interpretation when compared to most other bone assessment modalities. For example, dual energy x-ray absorptiometry gives a direct estimate of a well-understood physical parameter: bone mineral density. In contrast, QUS yields acoustic properties that cannot be so easily interpreted, and whose clinical value is derived largely from empirical relationships with physical properties of interest, such as bone density, or with other factors such as risk of fracture. Hence for QUS we cannot trace a clear causal path connecting an ultrasonic measurement to specific bone properties and on, in turn, to clinical outcomes and decisions. Without developing the theoretical framework required to achieve this, QUS will not achieve its full potential and is likely to remain sidelined and mistrusted in comparison to the x-ray-based modalities which are its principal competitors.

Contributed Papers

4:30

1pBB10. *In vitro* comparative study of three devices based on ultrasonic axial transmission. M. Muller, M. Talmant, P. Laugier (Laboratoire d'Imagerie Paramétrique, UMR CNRS 7623, 15 rue de l'école de Médecine, 75006 Paris, France), P. Moilanen, V. Kilappa, J. Timonen, P. Nicholson, and S. Cheng (Univ. of Jyväskylä, Jyväskylä, Finland)

In the axial transmission technique the velocity of waves propagating in the direction of bone axis is used to characterize cortical bone (radius, tibia, etc.). Corresponding clinically used devices are based on a long wavelength approach (typical frequency around 200 kHz) or shorter wavelength (typical frequency around 1 MHz) compared to bone thickness. They differ also by the methods of wave velocity evaluation. The aim of this study is to compare the sensitivity to bone properties of three representative devices using an *in vitro* investigation of the same specimens coupled with x-ray determination of bone properties. The moderate correlation between velocities suggests an important impact of site-matching. Among higher frequency devices, the one associated to the bidirectional probe provided generally higher correlation with bone properties than conventional axial transmission. The high-frequency devices are less sensitive to cortical thickness, CSA and trabecular BMD than the low-frequency device because high-frequency waves interrogate a thinner cortical layer than low-frequency waves. Our results suggest that different axial transmission approaches reflect different bone properties. Therefore, a multi-frequency technique might be useful in probing different bone properties at the same time (e.g., cortical thickness and BMD).

4:45

1pBB11. A new system for clincial ultrasound assessment of bone. Jonathan J. Kaufman, Gangming Luo (CyberLogic, Inc., 611 Broadway, Ste. 707, New York, NY 10012), Miriam Englander, and Robert S. Siffert (Mount Sinai School of Medicine, New York, NY 10029)

A new ultrasound device for noninvasive assessment of bone known as the QRT 2000 for Quantitative Real-Time—that is entirely self-contained, portable, and handheld is described. The QRT 2000 is powered by 4 AA rechargeable batteries and permits near real-time evaluation of a novel set of ultrasound parameters and their on-line display to the user. A clinical study has just been completed with the QRT 2000 in which 60 female subjects ranging in age from 25 to 88 were ultrasonically interrogated at their heels. The same heel was measured also using DEXA (PIXI, GE Medical Systems) and the bone mineral content (BMC) was compared with one ultrasound parameter which has been found to be extremely sensitive to bone mass. The parameter, known as the net time delay (NTD), and BMC had an associated R-squared value of 0.73, about a 13% improvement over presently marketed devices. This, coupled with the lower cost and portability of the system, makes the QRT 2000 ideally suited for use by primary care physicians in this country and abroad, and including for use in the developing world. Further improvements are being pursued through array methods (to improve reproducibility and correlations with BMD) and by incorporating other parameters particularly sensitive to architectural structure. [This research was supported by SBIR Grant No. 2R44AR045150 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the NIH.]

5:00–6:00

Panel Discussion

Session 1pPA

Physical Acoustics: Sound in Ground and Related Topics

Steven R. Baker, Chair
Naval Postgraduate School, Code PH/BA, Monterey, California 93943

Conbtributed Papers

2:00

1pPA1. Maximum effective aperture size for source direction estimates in a complex terrain. Mark Moran, D. Keith Wilson, and Roy Greenfield (USACE ERDC-CRREL, 72 Lyme Rd., Hanover, NH 03755)

Source resolution and clutter suppression from ground sensor arrays scales directly with sensor array aperture. However, boundary layer complexities can disrupt acoustic and seismic wavefronts at a variety of spatial scales. Hence, usage of an array dimension that is larger than these wavefront distortions can lead to degraded resolution and poor clutter suppression. Using experimental data, we demonstrate the maximum effective aperture for determining a line of bearing to a light moving ground vehicle. Our data were collected in mountainous terrain with open fields, dense forests, and steep hills. We deployed a field of 36 seismic sensors (geophones) and 36 high quality microphones. These sensors were distributed in triangular subclusters within a larger equilateral triangle that was roughly 100 m on each side. Our microphone separations were as small as 10 cm, with common separations of 1, 5, 10, and 25 m. Our seismic subclusters had common separations of approximately 5, 15, and 30 m. We apply a high resolution 2-D maximum-likelihood (ML) wavenumber estimation method to analyze array gain characteristics at multiple aperture sizes. Our results show large angular and range dependencies that correlate strongly with terrain complexity and disruptions of line-of-sight with a moving vehicle.

2:15

1pPA2. Theoretical analysis of a directional source for four dimensional seismic monitoring. Madjid Berraki, Bertrand Dubus (IEMN, Dept. ISEN, UMR CNRS 8520, 41 Blvd. Vauban 59046 Lille, Cedex France), and Axelle Baroni (Institut Frabcais du Ptrole, 92852 Rueil Malmaison, Cedex France)

The aim of four dimensional seismic monitoring is to monitor the changes in the subsurface in time; to achieve this, measurements are continuously repeated with the same space configuration. Reliability and repeatability of the acquisition footprint are thus key requirements for four dimensional surveys. The weathered zone (WZ) is the main cause of source signal fluctuations, polluting the signal coming from reservoir variations. This zone is directly located beneath the surface and its mechanical properties vary with weather. To eliminate this perturbation, a directional source, buried under the WZ and radiating downwards, is analyzed here. Assuming that the source size to wavelength ratio is small, a model is developed for the directional source buried in a layered half space: the Green's tensor in a layer is calculated using the discrete wavenumber method [M. Bouchon, Bull. Seismol. Soc. Am. **71**(4), 959-971 (1981)] and the solution is then propagated using the reflectivity method [J. Muller, Geophysics **58**, 153–174 (1985)]. Using a typical seismic signal called Ricker, synthetic seismograms are computed for the directional source buried in a half space covered by a layer (the WZ). The obtained results validate the directional source concept. [Work supported by an I.F.P. grant.]

2:30

1pPA3. Coupling finite element and iterative methods on a solid–solid boundary—Computation of the radiation of buried directional piezoelectric transducers for seismic applications. Madjid Berraki, Bertrand Dubus (IEMN, ISEN Dept., UMR CNRS 8520, 41 Blvd. Vauban, 59046 Lille Cedex, France), and Axelle Baroni (Institut Francais du Petrole, Malmaison Cedex, France)

Based on a theoretical directional source concept for four dimensional seismic monitoring (described in another communication), this work is dedicated to the design of the directional piezoelectric transducer. To describe the interaction of a piezoelectric source with its environment, a model coupling the finite element method with the impedance matrix of the soil is developed. The piezoelectric source and a surrounding spherical soil volume are described using finite element; on the surface of the spherical volume, a condition taking into account the external medium is applied by using an iterative method [Thiruvenkatachar *et al.*, Proc. R. Soc. London, Ser A **309**, 331–344 (1969)]. This numerical process is then implemented in the finite element code ATILA which is devoted to the modeling of piezoelectric structures. A transducer acting as the directional concept is designed and its far field radiation patterns are computed. [Work supported by an I.F.P. grant.]

2:45

1pPA4. Time-domain calculations of acoustic interactions with rigid porous surfaces. D. Keith Wilson (U.S. Army Eng. Res. Dev. Ctr., 72 Lyme Rd., Hanover, NH 03755), David F. Aldridge, Neill P. Symons (Sandia Natl. Labs., Albuquerque, NM 87185), Vladimir E. Ostashev (NOAA/Environ. Tech. Lab., Boulder, CO 80305), Sandra L. Collier, and David H. Marlin (U.S. Army Res. Lab., Adelphi, MD 20783)

In previous work, a set of time-domain equations for sound propagation in a rigid porous medium, including viscous and thermal dissipation effects, was derived [Ostashev *et al.*, J. Acoust. Soc. Am. **115**, 2624 (2004)]. From those equations, time-domain counterparts to frequency-domain impedance boundary conditions (BCs) were also derived. In this paper, several approaches to computational implementation of time-domain acoustic interactions with rigid porous surfaces are discussed that are based on these equations. Most of the approaches involve convolutions between relaxation or transformed impedance functions and the acoustic wavefield variables. For these approaches, when the relaxation times for viscous and thermal diffusion in the pores (which are inversely proportional to the static flow resistivity) are large, the convolution integrals must be evaluated over many periods of the acoustic disturbance and therefore become very computationally demanding. Some example finite-difference time-domain (FDTD) calculations show the significantly increased attenuation within the porous medium resulting from the convolution terms. Computational instabilities are observed when the wave is strongly attenuated over the FDTD spatial grid interval. Alternative ap-

proaches, based on Pade and series approximations to the time-domain impedance BC, are also considered that provide more compact numerical solutions.

3:00

1pPA5. The role of hysteresis in propagating acoustic wave in porous materials. Zhiqu Lu and James Sabatier (Natl. Ctr. for Phys. Acoust., The Univ. of Mississppi, Coliseum Dr., Univ., MS 38677)

Hysteresis and end-point memory are universal behaviors of porous materials including rock, sandstone, soil, sediment, ceramic, and some cracked materials. Recently, there have been growing interests in studying hysteresis and linking this static property to dynamic behaviors, particularly in modeling nonlinear acoustic phenomena in porous materials. This study aims at exploring the role of hysteresis in propagation of an acoustic wave in soils. A modified triaxial cell was used to measure stress/strain and sound speed/strain relationships in a triaxial test. Soil samples were subjected to a load stress path that was designed to undergo a series of meso-scopic unload-reload stress cycles with axial strain in the range of $10\mathrm{Exp}(-5)-10\mathrm{Exp}(-4)$. It is found that the slope of the hysteresis loop is linearly correlated with the dynamic elastic modulus, which implies that the fluctuation of acoustic pressure may create micro-hysteresis loops that make the dynamic elastic modulus higher than static elastic modulus. This discovery may lead to a correction in the classical equation of state by adding a hysteresis term. [Work supported by the U.S. Department of Agriculture.]

3:15

1pPA6. Impedance and Brewster angle measurement for thick porous layers. Del Leary, Craig J. Hickey (The Univ. of Mississippi, Coliseum Dr., Univ., MS 38677), Jean F. Allard, and Michel Henry (l'Université du Maine, 72085 Le Mans Cedex 9, France.)

For thin nonlocally reacting porous layers, a method derived from the work of Chien and Soroka [J. Sound Vib. **43**, 9–20 (1975)] has been used to localize the pole of the reflection coefficient located at an angle close to grazing incidence and to measure the surface impedance at this angle. A very simple experimental setup is used to obtain measurements on materials with large flow resistivity, at frequencies larger than 1 kHz, using samples with areas on the order of 1 m^2. This method is used in the present work to measure the surface impedance of acoustically thick porous layers of Ottawa sand and glass beads. There is good agreement between the measurements and predicted values. The method is also applied to study the effects of surface sealing. Sealing is modeled as a thin screen of higher flow resistivity at the surface. Reasonable agreement between the measured and predicted additional flow resistance is obtained.

3:30–3:45 Break

3:45

1pPA7. Modal analysis of broadband acoustic signals propagating in the top layer of the ground. Vladimir N. Fokin, James M. Sabatier, and Wheeler B. Howard (Natl. Ctr. for Phys. Acoust., 1 Coliseum Dr., University, MS 38655, mfok@olemiss.edu)

Experimental measurements of the frequency dependence of sound propagation losses in the top layers of the ground were performed in the frequency range between 80 and 420 Hz. These measurements revealed that the distribution of energy on the frequency-range plane has a regular structure. The features of this structure were modeled in the modal approach for a gradient model of the ground. It was shown that average sound speed in the ground and the approximate depth of the layers may be estimated through analysis of the experimentally measured frequency-range distribution of energy. The technique explores the waveguide properties of the top layer of the ground. According to the mode theory, two asymptotes coming from the point (0,0) should exist on the wave number-frequency plane. From the slope of these asymptotes, the minimum and maximal sound speeds in the waveguide were found to be 100 and 530 m/s. The approximate depth of the layers was estimated. Comparisons of obtained sound speeds and depth of layers with independently estimated values show satisfactory agreement. [Work supported by ONR Grant N00014-02-1-0878.]

4:00

1pPA8. Propagation of interface waves in a clay soil. Thomas G. Muir (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677), Richard Burgett (Planning Systems Inc., University, MS 38677), and Robert M. O'Neill (LD Consulting, Oxford, MS 38655)

Measurements and analyses on the propagation of interface waves, previously conducted on a sandy beach in Monterey, CA, are extended to a clay soil in Oxford, MS, in support of research on seismic sonar. Data are presented on attenuation versus lateral range and frequency, which may result from attenuation and scattering, as well as refraction and modal effects. Data are also presented on spatial coherence and group velocity, as a function of angle, along semi-circular arcs, also as a function of lateral range. The interface waves were generated in pulses, with both moving coil and moving mass sources (shakers), and were measured with three axis seismometers, at frequencies ranging from a few tens of Hz to several hundred Hz, at lateral ranges from a few meters to a few tens of meters. The sources excited numerous seismo-acoustic wave types; but the analysis focused on (1) Rayleigh waves having radial and vertical components in the vertical plane, and (2) transverse waves in the horizontal plane, thought to be Love waves. The data are interpreted with respect to environmental "ground truth" measurements. [Work supported by the U.S. Marine Corps Systems Command.]

4:15

1pPA9. Variation in acoustic to seismic signatures associated with natural depth variability of hardpans. Wheeler B. Howard, Craig J. Hickey, James M. Sabatier (Natl. Ctr. for Phy. Acoust./ Univ. of Mississippi, University, MS 38677), and David A. DiCarlo (Natl. Sedimentation Lab./ ARS)

Acoustic to seismic coupling techniques have been successfully used in the investigation of near-surface soils. Spatial variation in the soil profile and associated near-surface properties are manifested as differences in the acoustic to seismic transfer function. Two field sites were chosen which differ by the average depth of a naturally occurring hardpan. In this presentation, the measured acoustic to seismic response at multiple point locations within these two agricultural field sites is presented. Variability in attributes of the acoustic to seismic transfer function, namely the frequency, half-width, and amplitude of the resonance, are compared to variability of the depths to the hardpan.

4:30

1pPA10. Acoustic-to-seismic transfer function attribute for discrimination of false alarms and landmine detection. Margarita S. Fokina and James M. Sabatier (Natl. Ctr. for Phys. Acoust., 1 Coliseum Dr., University, MS 38655, vfok@olemiss.edu)

Acoustic detection of landmines based on the analysis of both spatial and frequency dependencies of the acoustic-to-seismic transfer function (A/S TF) utilizes the difference between mine impedance and the impedance of the surrounding ground. However, some deeply buried mines and some types of mines are hard to detect due to the natural variability of the ground. This work addresses the problem of false alarms and clutter (high values of the A/S TF in some frequency bands) that mimic the physics of a buried landmine. A time-scale, linear method (wavelet analysis) was used for improving the probability of landmines detection. Wavelet analysis of the calculated and measured signals permit one to find stable characteristics typical of the undisturbed ground, the disturbed ground, and the ground with a mine. The wavelet analysis showed that these characteristics may be used for the discrimination of landmines from false alarms. These characteristics also may be used as the additional criterion to find mines

which are hard to locate by traditional methods. The advantages of the suggested technique are illustrated using the experimental data. Possibilities for buried landmine classification are discussed. [Work supported by ONR Grant N00014-02-1-0878.]

4:45

1pPA11. Experiments on nonlinear acoustic landmine detection: Tuning curve studies of soil-mine and soil-mass oscillators. Murray S. Korman (Phys. Dept., U.S. Naval Acad., Annapolis, MD 21402), Thomas R. Witten, and Douglas J. Fenneman (U.S. Army Night Vision and Electron. Sensors Directorate, Fort Belvoir, VA 22060)

Donskoy [SPIE Proc. **3392**, 211–217 (1998); **3710**, 239–246 (1999)] has suggested a nonlinear technique that is insensitive to relatively noncompliant targets that can detect an acoustically compliant buried mine. Airborne sound at two primary frequencies eventually causes interactions with the soil and mine generating combination frequencies that can affect the vibration velocity at the surface. In current experiments, f_1 and f_2 are closely spaced near a mine resonance and a laser Doppler vibrometer profiles the surface. In profiling, certain combination frequencies have a much greater contrast ratio than the linear profiles at f_1 and f_2—but off the mine some nonlinearity exists. Near resonance, the bending (a softening) of a family of tuning curves (over the mine) exhibits a linear relationship between peak velocity and corresponding frequency, which is characteristic of nonlinear mesoscopic elasticity effects that are observed in geomaterials like rocks or granular media. Results are presented for inert plastic VS 1.6, VS 2.2 and M14 mines buried 3.6 cm in loose soil. Tuning curves for a rigid mass plate resting on a soil layer exhibit similar results, suggesting that nonresonant conditions off the mine are desirable. [Work supported by U.S. Army RDECOM, CERDEC, NVESD, Fort Belvoir, VA.]

5:00

1pPA12. Nonlinear acoustic experiments involving landmine detection: Connections with mesoscopic elasticity and slow dynamics in geomaterials. Murray S. Korman (Phys. Dept., U.S. Naval Acad., Annapolis, MD 21402), Douglas J. Fenneman (U.S. Army Night Vision and Electron. Sensors Directorate, Fort Belvoir, VA 22060), and James M. Sabatier (Univ. of Mississippi, Univ., MS 38677)

The vibration interaction between the top-plate of a buried VS 1.6 plastic, anti-tank landmine and the soil above it appears to exhibit similar characteristics to the nonlinear mesoscopic/nanoscale effects that are observed in geomaterials like rocks or granular materials. In nonlinear detection schemes, airborne sound at two primary frequencies f_1 and f_2 (chosen several Hz apart on either side of resonance) undergo acoustic-to-seismic coupling. Interactions with the compliant mine and soil generate combination frequencies that, through scattering, can effect the vibration velocity at the surface. Profiles at f_1, f_2, $f_1-(f_2-f_1)$ and $f_2+(f_2-f_1)$ exhibit a single peak while profiles at $2f_1-(f_2-f_1)$, f_1+f_2 and $2f_2+(f_2-f_1)$ are attributed to higher order mode shapes. Near resonance ($\sim$125 Hz for a mine buried 3.6 cm deep), the bending (softening) of a family of increasing amplitude tuning curves (involving the surface vibration over the landmine) exhibits a linear relationship between the peak particle velocity and corresponding frequency. Subsequent decreasing amplitude tuning curves exhibit hysteresis effects. Slow dynamics explains the amplitude difference in tuning curves for first sweeping upward and then downward through resonance, provided the soil modulus drops after periods of high strain. [Work supported by U.S. Army RDECOM, CERDEC, NVESD, Fort Belvoir, VA.]

Session 1pSC

Speech Communication and Signal Processing in Acoustics: Fifty Years of Progress in Speech Communication: Honoring the Contributions of James L. Flanagan

Sorin Dusan, Cochair
Rutgers University, CAIP Center, 96 Frelinghuysen Road, Piscataway, New Jersey 08854-8088

Lawrence Rabiner, Cochair
Rutgers University, 96 Frelinghuysen Road, Piscataway, New Jersey 08854

Chair's Introduction—1:30

Invited Papers

1:35

1pSC1. Fifty years of progress in acoustic phonetics. Kenneth N. Stevens (Res. Lab of Elec. and Dept. Elec. Eng., MIT, 50 Vassar St., Rm. 36-517, Cambridge, MA 02139-4307)

Three events that occurred 50 or 60 years ago shaped the study of acoustic phonetics, and in the following few decades these events influenced research and applications in speech disorders, speech development, speech synthesis, speech recognition, and other subareas in speech communication. These events were: (1) the source-filter theory of speech production (Chiba and Kajiyama; Fant); (2) the development of the sound spectrograph and its interpretation (Potter, Kopp, and Green; Joos); and (3) the birth of research that related distinctive features to acoustic patterns (Jakobson, Fant, and Halle). Following these events there has been systematic exploration of the articulatory, acoustic, and perceptual bases of phonological categories, and some quantification of the sources of variability in the transformation of this phonological representation of speech into its acoustic manifestations. This effort has been enhanced by studies of how children acquire language in spite of this variability and by research on speech disorders. Gaps in our knowledge of this inherent variability in speech have limited the directions of applications such as synthesis and recognition of speech, and have led to the implementation of data-driven techniques rather than theoretical principles. Some examples of advances in our knowledge, and limitations of this knowledge, are reviewed.

1:50

1pSC2. Fifty years of progress in speech waveform coding. Bishnu S. Atal (Dept. of Elec. Eng., Univ. of Washington, Seattle, WA 98195, bsatal@bishnu.net)

Over the past 50 years, sustained research in speech coding has made it possible to encode speech with high speech quality at rates as low as 4 kb/s. The technology is now used in many applications, such as digital cellular phones, personal computers, and packet telephony. The early research in speech coding was aimed at reproducing speech spectra using a small number of slowly varying parameters. The focus of research shifted later to accurate reproduction of speech waveforms at low bit rates. The introduction of linear predictive coding (LPC) led to the development of new algorithms, such as adaptive predictive coding, multipulse and code-excited LPC. Code-excited LPC has become the method of choice for low bit rate speech coding and is used in most voice transmission standards. Digital speech communication is rapidly moving away from traditional circuit-switched to packet-switched networks based on IP protocols (VoIP). The focus of speech coding research is now on providing to low cost, reliable, and secure transmission of high-quality speech on IP networks.

2:05

1pSC3. Fifty years of progress in speech analysis (LPC). Fumitada Itakura (Meijo Univ., Tempaku-ku, Nagoya 468-8502, Japan)

Although the history of statistical linear prediction is very long, it was first applied to speech analysis in 1966 at NTT to estimate the all-pole speech spectrum envelope in order to implement ASR and vocoder. All-pole spectral parameters are investigated in detail to find a better representation with respect to quantization and interpolation characteristics, partly at NTT and Bell Labs, leading to PARCOR and LSP. These parameters were applied to narrow band speech coders or LPC vocoders and speech synthesizer chips in the late 1970s and early 1980s, but the speech quality was insufficient for digital mobile telephone application. The problem was later solved by using hybrid CELP and MLP coding mainly at BTL. LPC analysis was also applied to acoustic front-end for ASR. Again it was found that LPC suffers with additive noise and linear/nonlinear distortions. Whereas computational efficiency of LPC was used to be the most prominent advantage, today we have gained thousands of times the processing power at a ten thousandth of the cost in 40 years. It is hoped a novel series of speech analysis methods, whose competence is comparable to human auditory system, should be developed at any expense of computational complexity.

2:20

1pSC4. Fifty years of progress in speech coding standards. Richard Cox (AT&T Labs, 180 Park Ave., Florham Park, NJ 07932, rvc@research.att.com)

Over the past 50 years, speech coding has taken root worldwide. Early applications were for the military and transmission for telephone networks. The military gave equal priority to intelligibility and low bit rate. The telephone network gave priority to high quality and low delay. These illustrate three of the four areas in which requirements must be set for any speech coder application: bit rate, quality, delay, and complexity. While the military could afford relatively expensive terminal equipment for secure communications, the telephone network needed low cost for massive deployment in switches and transmission equipment worldwide. Today speech coders are at the heart of the wireless phones and telephone answering systems we use every day. In addition to the technology and technical invention that has occurred, standards make it possible for all these different systems to interoperate. The primary areas of standardization are the public switched telephone network, wireless telephony, and secure telephony for government and military applications. With the advent of IP telephony there are additional standardization efforts and challenges. In this talk the progress in all areas is reviewed as well as a reflection on Jim Flanagan's impact on this field during the past half century.

2:35

1pSC5. Fifty years of progress in speech synthesis. Juergen Schroeter (AT&T Labs-Res., Rm. D163, 180 Park Ave., Florham Park, NJ 07932)

A common opinion is that progress in speech synthesis should be easier to discern than in other areas of speech communication: you just have to listen to the speech! Unfortunately, things are more complicated. It can be said, however, that early speech synthesis efforts were primarily concerned with providing intelligible speech, while, more recently, ''naturalness'' has been the focus. The field had its ''electronic'' roots in Homer Dudley's 1939 ''Voder,'' and it advanced in the 1950s and 1960s through progress in a number of labs including JSRU in England, Haskins Labs in the U.S., and Fant's Lab in Sweden. In the 1970s and 1980s significant progress came from efforts at Bell Labs (under Jim Flanagan's leadership) and at MIT (where Dennis Klatt created one of the first commercially viable systems). Finally, over the past 15 years, the methods of unit-selection synthesis were devised, primarily at ATR in Japan, and were advanced by work at AT&T Labs, Univ. of Edinburgh, and ATR. Today, TTS systems are able to ''convince some of the listeners some of the time'' that synthetic speech is as natural as live recordings. Ongoing efforts aim at replacing ''some'' with ''most'' for a wide range of real-world applications.

2:50

1pSC6. Fifty years of progress in speaker verification. Aaron E. Rosenberg (Ctr. for Adv. Information Processing, Rutgers Univ., CoRE Bldg, 96 Frelinghuysen Rd., Piscataway, NJ 08854-8088, aer@caip.rutgers.edu)

The modern era in speaker recognition started about 50 years ago at Bell Laboratories with the controversial invention of the voiceprint technique for speaker identification based on expert analysis of speech spectrograms. Early speaker recognition research concentrated on finding acoustic-phonetic features effective in discriminating speakers. The first truly automatic text dependent speaker verification systems were based on time contours or templates of speaker specific acoustic features. An important element of these systems was the ability to time warp sample templates with model templates in order to provide useful comparisons. Most modern text dependent speaker verification systems are based on statistical representations of acoustic features analyzed as a function of time over specified utterances, most particularly the hidden markov model (HMM) representation. Modern text independent systems are based on vector quantization representations and, more recently, on Gaussian mixture model (GMM) representations. An important ingredient of statistically based systems is likelihood ratio decision techniques making use of speaker background models. Some recent research has shown how to extract higher level features based on speaking behavior and combine it with lower level, acoustic features for improved performance. The talk will present these topics in historical order showing the evolution of techniques.

3:05–3:25 Break

3:25

1pSC7. Fifty years of progress in speech and speaker recognition. Sadaoki Furui (Dept. of Comput. Sci., Tokyo Inst. of Technol., 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8552 Japan)

Speech and speaker recognition technology has made very significant progress in the past 50 years. The progress can be summarized by the following changes: (1) from template matching to corpus-base statistical modeling, e.g., HMM and n-grams, (2) from filter bank/spectral resonance to Cepstral features (Cepstrum + DCepstrum + DDCepstrum), (3) from heuristic time-normalization to DTW/DP matching, (4) from gdistanceh-based to likelihood-based methods, (5) from maximum likelihood to discriminative approach, e.g., MCE/GPD and MMI, (6) from isolated word to continuous speech recognition, (7) from small vocabulary to large vocabulary recognition, (8) from context-independent units to context-dependent units for recognition, (9) from clean speech to noisy/telephone speech recognition, (10) from single speaker to speaker-independent/adaptive recognition, (11) from monologue to dialogue/conversation recognition, (12) from read speech to spontaneous speech recognition, (13) from recognition to understanding, (14) from single-modality (audio signal only) to multi-modal (audio/visual) speech recognition, (15) from hardware recognizer to software recognizer, and (16) from no commercial application to many practical commercial applications. Most of these advances have taken place in both the fields of speech recognition and speaker recognition. The majority of technological changes have been directed toward the purpose of increasing robustness of recognition, including many other additional important techniques not noted above.

3:40

1pSC8. Fifty years of progress in speech recognition. Raj Reddy (Wean Hall 5325, School of Computer Sci., Carnegie Mellon Univ., Pittsburgh, PA 15213-3891, rr@cmu.edu)

Human level speech recognition has proved to be an elusive goal because of the many sources of variability that affect speech: from stationary and dynamic noise, microphone variability, and speaker variability to variability at phonetic, prosodic, and grammatical levels. Over the past 50 years, Jim Flanagan has been a continuous source of encouragement and inspiration to the speech recognition community. While early isolated word systems primarily used acoustic knowledge, systems in the 1970s found mechanisms to represent and utilize syntactic (e.g., information retrieval) and semantic knowledge (e.g., Chess) in speech recognition systems. As vocabularies became larger, leading to greater ambiguity and perplexity, we had to explore the use task specific and context specific knowledge to reduce the branching factors. As the need arose for systems that can be used by open populations using telephone quality speech, we developed learning techniques that use very large data sets and noise adaptation methods. We still have a long way to go before we can satisfactorily handle unrehearsed spontaneous speech, speech from non-native speakers, and dynamic learning of new words, phrases, and grammatical forms.

3:55

1pSC9. Progress in speech research for telecommunications in the last five decades. Biing Hwang Juang (School of Elec. and Comput. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, juang@ece.gatech.edu)

Speech is the most natural form of human communications. A significant portion of speech research in the last century was devoted to the development of knowledge and technologies that aim at extending such a fundamental human capability over a long distance and at automating the support for such an extension. For the former, devices such as new acoustic transducers and techniques such as speech coding and echo cancellation have led to superb speech communication quality even for people located continents apart. For the latter, techniques in speech synthesis, recognition and understanding have been incorporated in various communication systems that automate billions of telephone calls or remote transactions of various sorts with remarkable success. Jim Flanagan, who had led the team on speech and acoustics in Bell Labs for over three decades, was very much in the center of this tremendous progress. In this talk, we highlight some of the milestones, many directly inspired by Jim Flanagan, that were achieved in the last few decades in the area of speech technologies and systems that have brought us benefits in our everyday life.

4:10

1pSC10. Fifty years of progress in speech understanding systems. Victor Zue (Comput. Sci. and Artificial Intelligence Lab., MIT, Cambridge, MA 02139)

Researchers working on human-machine interfaces realized nearly 50 years ago that automatic speech recognition (ASR) alone is not sufficient; one needs to impart linguistic knowledge to the system such that the signal could ultimately be understood. A speech understanding system combines speech recognition (i.e., the speech to symbols conversion) with natural language processing (i.e., the symbol to meaning transformation) to achieve understanding. Speech understanding research dates back to the DARPA Speech Understanding Project in the early 1970s. However, large-scale efforts only began in earnest in the late 1980s, with government research programs in the U.S. and Europe providing the impetus. This has resulted in many innovations including novel approaches to natural language understanding (NLU) for speech input, and integration techniques for ASR and NLU. In the past decade, speech understanding systems have become major building blocks of conversational interfaces that enable users to access and manage information using spoken dialogue, incorporating language generation, discourse modeling, dialogue management, and speech synthesis. Today, we are at the threshold of developing multimodal interfaces, augmenting sound with sight and touch. This talk will highlight past work and speculate on the future. [Work supported by an industrial consortium of the MIT Oxygen Project.]

4:25

1pSC11. Microphones for speech and speech recognition. James E. West (Electr. and Comput. Eng., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218)

Automatic speech recognition (ASR) requires about a 15- to 20-dB signal-to-noise ratio (S/N) for high accuracy even for small vocabulary systems. This S/N is generally achievable using a telephone handset in normal office or home environments. In the early 1990s ATT and the regional telephone companies began using speaker-independent ASR to replace several operator services. The variable distortion in the carbon microphone was not transparent and resulted in reduced ASR accuracy. The linear electret condenser microphone, common in most modern telephones, improved handset performance both in sound quality and ASR accuracy. Hands-free ASR in quiet conditions is a bit more complex because of the increased distance between the microphone and the speech source. Cardioid directional microphones offer some improvement in noisy locations when the noise and desired signals are spatially separated, but this is not the general case and the resulting S/N is not adequate for seamless machine translation. Higher-order directional microphones, when properly oriented with respect to the talker and noise, have shown good improvement over omnidirectional microphones. Some ASR results measured in simulated car noise will be presented.

4:40

1pSC12. 50 years of progress in microphone arrays for speech processing. Gary W. Elko (Avaya Labs, 233 Mt. Airy Rd., Basking Ridge, NJ 07920, gwe@ieee.org)

In the early 1980s, Jim Flanagan had a dream of covering the walls of a room with microphones. He occasionally referred to this concept as acoustic wallpaper. Being a new graduate in the field of acoustics and signal processing, it was fortunate that Bell Labs was looking for someone to investigate this area of microphone arrays for telecommunication. The job interview was exciting, with all of the big names in speech signal processing and acoustics sitting in the audience, many of whom were the authors of books and articles

that were seminal contributions to the fields of acoustics and signal processing. If there ever was an opportunity of a lifetime, this was it. Fortunately, some of the work had already begun, and Sessler and West had already laid the groundwork for directional electret microphones. This talk will describe some of the very early work done at Bell Labs on microphone arrays and reflect on some of the many systems, from large 400-element arrays, to small two-microphone arrays. These microphone array systems were built under Jim Flanagan's leadership in an attempt to realize his vision of seamless hands-free speech communication between people and the communication of people with machines.

NOTE: Separate registration required.

MONDAY EVENING, 15 NOVEMBER 2004 PACIFIC SALON 3, 7:00 TO 9:00 P.M.

Session 1eID

Interdisciplinary: Tutorial Lecture on Ocean Noise and Marine Mammals

George V. Frisk, Chair
Department of Ocean Engineering, Florida Atlantic University, 101 North Beach Road, Dania Beach, Florida 33004-3023

Chair's Introduction—7:00

Invited Paper

7:05

1eID1. Ocean noise and marine mammals: A tutorial lecture. Gerald D'Spain (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., La Jolla, CA, 92093-0704, Mail Code: 0704, gld@mpl.ucsd.edu) and Douglas Wartzok (Florida Intl. Univ., Miami, FL 33199)

The effect of man-made sound on marine mammals has been surrounded by controversy over the past decade. Much of this controversy stems from our lack of knowledge of the effects of noise on marine life. Ocean sound is produced during activities of great benefit to humans: commerce, exploration for energy reserves, national defense, and the study of the ocean environment itself. However, some recent strandings of marine mammals have been associated with the occurrence of human-generated sound. The documented increase of man-made sound in the ocean suggests the potential for more extensive though subtler effects than those observed in the mass strandings. The purpose of this tutorial is to present the scientific issues pertaining to ocean noise and marine mammals. Basic physics of sound in the ocean and long term trends of ocean sound will be presented. The biology of marine mammals, particularly their production, reception and use of sound in monitoring their environment, social interactions, and echolocation, will be reviewed. This background information sets the stage for understanding the effects of man-made sound on marine mammals. The extensive gaps in current knowledge with respect to marine mammal distribution and behavioral and physiological responses to sound will highlight research needs.

Session 2aAA

Architectural Acoustics: General Topics in Architectural Acoustics

Neil A. Shaw, Chair
Menlo Scientific Acoustics, Inc., P.O. Box 1610, Topanga, California 90290-1610

Contributed Papers

8:45

2aAA1. The ancients have stolen our inventions—Recently discovered documents of Vern O. Knudsen. Neil A. Shaw (Menlo Sci. Acoust., Inc., P.O. Box 1610, Topanga, CA 90290-1610, menlo@ieee.org) and Charlotte B. Brown (UCLA Library, Los Angeles, CA 90095-1575, cbbrown@library.ucla.edu)

Vern O. Knudsen was a founder of the Acoustical Society of America (1928), a member of the Bell Telephone Laboratories staff (1918), a respected professor of physics at UCLA (1922–1958), a Chancellor of UCLA (1959–1960), and a consultant in acoustics. Recently, a trove of Dr. Knudsen's papers and books was gathered by his family and donated to the UCLA Dept. of Special Collections where they will be added to the existing 44 boxes of the Vern Oliver Knudsen Papers (UCLA Manuscript Collection No. 1153). These newly acquired documents and architectural drawings illustrate Dr. Knudsen's pioneering work in classroom acoustics, room acoustic design, and sound propagation–aspects of acoustics that are still investigated today, 80 years later. The presentation will include a review of selected items from the Knudsen Papers including: his work for clients such as the Hollywood Bowl (1926–1929) and MGM Studios; his work with architects in the design of classrooms, performance and production spaces; Dr. Knudsen's basic physical research; and his correspondence with luminaries from numerous fields. Many in the audience will rediscover that they do, indeed, follow in the footsteps of a giant.

9:00

2aAA2. Rock art acoustics. Lauren M. Ronsse (Univ. of Kansas, 1312 Louisiana St., Lawrence, KS 66044, lronsse@ku.edu)

The relationship between the location of rock art and the acoustical properties of its immediate environment has been a source of previous investigation. The markings the early peoples created on rock formations open a fascinating portal into the exploration of their lifestyles. Previous research has shown that often ancient rock art was placed on surfaces or in locations that echoed, whereas locations without such echoes were undecorated [S. J. Waller, 2002 Rock Art Acoustics in the Past, Present, and Future. 1999 International Rock Art Congress Proceedings **2**, 11–20]. Could the acoustical characteristics of the decorated outcroppings have been perceived by the early peoples of the American Plains? A detailed study of six rock art sites located in Ellsworth County, KS has been conducted to determine the acoustical properties of these sites. At each site, various impulsive sounds were created to energize the space. The impulse responses were recorded and analyzed using two acoustic recording and analysis computer programs. This study did find echoes occurring at the decorated sites. These echoes were quantifiably louder than any reflection of sound measured at the surrounding undecorated locations.

9:15

2aAA3. Acoustics at the shrine of St. Werburgh. David Lubman (Acoust. Consultant, 14301 Middletown Ln., Westminster, CA 92683-4514)

England's Chester Cathedral (Anglican) contains a shrine to its patron saint, St. Werburgh, a 7th century Saxon princess of Mercia who became a nun and abbess. The 8th or 9th century shrine is far older than the 16th century cathedral. The shrine was enlarged around 1340, apparently because of its popularity as a place of pilgrimage and reported miracles. It was smashed by 16th century Henrician reformers and restored in the 19th century. The present shrine contains six recesses. In pre-Reformation times, kneeling pilgrims placed their heads in recesses to deliver spoken petitions to the saint. Analyses of binaural recordings made in a recess reveal that vocalizations are dramatically reinforced (10.4 dBA, 15.5 dB overall) and distorted by resonances (the strongest is 25.1 dB @ 125 Hz). Documented acoustical features suggest the shrine comprises a forgotten or previously unknown form of religious theater-for-one in which the sound experience was important. Modern listeners can enjoy a pre-Reformation auditory virtual reality experience via binaural recordings made at the shrine (best with quality headphones). [The author gratefully acknowledges architect Peter Howell of the Dog Rose Trust, a registered English charity, who reported the unusual shrine acoustics; and the cooperation of Chester Cathedral authorities.]

9:30

2aAA4. The penguin and puffin coast: Application of a micro-perforated membrane stretch ceiling. Jeff B. Pride and Jason T. Weissenburger (Eng. Dynam. Intl., 8420 Delmar Blvd., Ste. 303, St. Louis, MO 63124, edi@edi-stl.com)

The St. Louis Zoo has been undergoing major renovations that make it one of the premier zoos in the world. The Penguin and Puffin Coast, a walk through exhibit, is as near to a perfect habitat reproduction as is technologically feasible. Reproduced are a rugged coastline with craggy rock outcroppings, frigid water and a barrel vault ceiling that has theatrical lighting that can be used to simulate a colorful sunrise, a sunset over the horizon or the reversed seasons. Sounds of crashing waves and a sea lion's bark can be heard in the distance. A micro-perforated membrane stretch ceiling was chosen for its ability to easily conform to the undulations required for the sky, its durability in the moist confines of the exhibit and, last but not least, its ability to provide much needed acoustical absorption, all with the same product. This talk will discuss experiences with the micro-perforated material, review the acoustical efficacy of the installation and show photographs of the construction from essential beginning to final product. A photo or two of the penguins and puffins will also be shown.

9:45

2aAA5. Influence of surface scattering characteristics on the sound quality of reverberation. Jacob Mueller, Mendel Kleiner, Ning Xiang, and Rendell Torres (Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180)

The influence of the scattered sound from different sound diffusing surfaces on the subjective characteristics of the late reverberation of a room was studied using a simulation approach. The impulse response of various scale model scatterers was measured for a set of angles and used in combination with a simple room model for auralization. As shown by Kleiner one can quite easily hear the differences between the scattered sound from different scatterers [M. Kleiner *et al.*, Proc. 93rd Audio Eng. Soc. Convention, San Fransisco. Vol. **43**, ''Auralization of QRD and Other Diffusing Surfaces using Scale Modelling'' (1992)]. The results obtained here, for the late reverberation, are not as clear, indicating that the ''individual'' sound of scattering surfaces will influence primarily the sound quality of the early part of the reverberation. [Work supported by RPI.]

10:00–10:15 Break

10:15

2aAA6. Measured effects of diffusers and absorbers on the low-frequency modal structures in two reverberation chambers. David Nutter, Micah Shepherd, Timothy Leishman, and Benjamin Shafer (Acoust. Res. Group, Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602, dave_nutter@hotmail.com)

The Brigham Young University Acoustics Research Group has recently constructed two rectangular reverberation chambers for its research efforts. Steps have been taken to qualify the chambers for a variety of applications. Stationary diffusers and low-frequency absorbers were designed and installed as part of this process. This paper presents measurements taken at various stages of the installation to assess the impact of the diffusers and absorbers on the modal structure of the low-frequency fields.

10:30

2aAA7. Use of surrogate samples to study variation of diffuse field absorption coefficients of fiberglass with altitude. Richard D. Godfrey (Owens Corning, Sci. & Technol., 2790 Columbus Rd., Granville, OH 43023)

ASTM C 423 identifies air temperature and relative humidity as significant parameters, but does not address air density effects. In previous papers, normal and diffuse field analysis showed significant changes in predicted absorption coefficients with altitude. These predictions were validated experimentally for normal incidence in a vacuum chamber, and by using surrogate samples, thus showing the feasibility to study altitude effects in a single laboratory. Mechel design charts are normalized by two parameters. One is not dependent on air density. The other (R) is the ratio of flow resistance and the impedance of air. At constant thickness, the effect of lowering air density can be studied by increasing the sample flow resistivity. Samples with flow resistivity ratios of 1.25 and 1.5 were studied in a diffuse field following ASTM C 423 methodology. These values correspond to altitudes of 6000 and 10 700 ft in altitude, respectively. These results followed the predicted trends. At both altitudes, samples with R's of 2 and 4 had higher average absorption coefficients. For $R = 8$, the reverse effect was observed. It was concluded that the effects of altitude are not limited to the impedance tube method, but are also present in the reverberation room method.

10:45

2aAA8. Prediction of sound transmission loss of honeycomb sandwich panel by higher order approach. Tongan Wang, Shankar Rajaram, and Steven R. Nutt (Mater. Sci. Dept., Gill Foundation Composite Ctr., Univ. of Southern California, 3651 Watt Way, VHE602, Los Angeles, CA 90089)

People have studied the sound transmission loss (STL) of sandwich panels since the 1970s. However, most of the existing prediction methods have been based on single-layer dynamical models, neglecting the symmetric (dilatational) movements of the skins. Consequently, the symmetric coincident frequency of the sandwich panel cannot be predicted using those approaches. To account for this dilatational motion of the sandwich structures, different methods were utilized. However, most of them were based on one dimensional sandwich beams theories. The authors have also applied the consistent higher order beam approach to calculate the sound transmission loss of a unidirectional sandwich panel. Although the one dimensional approximation is good in predicting STL, the effects of some factors, such as the anisotropy and orientation of the principle axis of the panel, cannot be estimated. In the current work, the authors extended that one dimensional beam model into two dimensions, which allows us to calculate the STL of sandwich plates with composite laminate face sheets and honeycomb core. Both flexural (antisymmetric) and dilatational (symmetric) motions of the sandwich panel were considered in the study. The predictions were finally compared with our experimental data.

11:00

2aAA9. Measurement of transmission loss trends for orthotropic sandwich panels at a subscale facility. Shankar Rajaram, Tongan Wang, and Steve Nutt (Dept. of Mater. Sci., Univ. of Southern California, 3651 Watt Way, VHE-602, Los Angeles, CA 90089)

Interior noise studies in airplanes have identified floors as one of the primary noise paths. The acoustic barrier properties of floor panels are typically quantified by sound transmission loss (STL) measurements. A subscale transmission loss suite consisting of a reverberant room and an anechoic room was constructed and qualified to study the relative transmission loss trends of orthotropic sandwich panels used in airplane floors. A host of material combinations accounting for a variety of mechanical properties were tested for their acoustic performance. The transmission loss measurements were based on a sound intensity technique (ASTM E 2249-02) and a sound pressure technique (SAE J 1400-90). The results from these two techniques were compared to results from a full-scale accredited facility using the two-reverberation room method (ASTM E 90-02). The STL trends for orthotropic sandwich panels from the sub-scale facility were comparable to trends from the full-scale facility between 315 and 5000 Hz. The measurements permitted a ranking of several standard floor panels with respect to acoustic performance. [Support for this research from Mervyn C. Gill Foundation is gratefully acknowledged.]

11:15

2aAA10. Noise control computer modeling for architectural academic education. Michael Salameh (712 E. St. Andrews, Midland, MI 48642, ms@tir.com)

Commercial software is widely used in professional acoustical design and consultation. However, there are few educational computer programs that are designed mainly for academic teaching of architectural acoustics. Some of these computer programs, such as ACOUSTIC2D and ACOUSTIC3D [J. Turner and N. Barnett, University of Michigan], do not include noise control modeling. A new educational program, NOISE CONTROL 2D, is presented. The approach used in the modeling of noise control concepts as well as the design criteria related to this program are explored. The visual user interface and the ability to study the impact of various parameters on room noise are demonstrated. These parameters include architectural factors such as the architectural layout and acoustical parameters such as transmission losses of the room boundaries. Many aspects of noise control in buildings are covered in the program, including noise levels, reverberant and direct sound, NC, NCB, STC, and composite STC.

11:30

2aAA11. Acoustic3D Teaching Program. Norman E. Barnett and James A. Turner (Taubman College of Architecture and Urban Planning, 2000 Bonisteel Blvd., Ann Arbor, MI 48109)

A computer program has been under development for several years to assist the teaching of architectural acoustics to architecture students. This program is an outgrowth of a 2D program used routinely by the students since 1999. Emphasis is on where, when and how much sound distributes within a geometrical envelope. The initial geometry of a space is drawn or modeled with commercial programs that students typically would use to represent a design. That geometry is exported in the dxf file format and imported into the acoustic program. There it is combined with at least one sound source and initial sound reflection characteristics representing the room surfaces. As such, an acoustical simulation is produced that can be manipulated to reveal the implicit acoustical behavior of the representation. A description is given about how the program is organized and how it functions. Two brief examples are provided. The first is for a simple rectangular schoolroom such as might be used to develop some introductory learning about architectural acoustics. The second is for a more complicated geometry more appropriate for students who are reasonably well grounded in acoustics. They are presumably ready to undertake some acoustical investigation at a design development level.

Session 2aAB

Animal Bioacoustics: Marine Mammal Acoustics: Session in Honor of Ron Schusterman I

Whitlow W. L. Au, Chair
Hawaii Institute of Marine Biology, P.O. Box 1106, Kailua, Hawaii 96734

Chair's Introduction—7:55

Invited Papers

8:00

2aAB1. Dr. Ronald Schusterman's contributions to national acoustic policy. Roger L. Gentry (NOAA Fisheries, 1315 East West Hwy., Silver Spring, MD, roger.gentry@noaa.gov)

When the effects of underwater noise on marine mammals became a national issue in the 1990's, the federal government needed people trained in those fields to manage programs. No trained managers with that background existed, so a few scientists took on management roles. Of today's five Washington, DC managers in this issue, three received advanced degrees under Ron Schusterman. One started ONR's comprehensive research program on this topic, and two run NMFS's regulatory policy program on acoustics. Dr. Schusterman intended his students to work in basic science, specifically cognition, learning, sensory perception, and social behavior of marine mammals. Ironically, and despite his aversion to bureaucracy, this background equipped his students to become government decision makers. Such are the twists of history. He also taught his students to rigorously apply the scientific method and to assume nothing. If his science castaways in Washington, DC are worthy of his training, the national noise issue will be the better for it.

8:20

2aAB2. Audiology to ecology: Auditory scene analysis goes underwater. Robert Gisiner (ONR, 800 N. Quincy St., Arlington, VA 22217)

For decades the study of marine mammal audiology was considered an arcane branch of comparative psychophysics with little to offer the marine mammal ecologist. But in the past decade the oval window of the marine mammal ear has effectively become the ecologist's window into a marine ecosystem that truly is a world of sound. We are learning to listen to the ocean as marine mammals do, in order to better understand an environment long hidden from our view. And we are learning about the importance of sound to marine mammals, including the effects of our own noisy entry into their world. The past decade of revolutionary change in the use of sound to study marine mammals, and the associated revolution in our appreciation of marine mammal uses of sound, will be reviewed. We now see the ocean world through a marine mammal's ears, thanks to Ron Schusterman and a few dedicated colleagues who have opened our eyes, and ears, to the importance of audiology in ecology.

8:40

2aAB3. Studying social cognition in marine mammals. Peter Tyack (Biol. Dept., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, ptyack@whoi.edu)

Ron Schusterman has played an important role in broadening the perspective of marine mammal cognition studies from narrow comparisons to human language to more general cognitive concepts. He has also contributed to our understanding of learning mechanisms for individual recognition in pinnipeds, linking naturalistic observations with controlled studies in captive settings. I discuss how odontocetes learn to develop signals for individual and group recognition. Bottlenose dolphins use vocal learning to develop individually distinctive whistles in the first 1–2 years of life, but they also maintain the ability to imitate whistles throughout their lifetime. As maturing males form a coalition, their whistle repertoires converge. Species with more stable groups than dolphins use vocal learning to develop repertoires that are group distinctive. Schusterman has recently developed theoretical approaches to thinking about how animals form categories of social knowledge such as coalition or group. Depending upon the social context, animals that modify their vocalizations based upon auditory input and social relationships may use similar vocal learning mechanisms to develop quite different vocal repertoires. I will discuss the interaction between communication, social knowledge, and cognition in marine mammals from the approaches suggested by Schusterman for the study of social knowledge.

9:00

2aAB4. How acoustic signals become meaningful to listeners: An experimental approach. Colleen Reichmuth Kastak, Kristy Lindemann, and Ronald Schusterman (UCSC Long Marine Lab., 100 Shaffer Rd., Santa Cruz, CA 95060)

Most models of animal acoustic communication describe how vocal cues produced by a signaler influence the behavior of a listener. The response made by a listener depends in large part on the perceived meaning of the signal. But, how do signals become meaningful to listeners? In some cases, such as imprinting, signal meaning can be attributed to structural cues that are perceived and acted upon through an innate releasing mechanism. In other instances, signals may be arbitrarily related to objects, individuals, or species. Equivalence theory provides a model describing how some arbitrary signals may acquire meaning. Here, we describe theory and experimental evidence in the form of cross-modal matching-to-sample tasks showing how acoustic signals can become referents

for visual stimuli. The subject of these behavioral experiments is a California sea lion with extensive experience in performing associative learning tasks. The aim of the experiments is to establish multiple auditory-visual discriminations and then test for the emergence of untrained relationships between disparate visual stimuli linked by a common auditory signal. Preliminary data show successful emergent matching across visual and auditory modalities. These findings suggest that acoustic signals become meaningful to listeners when learned associations lead to the formation of equivalence classes.

9:20

2aAB5. White whale echolocation pulses in the open sea at the surface and at depth. Sam Ridgway and Don Carder (U.S. Navy Marine Mammal Program, Space and Naval Warfare Syst. Ctr., San Diego, 53560 Hull St., San Diego, CA 92152-5001)

Previously we reported on the first ever hearing tests of trained cetaceans in the open ocean demonstrating that zones of audibility for sound were just as great throughout the depths to which white whales dive, down to at least 300 m. The tests also showed that the whale's response whistles changed with increasing depth, overall amplitude decreased and frequency emphasis shifted higher with increasing depth from 5 to 300 m. Subsequently a door was installed in the test apparatus and the whales were taught to whistle in response to the presence of a small cylindrical target 2 m away. When the door opened the whales would utter a train of pulses and then whistle if the target were present. There was no statistical difference in echolocation pulse frequencies or amplitudes between depths of 5, 100, 200, and 300 m. Surprisingly, all pulses recorded at the open ocean test site had peak frequencies between 4 and 40 kHz. These differed markedly from pulses recorded with the same cable and apparatus in San Diego Bay where the whale's pulses usually exhibited two peaks, one in the 30–80 kHz range and the other often around 100–120 kHz.

9:40

2aAB6. Hearing loss and echolocation signal change in dolphins. Patrick W. Moore, James Finneran (SPAWARSYSCEN San Diego, Code 2351, 53560 Hull St., San Diego, CA 92152-5001), and Dorian S. Houser (BIOMIMETICA5750, La Mesa, CA 91942)

Recent studies and ongoing research have shown that echolocating dolphins can change the structure of their emitted echolocation signals during active echo-investigation of targets. The presumption has been that the animal adjusts various parameters (source level, peak frequency, etc.) of the emitted signal to maximize the information return in the target echo as a function of task or environmental constraints and requirements. Other work has suggested that the frequency range over which this dynamic control is exerted may change due changes in the animals hearing ability. Specifically, dolphins that develop high frequency hearing loss, for example from age, noise exposure or ototoxic drugs, shift the center frequency of the emitted echolocation click to lower frequency ranges. Observations of several Navy Marine Mammal Program animals with known high frequency hearing loss have demonstrated these frequency shifts. In this paper we will elaborate and extend ongoing analysis of emitted echolocation signals of several dolphins that show hearing loss associated changes in emitted signal structure, discuss the implications of these measures and suggest approaches that may prove useful for evaluating basic hearing capabilities from collected echolocation signals.

10:00–10:10 Break

10:10

2aAB7. Relationship between auditory evoked potential (AEP) and behavioral audiograms in odontocete cetaceans. Dorian S. Houser (BIOMIMETICA, 7951 Shantung Dr., Santee, CA 92071), James J. Finneran, Donald A. Carder, Sam H. Ridgway, and Patrick W. Moore (SPAWARSYSCEN San Diego, San Diego, CA 92152)

Auditory evoked potentials (AEPs) offer an alternative to behavioral methods of determining auditory sensitivity in marine mammals. The technique can be performed without the need for animal training, substantially expediting the process, and has the potential for application to stranded and rehabilitating marine mammals, thus providing an opportunity to determine hearing sensitivity in animals not likely to be kept in captivity. As an emerging technology in the field of marine mammalogy, the equivalence of AEP and behavioral thresholds remains to be quantitatively assessed. Human and laboratory animal AEPs are typically -5 to $+20$ dB of behaviorally determined thresholds and vary by technique and frequency tested. To be an effective tool in the field of marine mammalogy, the expected variation in AEP thresholds relative to behavioral thresholds in marine mammal species needs to be determined. We compare the behavioral and AEP audiograms of several odontocetes covering a range of normal hearing to profound hearing loss and demonstrate the offsets between results obtained with the two methods. Thresholds determined by the two methods show generally good agreement and demonstrate the utility of AEPs as an emerging technology in the study of marine mammal audiometry.

10:30

2aAB8. Acoustic basis for fish prey selection by echolocating odontocetes. Whitlow W. L. Au, Kelly J. Benoit-Bird (Hawaii Inst. of Marine Biol., Univ. of Hawaii, P.O. 1106, Kailua, HI 96734), Ronald Kastelein, and Sander van de Heul (SEAMARCO, 3843 CC Harderwijk, The Netherlands)

Acoustic backscatter data were obtained from four fish species, sea bass (*Dicentrarchus labras*), pollack, (*Pollachius pollachius*), grey mullet (*Chelon labrosus*), and Atlantic cod (*Gadus morhua*), using broadband bottlenose dolphin and narrow-band harbor porpoise signals. The fishes were anesthetized and attached to a monofilament net that was in turn attached to a rotor so echoes could be collected along the lateral axis of each fish. The echo waveforms were complex with many highlights and varied with the orientation of the fish. The highlight structure was determined by calculating the envelope of the cross-correlation function between the incident signal and the echoes. The strongest echo occurred when the incident angle was perpendicular to the long axis of the swim bladder, however, the number of highlights was the fewest at this perpendicular orientation and increased as the fish orientation moved

away from the perpendicular aspect. The echo structures were easily distinguishable between species and were generally consistent within species. The highlight structure of the echoes resulted in the spectrum being rippled, with local maxima and minima at different frequencies. However, differences in species were more obvious with the broadband dolphin signal than the narrow-band porpoise signal which had a much lower spatial resolution.

10:50

2aAB9. Hey Ron manatees do not echolocate either; the underwater hearing and acoustical behavior of West Indian manatees. Edmund Gerstein, Laura Gerstein (Leviathan Legacy Inc., 1318 SW 14th St., Boca Raton, FL 33486), Steve Forsythe (Naval Undersea Warfare Ctr., Newport, RI 02841), and Joseph Blue (Leviathan Legacy Inc., Boca Raton, FL 33486)

A comprehensive series of underwater psychoacoustic tests were conducted with captive manatees to measure their hearing abilities under varying acoustic conditions. Forced-choice paradigms with either a staircase, or method of constants, psychometric were used to define their audiogram, critical ratios, temporal integration, and directional hearing abilities. Pure tones, complex, broad and narrow-band noise were presented with a masker, at different intensities, to measure simultaneous masking effects at ambient levels recorded in manatee habitats. Masked thresholds across frequencies increased linearly with masker intensity. Critical ratios for pulsed signals were lower than nonpulsed, suggesting an inhibitory process affecting perception of nonpulsed tones. Comparisons with other mammals indicate manatees have acute filtering abilities for detecting pulsed sounds. While manatees do not exhibit a vocal repertoire to account for acute filtering, they are passive listeners, well adapted to selectively filter out continuous noise in favor of biologically significant sounds like their own 200-ms calls. Playbacks of band-limited calls suggest loudness summation across multiple critical bands may enable manatees to detect and locate their calls near or below ambient levels. This may explain why calibrated calls recorded in the wild exhibit no Lombard shifts. Low source levels and pulse rates negate their utility for active echolocation.

11:10

2aAB10. Underwater hearing thresholds in pinnipeds measured over a 6-year period. Brandon L. Southall (Long Marine Lab., Univ. of California, Santa Cruz, 100 Shaffer Rd., Santa Cruz, CA 95060 and NOAA Fisheries Acoust. Program), Ronald J. Schusterman, David Kastak, and Colleen Reichmuth Kastak (Univ. of California, Santa Cruz, CA)

While absolute hearing thresholds have been obtained for some marine mammals, few published data are available on how measurements of individual auditory sensitivity may change over relatively long periods of time. Studies that have investigated temporal changes in sensitivity have typically focused on animals in which differences in hearing are anticipated (age-related hearing loss). This study investigated the replicability of underwater hearing thresholds in prime-aged individuals of three pinniped species over a 6-year period. Aside from their age and experience with behavioral signal detection tasks, test subjects were of similar physical condition throughout this experiment. They were tested in the same enclosure at similar test frequencies (0.1–6.4 kHz) using identical methodology and criteria. Underwater hearing thresholds obtained throughout this testing period were not significantly different. These data indicate that underwater hearing sensitivity may remain relatively stable over long periods in nonsenescent marine mammals, including those regularly exposed to noise. Further, our results suggest that variability in testing equipment and experimental personnel may have little impact on behavioral hearing data, as long as similar testing methodologies and subject response bias are carefully maintained.

Contributed Papers

11:30

2aAB11. Testing the acoustic prey debilitation hypothesis: No stunning results. Kelly Benoit-Bird (College of Oceanic and Atmospheric Sci., Oregon State Univ., 104 Ocean Admin. Bldg., Corvallis, OR 97331), Whitlow Au (Hawaii Inst. of Marine Biol., Kailua, HI 96734), Ronald Kastelein, and Sander van de Huel (SeaMarco, 3843 CC Harderwijk, The Netherlands)

We examined the hypothesis that sounds produced by odontocetes can debilitate fish by testing the effects of three odontocete-like pulsed signals on three individuals of each of three fish species: sea bass, cod, and herring. We used a high-frequency click with a center frequency of 120 kHz exposing the fish to approximately 112 dB, a mid-frequency click with a center frequency of 70 kHz and 208 dB exposure level, and a low-frequency click with a center frequency of 40 kHz and 193 dB exposure level. Individual fish were placed in a 0.3-m-diam net enclosure immediately in front of a transducer. Each fish was allowed to remain in the experimental set up for at least 3 min prior to exposure to the clicks which were presented at a rate of 100 pulses/s grading to 700 pulses/s in 1.1, 2.2, and 3.3 s. Sea bass were also exposed to a constant pulse rate of 700 pulses/s for exposures of up to 30 s. No effect was observed in any of the fish for any signal type or pulse modulation rate. Based on our results, the hypothesis that acoustic signals of odontocetes alone can disorient or stun prey cannot be supported.

11:45

2aAB12. Auditory brainstem response hearing measurements in free-ranging bottlenose dolphins (*Tursiops truncatus*). Mandy L. H. Cook (USF College of Marine Sci., 140 Seventh Ave. S., St. Petersburg, FL 33701-5016), Randall S. Wells (Mote Marine Lab., Sarasota, FL 34236), and David A. Mann (USF College of Marine Sci., St. Petersburg, FL 33701-5016)

Bottlenose dolphins (*Tursiops truncatus*) rely on sound for communication, navigation, and foraging. Both natural and anthropogenic noise in the marine environment could mask the ability of wild dolphins to detect sounds, and chronic noise exposure could cause permanent hearing loss. The hearing abilities of a wild population of bottlenose dolphins in Sarasota Bay, FL are being investigated to determine whether they suffer hearing losses in comparison to animals living in quieter environments. This study is the first to measure the hearing sensitivity of a large population of wild dolphins that are exposed to significant levels of noise. Data on hearing sensitivities at frequencies used for acoustic communication (5–20 kHz) and echolocation (20–100 kHz) are reported. Hearing sensitivity was measured in the field using the non-invasive auditory brainstem response (ABR) procedure. ABR responses were evoked by the presentation of amplitude-modulated (AM) tones (carrier frequencies of 5, 10, 20, 30, 60, and 80 kHz) through a jawphone. The tones were modulated at 600 Hz, which elicited a robust envelope following response. A rapid ABR procedure was employed so that an entire audiogram could be obtained in approximately 30 min. This study also provides baseline data for longitudinal hearing studies in known individuals.

Session 2aAO

Acoustical Oceanography and Underwater Acoustics: Acoustic Sensing of Internal Waves I

James F. Lynch, Chair
Woods Hole Oceanographic Institute, 203 Bigelow Building, Woods Hole, Massachusetts 02543

Chair's Introduction—8:25

Invited Paper

8:30

2aAO1. Acoustic effects in presence of internal solitons in shallow water. Boris Katsnelson (Voronezh Univ., 1, Universitetskaya sq., Voronezh, 394006, Russia)

The given lecture (mini-tutorial) is dedicated to description of recent efforts (last 10–15 years) in experimental observation and theoretical modeling of the sound propagation in shallow water in the presence of traveling internal solitonlike waves (ISW). It contains a review of the main experimental observation of acoustical effects (Rubinstein and Rubinstein and Brill, Zhou *et al.*, Badiey *et al.*, etc.) The main theoretical approaches can be divided in dependence on orientation of ISW wave front: (1) Modes coupling due to ISW, traveling along acoustic track. Resonance effects in sound propagation, resonance coupling by packets and separate solitons (J-X Zhou, Preisig and Duda, etc.) are considered. (2) Horizontal refraction (HR) caused by ISW crossing acoustic track. Technique of vertical modes and horizontal rays, vertical modes and PE in horizontal plane and modeling of acoustic effects in typical conditions of Barents sea are shown. Analysis of specific peculiarities due to HR (synchronicity, periodicity, and depth dependence of intensity fluctuations) is given. Experimental observations of acoustic effects in the SWARM'95 (broadband sound propagation, shot and LFM signals) are presented. New specific features of signals, passing through ISW space-frequency horizontal structure, and fluctuations of modal spectrum are demonstrated, and experimental setup to register these is discussed. [Work was supported by RFBR and CRDF.]

9:30–9:45 Break

Contributed Papers

9:45

2aAO2. Focusing effects due to solitons: 3D Gaussian beam modeling. Paul Hursky, Michael B. Porter (Ctr. for Ocean Res., SAIC, San Diego, CA 92121), and Brian J. Sperry (SAIC, McLean, VA 22102)

As is well known, the SOFAR channel in deep water reduces spherical spreading, to cylindrical spreading, allowing sound to propagate to enormous distances. Similarly, soliton packets can produce channels in shallow water, forming acoustic corridors, their walls consecutive solitons. As the solitons pass over a propagation path they can generate dramatic focusing and defocusing effects. Acoustic modeling of such phenomena is challenging in that 3D (horizontal refraction) effects are clearly important (Nx2D approaches fail here). The geotime evolution is equally important—we must model a series of frozen oceans as the soliton packet passes by. Gaussian beam tracing models are ideally suited for such 4D modeling. We have developed a MATLAB Gaussian beam-tracing model to address these problems. It includes capabilities for a variety of useful beam options, from ''geometric beams'' to the most formal beam theory based on paraxial approximations. The latter is implemented using a novel ''reduced delta-matrix formulation'' that greatly simplifies the algorithms. The new model also allows for broadband calculations, 3D bathymetry, and 3D oceanography. We will discuss a variety of applications, with particular emphasis on the effects of solitons.

10:00

2aAO3. Evolution of internal soliton groups. Lev A. Ostrovsky (Zel Technologies and Univ. of Colorado, 325 Broadway, R./ET0, Boulder, CO 80305), Konstantin A. Gorshkov, and Irina A. Soustova (Russian Acad. Sci., Nizhny Novgorod 603095, Russia)

As known, acoustic wave propagation through a group of internal solitary waves (a solibore) in a coastal zone has a number of peculiarities; among them is a possible damping of sound upon passing a periodic group of solitons when the group period resonates with the interference distance of acoustic modes. However, solibores are typically not periodic, and the distance between solitons and possibly their order in the group can vary upon the onshore propagation. In this presentation, an evolution of a multisoliton group is considered in the framework of an evolution Gardner equation that takes both quadratic and cubic nonlinearity into account. For that, a perturbation method is used which allows the description of solitons as compounds of interacting fronts-kinks, and reduces the problem to a set of ordinary differential equations. The results are applied to strong solitons observed near the Oregon coast in 1995 where the same wave group was registered at two sites separated by 20 km. Although nonperiodic, the group retains its quasiperiodicity with a characteristic scale of order 1 km, which can affect sound propagation.

10:15

2aAO4. Uncertainty due to chaotic effects of internal waves in shallow water. Robert R. Luter, Jr. and Brian La Cour (Appl. Res. Labs., Univ. of Texas., Austin, TX 78713, rluter@arlut.utexas.edu)

Acoustic ray propagation is investigated in shallow water environments that include internal waves. Internal waves cause perturbations to the sound speed profile which results in ray chaos and causes uncertainty in the expected dynamics. In shallow water, the deep-water resonance structure exhibited in surfaces of section is largely missing due to the interaction of the rays with the ocean bottom. Instead, resonances caused by reflections create bands of chaos within large regions of regular dynamics. Lyapunov exponents are used to determine the extent of chaotic regions and to distinguish between regular and chaotic acoustic ray properties. Floquet theory is used to analyze range-periodic Hamiltonians that result from inclusion of the internal waves.

10:30

2aAO5. Matched-field replica vector stability in the winter ocean on the New Jersey (USA) shelf—RAGS03. Peter C. Mignerey and Marshall H. Orr (Acoust. Div. 7120, Naval Res. Lab., Washington, DC 20375-5350, mignerey@nrl.navy.mil)

A measure of the environmental influence on matched-field processing is the temporal autocorrelation of the coherent acoustic field passing through a vertical aperture. The matched-field autocorrelation times provide estimates of the time interval over which matched-field replica vectors will remain valid. In December 2003 the Naval Research Laboratory moored three vertical arrays at ranges of 10, 20 and 30 km distant from fixed 300- and 500-Hz CW acoustic sources. The purpose of the experiment was to measure the relationship of array gain to shelf break fluid processes (RAGS03). Data was recorded continuously for more than 20 days. Range and time dependent temporal autocorrelation times derived from this data will be presented and compared to similar measurements made at the South China Sea shelf break during the ONR AsiaEx01 acoustic propagation experiment. Preliminary RAGS03 results show that matched-field autocorrelation times are shortened during strong wind events and that sound speed fluctuations caused by semi-diurnal tidal processes forced quarter-diurnal fluctuations in the matched-field processor output. [Work supported by ONR.]

10:45

2aAO6. Estimating internal wave statistics from underwater acoustic transmission scintillation measurements on the New Jersey shelf with a 3-D stochastic model. Purnima Ratilal, Tianrun Chen, and Nicholas Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

The acoustic intensity expected after transmission through random inhomogeneities in an ocean waveguide is analytically expressed in terms of modal dispersion, attenuation, and energy redistribution in a 3-D multiple forward scattering formulation [Ratilal and Makris, J. Acoust. Soc. Am. **114**, 2428 (2003)]. This approach is used to model forward scattering through a random internal wavefield. Scattering from the density and compressibility inhomogeneities caused by the internal waves is approximated with the Rayleigh–Born series. The model is used to estimate the rms internal-wave height from low-to-mid-frequency underwater acoustic transmission scintillations measured during the Main Acoustic Clutter experiment of 2003 in the New Jersey Strataform area. Estimated internal wave height standard deviations matched those obtained from independent temperature and CTD measurements and suggest that the internal wave field was not temporally stationary.

11:00

2aAO7. Transmission loss and signal coherence statistics in the northeastern South China Sea shelf edge. Ching-Sang Chiu and Christopher Miller (Dept. Oceanogr., Naval Postgrad. School, Monterey, CA 93943, chiu@nps.edu)

Observations from the Northeastern South China Sea shelf edge in May 2001 showed spectacular changes in the sound-speed profile, transmission loss, and signal coherence at low frequency. These significant acoustical fluctuations were induced by the passage of large-amplitude, nonlinear internal waves that depressed the shallow isotherms to the ocean bottom along the transmission path. In this talk, the measured statistics of transmission loss, temporal coherence, and horizontal coherence of a 400-Hz signal transmitted upslope from a moored sound source to an L-shaped hydrophone array are presented and discussed. Specifically, the discussion is focused on both the inter- and intradaily variability of these observed statistics of the sound field and their dependence on the strength and timing of the nonlinear internal waves. [The research is sponsored by ONR.]

11:15

2aAO8. Shelf-break tidally induced environmental influences on acoustic propagation. Roger Oba, Steven Finette, Colin Shen, and Thomas Evans (Naval Res. Lab., Washington DC 20375)

Continuous wave propagation in the 100–500 Hz band in littoral regions depends upon both time-dependent oceanography and bathymetry. The environmental influences interact nonlinearly in the acoustical time variation, especially since the diurnal tide surface height changes creates time-dependent total water depth. A submesoscale hydrodynamic model developed by Shen and Evans is used with tidal forcing and a simple shelf-break bathymetry to produce surface height variation and internal wave activity due to internal tide in a stratified ocean environment. A three-dimensional parabolic equation acoustic model is used to acoustically probe this environment at various bearings relative to the shelf break and the resulting internal tidal dynamics. In particular, the acoustical results are examined for three-dimensional effects such as horizontal refraction. First, the influence of bathymetry alone is shown, and then compared to the full environment due to hydrodynamic action. The relative influences will then be compared by various measures such as modal decomposition, acoustic energy summed over depth, and signal gain degradation. [This research is sponsored by the ONR.]

Session 2aBB

Biomedical Ultrasound/Bioresponse to Vibration: Cavitation and Lithotripsy

R. Glynn Holt, Chair
Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Contributed Papers

8:30

2aBB1. Nonlinear oscillations of encapsulated gas bubbles in incompressible elastic media. E. A. Zabolotskaya, Yu. A. Ilinskii, G. D. Meegan, and M. F. Hamilton (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029)

An equation derived previously to describe nonlinear bubble oscillations in incompressible elastic media [Emelianov *et al.*, J. Acoust. Soc. Am. **115**, 581 (2004)] is modified to include an elastic shell with properties that are different from those of the host elastic medium. The new model equation is derived using Lagrangian mechanics. Nonlinearity is taken into account in the motion of the shell and surrounding medium, the compression of the gas, and the strain deformation in both elastic media. Two cases are considered for small but nonlinear bubble oscillations. The first applies to the case in which the equilibrium gas pressure is equal to the pressure at infinity, such that there is no equilibrium strain deformation in the medium. In this case, the Landau strain energy expansion can be used to obtain explicit expressions for the nonlinearity coefficients. For the second case, the equilibrium gas pressure in the bubble is different from the pressure at infinity, which results in strain deformation of both the shell and surrounding medium in equilibrium. Linear and nonlinear characteristics of the bubble oscillations are estimated for both cases. Limiting forms of the results are compared with those obtained by others. [Work supported by ARL:UT IR&D.]

8:45

2aBB2. Pressure and temperature fields from high-intensity focused ultrasound: Modeling the impact of bubbles and cavitation. Tianming Wu, Ronald. A Roy, and R. Glynn Holt (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, twu@bu.edu)

The propagation of high-intensity focused ultrasound (HIFU) in tissue-mimicking phantoms is modeled via a finite difference time-domain simulation. Above a threshold pressure, cavitation activity results and the HIFU focal zone becomes a bubbly medium. We assume an effective medium and account for the impact of bubbles by computing modified effective sound-speed and attenuation coefficients, where the latter includes cavitation-related dissipation mechanisms. Stability criteria establish the bubble equilibrium sizes, and comparison with experiments provides an estimate of the bubble number density. Nonlinear bubble responses are computed numerically and the resulting cycle-averaged void fraction is used to estimate the effective sound speed using a Woods approximation. Computed absorption cross sections related to viscous dissipation and the absorption of reradiated sound yield the effective attenuation coefficient. Using the updated sound-speed and attenuation coefficients, the pressure field is recomputed in an iterative process. Heat deposition is estimated using the averaged acoustic intensity as the heat source along with the evolving attenuation coefficient. The space-time-dependent temperature field and thermal dose is then calculated. Results indicate enhanced heating rates as well as a tadpole-shaped lesion that grows towards the HIFU transducer. [Work supported by the US Army.]

9:00

2aBB3. Observations of cavitation activity and lesion growth in optically clear tissue phantoms. Charles R. Thomas, Caleb H. Farny, Ronald A. Roy, and R. Glynn Holt (Dept. Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Above a certain acoustic pressure threshold the heating rate of a tissue-mimicking phantom (as well as *in vivo* tissue) by high-intensity focused ultrasound (HIFU) is greatly enhanced. This enhanced heating regime has been shown to correlate well with an increase in cavitation activity; thus, it is believed that the enhanced heating is the result of bubbles formed at the focus of the HIFU source. In this talk we report the results of work carried out to observe the cavitation activity in the focus of a 1.1-MHz source, using optically clear acrylamide/BSA tissue phantoms. Three different methods were employed to make the measurements: simultaneous passive cavitation detection (PCD) and video imaging, simultaneous PCD and light emission measurements (using a photomultiplier tube), and video imaging with back light. Results complement previous work by other groups which showed that thermal lesion growth progresses towards the HIFU source; however in contrast to those studies, our results indicate that at some level cavitation is always present during the formation of thermal lesions in these particular tissue phantoms. [Work supported by the US Army and the Center for Subsurface Sensing and Imaging Systems via NSF ERC Award Number EEC-9986821.]

9:15

2aBB4. Measurement and correlation of acoustic cavitation with cellular bioeffects. Daniel M. Hallow (School of Chemical & Biomolecular Eng., Georgia Inst. of Technol., 311 Ferst Dr., Atlanta, GA 30332, daniel.hallow@chbe.gatech.edu), Todd E. McCutchen, Anuj D. Mahajan, Vladimir G. Zarnitsyn, and Mark R. Prausnitz (Georgia Inst. of Technol., Atlanta, GA 30332)

Noninvasive methods to measure and predict ultrasound effects on cells are needed to realize applications of ultrasound-mediated drug delivery to improve chemotherapy, gene therapy and targeted delivery. This study tested the hypothesis that (*i*) cellular bioeffects of ultrasound correlate with cavitation dose, (*ii*) broadband noise provides a measure of cavitation dose, and, thus, (*iii*) cellular bioeffects can be predicted by noninvasively measuring broadband noise. After exposing cell suspensions to ultrasound and measuring intracellular molecular uptake and loss of cell viability (bioeffects), a broad range of bioeffects were achieved by varying frequency, pressure, exposure time, cavitation nucleation site (Optison) concentration, and cell type. As a measure of cavitation activity, broadband noise measurements from acoustic spectra were collected during cell sonication and shown to be larger at elevated pressure and, after a high initial value, sharply decayed to a constant, background value at long exposure times. Combining these results, we found that broadband noise correlated well with molecular uptake and viability over the broad range of experimental conditions used (p-value <0.0001). This indicates that acoustic spectrum analysis provides a unifying parameter to correlate with bioeffects over a wide range of acoustic and experimental conditions. [Work supported by NIH, EKOS, DoEd GAANN Program.]

9:30

2aBB5. High-speed photography and acoustic emission synchronized observation of ultrasound induced acoustic cloud cavitation. Teiichiro Ikeda, Masataka Tosaki, Shin Yoshizawa, and Yoichiro Matsumoto (Mech. Eng., The Univ. of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo, 113-8656, Japan)

Though the violent collapse of the acoustic cavitation during ultrasound therapy may cause tissue traumas, it has a potential for therapeutic benefits if it is carefully controlled. The investigation of a two-frequency focused ultrasound method for the acoustic cloud cavitation control for lithotripsy is discussed. In the forcing cycle, cavitation is controlled by generation, growth, and shape stabilization during a high-frequency (1–4 MHz) phase, and a violent forced collapse that produces very high pressure during a low-frequency phase (400–550 kHz). Ultra high-speed photography for various conditions, gas concentration in the media and the acoustic field properties, are conducted. The photography (up to 200 MHz) framing rate is synchronized with the measurement of the acoustic emission from the cavitation bubbles by a concave PVDF hydrophone with a high directivity and a broadband response up to 10 MHz. The behavior of the cloud cavitation during the two-frequency cavitation control cycle is investigated with respect to the acoustic emission of the cavitation bubbles. The shape stabilization of the cloud cavitation and the high-pressure concentration on the solid surface by the forced collapse of the bubble cloud are discussed.

9:45

2aBB6. Controlling a high intensity focused ultrasound induced cavitation field via duty cycle. Caleb H. Farny, Charles R. Thomas, R. G. Holt, and Ronald A. Roy (Boston Univ., Dept. of Aerosp. and Mech. Eng., 110 Cummington St., Boston, MA 02215, cfarny@bu.edu)

Cavitation has been implicated in the lack of control over the shape of thermal lesions generated by high-intensity focused ultrasound (HIFU). A coincident effect the decline in the acoustic emissions from cavitation at the focus suggests that the HIFU energy is shielded from the focal region, possibly by prefocal bubble activity. Most clinical techniques employ continuous-wave (CW) ultrasound, which can exacerbate the problem depending on the acoustic intensities employed. This talk presents a series of experiments investigating techniques to control HIFU energy delivered to, and cavitation activity within, a tissue phantom. A passive cavitation detector (PCD) is employed as a sensor of cavitation activity. For 1.1-MHz CW ultrasound at focal pressures above 3 MPa, bubble shielding was inferred from a steady decline in the PCD signal over time. By lowering the duty cycle the PCD output remained constant over time. Finally, driving the HIFU source initially with a CW signal and then switching to a pulsed signal resulted in shielding, recovery, and a stable PCD signal, thus demonstrating our ability to control cavitation activity during HIFU exposure. [Work supported by the U.S. Army and the Center for Subsurface Sensing and Imaging Systems via NSF ERC Award No. EEC-9986821.]

10:00

2aBB7. Correlation of ultrasound-induced premature beats and cavitation *in vivo*. Claudio Rota, Carol H. Raeman, and Diane Dalecki (Dept. Biomed. Eng. and Rochester Ctr. for Biomed. Ultrasound, Univ. of Rochester, 319 Hopeman, Rochester, NY 14627, rota@bme.rochester.edu)

Exposure of the heart to pulsed ultrasound can produce arrhythmias such as premature ventricular contractions (PVCs). Recently, studies indicate that microbubble ultrasound contrast agents can further increase the sensitivity of the heart to ultrasound by lowering the pressure threshold for PVCs. It was hypothesized that the physical mechanisms for ultrasound-induced PVCs involve acoustic cavitation. To test this hypothesis, a passive cavitation detector (PCD) was used to directly measure cavitation *in vivo* and to correlate the detected output with the occurrence of a premature beat. Experiments were performed with adult anesthetized mice. Boluses of either a contrast agent (Optison) or saline were delivered via tail vein injections. Pulsed ultrasound exposures were performed at 200 kHz with pulse durations of 1 ms and peak negative pressures ranging between 0.1 and 0.25 MPa. A 5-MHz focused transducer was used as a passive listening device of acoustic signals. For the acoustic conditions above, premature beats were found in all mice injected with Optison and the effect correlated with PCD signal amplitude. Neither premature beats nor cavitation activity were observed among animals injected with saline and exposed to ultrasound. These results are consistent with cavitation as a mechanism for this bioeffect.

10:15–10:30 Break

10:30

2aBB8. A contrast source inversion method for imaging acoustic contrasts. Koen W. A. van Dongen and William M. D. Wright (Dept. of Elec. and Electron. Eng., Univ. College Cork, Cork, Ireland)

The propagation and scattering of acoustic wavefields is described by an integral equation of the second kind. In the forward problem, the kernel contains known Green's tensors and contrast functions and is applied on the unknown total wavefields. Since the contrast functions, given by changes in density and compressibility, and the incident wavefields are known, the total wavefields can be obtained by solving the integral equation iteratively via a conjugate gradient (CG) method. In the inverse problem, the complete incident wavefields and the total wavefields at a limited number of positions are known, while the contrast function is unknown. To solve the inverse problem, the contrast function can be obtained from the integral equation by using the same CG scheme. However, for each update of the contrast function the forward problem must be solved. To avoid this problem, contrast sources are introduced, defined by the product of contrast functions and corresponding total wavefields. Hence, the integral equation will be solved by reconstructing contrast sources, from which the true amplitude contrast function is obtained via a single step minimization procedure. Results obtained with standard imaging methods like back propagation will be compared with images obtained via the contrast source formulation.

10:45

2aBB9. A model for image formation in vibro-acoustography. Glauber T. Silva (Dept. de Tecnologia da Informacao, Universidade Federal de Alagoas, BR104N, km 14, Maceio, AL, Brasil, 57072-970) and Mostafa Fatemi (Mayo Clinic College of Medicine, Rochester, MN 55905)

Vibro-acoustography is a technique that images the vibro-acoustic response of an object to the harmonic ultrasound radiation force. The system point-spread function is associated to the radiation force exerted on a point-target by a dual-frequency ultrasound beam. So far the radiation force was calculated by assuming a dual-frequency beam composed by two plane waves. This reduces the radiation force to a one-dimensional quantity neglecting transverse components. Here, we model the radiation force as a three-dimensional vector. In this model, the scattering of a dual-frequency beam with any spatial distribution by a point-target is solved. The incident and scattered fields are used to calculate the radiation force on a sphere whose radius approaches to zero. The force is proportional to the gradient of the product of the incident wave amplitudes. Evaluation of the radiation force produced by a two-element co-focused transducer shows that the transverse component of the force is about -20 dB smaller than the corresponding axial component. Effects of the transverse force on image formation are still under investigation. In conclusion, the presented model describes vibro-acoustography systems with dual-frequency beams of any spatial distribution. [Work partially supported by Grant No. DCR2003.013-FAPEAL/CNPq.]

11:00

2aBB10. Microbubble contrast agents in vibro-acoutography. Prasika Manilal, Ahmed Al-Jumaily (Diagnostics and Control Res. Ctr., Auckland Univ. of Technol., Auckland, New Zealand), and Mostafa Fatemi (Mayo Clinic College of Medicine, Rochester, MN 55905)

The use of contrast agents is becoming a routine practice in diagnostic ultrasound imaging. Microbubble contrast agents are proving to be a safe and practical way of enhancing conventional ultrasound images. This paper discusses the concept of using microbubble contrast agents in the

vibro-acoustography (VA) technique. A brief background of microbubbles and their clinical applications is given as well as a summary of the different attempts at modeling microbubble contrast agents in ultrasound fields. A general Keller–Herring equation is modified to mathematically model the interaction of encapsulated microbubbles in blood with VA ultrasound. The frequency response and scattered pressure of microbubbles with initial radius of 1 and 3 m are presented. This method is accurate for pressure amplitudes up to 500 kPa. Understanding this interaction is important for the correct interpretation of clinical investigations when using VA.

11:15

2aBB11. Vibro-acoustography for targeting kidney stones during lithotripsy. Neil R. Owen, Michael R. Bailey, Adam Maxwell, Brian MacConaghy (Ctr. for Industrial and Med. Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, cimu@apl.washington.edu), Tatiana D. Khokhlova, and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

Vibro-acoustography can be used to measure material properties and detect calcifications within the body. Two transducers (diameter 10 cm, curvature 20 cm, frequency 1.1 MHz) are placed with overlapping foci in degassed water and driven at different frequencies to produce a dynamic radiation force in the range of 5–50 kHz. A LABVIEW program instructs the transducers to sweep through this frequency range at 500-Hz increments while a synthetic cylindrical kidney stone is held in the focus in one of three ways: with a rubber band, within an acrylamide gel, or within a finger cot. A low-frequency hydrophone, 10 cm from the focus and 90 deg from the direction of propagation, detects the radiated acoustic emission from the stone. The average amplitude of five signals is recorded to measure the frequency response of the stone. Unbroken stones exhibited higher amplitude response at frequencies near 10, 25, and 35 kHz. Stones are moved to simulate patient breathing and different in-focus and out-of-focus acoustic emissions indicate that vibro-acoustography may potentially target kidney stones during lithotripsy. Comminution is improved *in vitro* by gating SWs with targeting. [Work supported by NIH Grants DK43881 and DK55674, NSBRI Grant SMS00203, and CRDF.]

11:30

2aBB12. Finite difference time domain simulation of nonlinear ultrasonic pulse propagation. Keisuke Fukuhara and Nagayoshi Morita (Dept. of Elec., Electron., and Comput. Eng., Chiba Inst. of Tech., 2-17-1 Tsudanuma, Narashino, 275-0016, Japan, g0274502@cc.it-chiba.ac.jp)

The extracorporeal shock wave lithotripsy has come into wide use rapidly owing to its advantage of non-invasiveness. However, as for the shock wave propagation in the human body in relation to practical lithotripsy, reports of numerical simulation based on models conforming to a practical situation are very few. In this paper, excited pulse wave forms are measured by using a water tank model based on a practical lithotriptor and numerical simulation is made on the basis of these measured data. A new FDTD algorithm is proposed and used for this simulation, the problem being connected with extracorporeal shock wave lithotripsy. In this simulation method, conventional plane wave approximation is not used but original equations including convection terms are directly employed to derive new FDTD algorithms. This method is applied to an experimental setup and its numerical model that resemble an actual treatment situation to compare sound pressure distributions obtained numerically with those obtained experimentally. It is shown that the present method gives clearly better results than the earlier method does, in the viewpoint of numerical reappearance of strongly nonlinear wave forms.

11:45

2aBB13. Potential mechanism for the effect of shock wave rate in shock wave lithotripsy. Yuri A. Pishchalnikov, Richard J. VonDerHaar, James A. McAteer, Irina V. Pishchalnikova (Dept. of Anatomy and Cell Biol., Indiana Univ., School of Medicine, 635 Barnhill Dr., Indianapolis, IN 46202-5120, yura@anatomy.iupui.edu), Michael R. Bailey (Univ. of Washington, Seattle, WA 98105-6698), James C. Williams, Jr., and Andrew P. Evan (Indiana Univ., School of Medicine, Indianapolis, IN 46202-5120)

Artificial stones break significantly better when shock waves (SWs) are delivered at 0.5 Hz than at 2 Hz, and patients treated at slower rates have improved stone-free rates. One possible explanation may be cavitation bubbles that might persist between SWs at high rate and distort subsequent SWs sufficiently to reduce their effectiveness at stone comminution. High-speed photography gives evidence that bubble numbers are greater at higher rates. B-mode ultrasound echo in the free field typically disappears between pulses administered at 0.5 Hz but persists at 2 Hz. Fiberoptic hydrophone measurements at 2 Hz showed, in the free field of an electrohydraulic lithotripter, that SW negative tail was truncated, and proximal to a stone, that SW waveform varied and was distorted such that often the positive pressure amplitude was reduced. Changes in waveform proximal to stone declined as the stone disintegrated and fell away. Thus, data support the persistence of cavitation bubbles at high SW rate, and consequent distortion of waveform. Additionally, debris from a stone appears to accentuate this effect. These measurements may help not only define an effective SW rate and shape, but may further improve our understanding of the comminution process. [Work Supported by NIH-DK43881, DK55674, and ONRIFO-N00014-04-1-4010.]

Session 2aEA

Engineering Acoustics and Committee on Standards: MEMS Microphones: Fabrication, Calibration, and Application to High-Density Arrays

Allan J. Zuckerwar, Cochair
NASA Langley Research Center, Hampton, Virginia 23681

Qamar A. Shams, Cochair
NASA Langley Research Center, Hampton, Virginia 23681

Chair's Introduction—8:30

Invited Papers

8:35

2aEA1. A commercialized MEMS microphone for high-volume consumer electronics. Peter V. Loeppert and Sung B. Lee (Knowles Acoust., 1151 Maplewood Dr., Itasca, IL 60143, pete.loeppert@knowles.com)

More than a billion microphones will be used this year in consumer electronic items such as cell phones, PDAs, MP3 players, and cameras. Because standard electret condenser microphones (ECMs) are temperature sensitive, these microphones are either hand placed or inserted with specialized equipment into the applications. MEMS microphones are tolerant of the high temperatures used in a lead-free solder process and hence are ideal for surface mounting with standard pick and place equipment. The challenge has been to develop a stable, low-cost MEMS microphone. This paper will present the design of the Knowles $SiSonic^{TM}$ microphone, which has been successfully commercialized over the past 2 years. Cost is a key driver in this market and total silicon area is a concern; however, the $SiSonic^{TM}$ microphone utilizes separate CMOS and MEMS dies because this has yielded a lower cost than proposed integrated solutions. The MEMS die size is driven by scaling issues and performance requirements. To achieve consistent performance, the $SiSonic^{TM}$ microphone uses a patented free-plate diaphragm. The microphone packaging has been developed to facilitate batch fabrication and a high-speed automated testing system has been developed to deliver a low cost component.

9:00

2aEA2. MEMS-based acoustic arrays: Promise and challenges. Mark Sheplak, Toshikazu Nishida, and Louis Cattafesta (Interdisciplinary Microsystems Group, Univ. of Florida, 231 MAE-A Bldg., P.O. Box 116250, Gainesville, FL 32611-6250, sheplak@ufl.edu)

A review of microelectromechanical system (MEMS)-based directional acoustic array technology is presented. The prospects for reducing cost, improving speed, and increasing mobility over conventional array technologies is critically reviewed. The advantages and limitations of existing devices are discussed. Finally, unresolved technical issues are summarized for future sensor development. A specific example of a system is presented that uses 16 hybrid-packaged silicon-micromachined piezoresistive microphones mounted to a printed-circuit board and a high-speed signal processing system to generate the array response over 2400 scan locations in under 20 s. The hybrid microphone packages show an average sensitivity of 0.8 mV/Pa with matched magnitude (0.6 dB) and phase (1 deg) responses between devices. The measured array response matches the theoretical response over the frequency range of 3 to 8 kHz with a localization error of 0.3 in. The array has a minimum detectable signal of 43.5 dB SPL for a 1-Hz bin at 6 kHz and a maximum pressure input of at least 160 dB SPL. These results represent a proof-of-concept demonstration of a high-speed, low-cost directional acoustic array system.

9:25

2aEA3. Design and fabrication of 128-channel MEMS-based acoustic array. Qamar A. Shams, William M. Humphreys, Bradley S. Sealey, Jimmy K. Adams, Toby Comeaux (NASA Langley Res. Ctr., M.S. 238, Hampton, VA 23681), John C. Ingham (Old Dominion Univ., Norfolk VA 23529), and Walter C. Babel (SAIC, Hampton, VA 23681)

Surface-mount microphones based on MEMS (micro-electromechanical system) technologies have recently become viable as component-level engineering solutions for acoustic measurements. In addition, advances in microelectronics, flexible circuitry, and array processing have motivated the design of a high-speed, low-cost acoustic array system for aeroacoustic measurements. A variety of microphones are available in the market today. Each type of microphone has its benefits and drawbacks. For example, standard condenser microphones have excellent sensitivity, stability, and high frequency response, but tend to be unwieldy, expensive, and require relatively high operational voltages. Electret microphones are small and fairly inexpensive but their performance deteriorates if exposed to moderately elevated temperatures. MEMS microphones combine the best features of the electret and condenser microphones while occupying a volume of less than 20 cubic millimeters. This paper details the custom-made 128-channel MEMS-based acoustic array for wind tunnel applications as well as the electrical, mechanical, and acoustic properties of the MEMS microphones utilized here.

9:50–10:00 Break

10:00

2aEA4. Development of MEMS microphone array technology for aeroacoustic testing. Qamar A. Shams, Sharon S. Graves, Scott M. Bartram, Bradley S. Sealey, and Toby Comeaux (NASA Langley Res. Ctr., Hampton, VA 23681)

A new approach to aeroacoustic microphone array design and implementation is described and demonstrated. Using commercially available, low-cost MEMS microphones exhibiting a suitable low-frequency response, a series of 128-channel arrays were constructed on flexible Kapton circuit boards which were bonded to rigid aluminum backplates. Cover panels with precision cutouts for the microphones were bonded on top of the Kapton circuit boards to create a smooth surface providing flush-mounting for all microphones. Connections for the microphones were created by extending strips of Kapton containing power and signal busses to the rear of the backplates. All channels were powered from a common 3 V power source, and all signals were conditioned using custom-manufactured filtering and line-driving hardware. The conditioned signals were digitized and processed in near real-time using both commercially available and customized data acquisition and analysis hardware. This new type of array construction addresses two challenges which currently limit the widespread use of large channel-count arrays for aeroacoustic applications, namely by providing a lower cost-per-channel solution and by providing a simpler method for mounting microphones in wind tunnels. The MEMS arrays have been extensively tested in anechoic and hard-walled facilities, and their performance has been found comparable to that of condenser microphone arrays.

10:25

2aEA5. Pressure calibration of MEM microphones. George S. K. Wong (Inst. for Natl. Measurement Standards, Natl. Res. Council, Ottawa, ON, Canada, K1A 0R6)

Micro-electro-mechanical microphones (MEMS) are micro devices approximately $4\times6\times1.5$ mm in size, with relatively low sensitivities (typically -40 dB *re* 1V/Pa) and operate to ultrasonic frequencies. However, the calibration of these microphones has been a challenge to most Metrology Institutes. One obvious route is to perform a free-field calibration. This presentation will discuss a preliminary coupler design for pressure comparison calibration of microphones to frequencies as high as 80 kHz.

Contributed Papers

10:50

2aEA6. Micromachined microphones with integrated optical nanoscale displacement sensors. Neal A. Hall, Wook Lee, Mohammad K. Jeelani, F. Levent Degertekin (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, gte802s@mail.gatech.edu), and Murat Okandan (Sandia Natl. Labs., Albuquerque, NM 87123)

Micromachined microphones with diffraction-based optical displacement detection are presented. A compliant membrane is made part of a phase-sensitive diffraction grating, and the deflection resulting from external acoustic pressure alters the intensities of the diffracted orders which are monitored with integrated photodiodes. The scheme provides the displacement sensitivity of a Michelson interferometer and can be integrated without beam splitters or critical alignment problems into volumes on the order of 1 mm^3. Preliminary characterization with ultrasonic sensors shows a displacement resolution of 1×10^{-4} Å/$Hz^{1/2}$ near 100 kHz with 60 μW of laser power incident on the photodetector. Current research is aimed at achieving similar displacement resolution in the audio frequency range while demonstrating the potential for high-fidelity miniature microphone arrays for hearing and measurement applications. The approach is implemented and characterized using microphone membranes with integrated diffraction grating bottom electrodes fabricated on silicon using Sandia National Laboratories' dedicated processing platform. Preliminary results on nonoptimized implementations show a flat frequency response to 15 kHz with internal noise levels below 40 dBA. [The authors would like to thank NIH and DARPA for supporting this research.]

11:05

2aEA7. Application MEMS microphones in photoacoustic instrumentation for bio-chemical detection. Michael Pedersen (CNRI, 1895 Preston White Dr., Ste. 100, Reston, VA 20191)

In this paper we discuss the design and application of MEMS-based microphones in photoacoustic (PAS) instrumentation, and it is demonstrated that by tailoring the properties of the microphone, tremendous improvements can be achieved in performance of particular importance to photoacoustic detection, such as sensitivity and noise level. Microphones typically used for PAS applications have bandwidths that exceed the requirements by an order of magnitude or more. New microphone designs, with bandwidths specifically targeted towards PAS applications, are shown to have open-circuit sensitivities in excess of 400 mV/Pa and noise levels around 0 dB SPL for a microphone diaphragm as small as 1×1 mm. This gain in performance may lead to a $10\times-100\times$ reduction of the detection limit in a state-of-the-art photoacoustic cell. Since the diaphragm in the MEMS microphone is very small, the mass of inertia is also greatly reduced, which leads to a reduction in vibration sensitivity of >10 dB over miniature hearing aid microphones and >20–30 dB over measurement microphones. The reduced vibration sensitivity is critical to push photoacoustic instruments from the laboratory into the field, where high performance rugged instruments are needed for bio-chemical detection.

11:20

2aEA8. Sources of excess noise in silicon piezoresistive microphones. Robert Dieme, Mark Sheplak, and Toshikazu Nishida (Interdisciplinary Microsystems Group, Univ. of Florida, Gainesville, FL 32611-6200)

The reduction of acoustic microphone size using microelectromechanical systems (MEMS) technology enables increased spatial and temporal resolution. Whether the small size can be effectively utilized depends on the signal-to-noise ratio and minimum detectable signal (MDS) that are a function of the structural geometry, material properties, and transduction method. The optimal MDS depends on both electronic and thermomechanical noise sources. Fundamental noise sources may be divided into frequency independent thermal noise and frequency dependent excess noise dominating at low frequencies. There have been some questions regarding the dominance of electrical or mechanical sources of the excess noise in piezoresistive microphones [A. Zuckerwar *et al.*, J. Acoust. Soc. Am. **113**, p. 3179–3187 (2003)]. Noise power spectra have been measured for various piezoresistive microphones. We present results on the bias

dependence of the excess noise that indicate that the primary source of excess noise is electrical. The relative contributions of mechanical and electrical noise sources will be discussed.

11:35

2aEA9. Free-field calibration of the pressure sensitivity of air-condenser, electret, MEMS, and piezoresistive microphones at frequencies up to 80 kHz. Gregory C. Herring, Allan J. Zuckerwar (NASA Langley Res. Ctr., Hampton, VA 23681), and Brian R. Elbing (Univ. of Michigan, Ann Arbor, MI 48109)

A free-field substitution method for calibrating measurement microphones at frequencies up to 80 kHz is demonstrated with both grazing and normal-incidence geometries. The method is suitable for newer technology microphones that cannot be calibrated with the electrostatic actuator (EA), currently the industry standard for high frequency calibrations. A substitution-based method, as opposed to a simultaneous method, avoids problems associated with the nonuniformity of the sound field and uses a $\frac{1}{4}$-inch pressure sensitive microphone as a known reference. A commercially available reference sound source (centrifugal fan) is used as a broadband acoustic source. Although the broadband excitation produces smaller instantaneous signal-to-noise ratios than tonal excitation, it minimizes reflection-related interferences that often plague free-field measurements. Calibrations were performed on $\frac{1}{4}$-in. air-condenser, electret, piezoresistive, and MEMS microphones in an anechoic chamber. Five repetitions of a single microphone, over three months, give a typical reproducibility of ± 0.2 dB. Because the air-condenser microphone can be calibrated with the EA, it was possible to estimate the accuracy of this free-field method by comparing the pressure sensitivity, as derived from the free-field measurement, with that of the EA calibration. A typical comparison gives a rms difference of ± 0.4 dB, over the range 2–80 kHz.

TUESDAY MORNING, 16 NOVEMBER 2004 SAN DIEGO ROOM, 10:00 A.M. TO 12:00 NOON

Session 2aED

Education in Acoustics: Hands on Demonstrations for High-School Students

Paul A. Wheeler, Chair
Utah State University, 1595 N 1600 E, Logan, Utah 84322

Chair's Introduction—10:00

Acoustics demonstrations will be distributed around the room. All demonstrations will be available for high school students' hands-on experimentation. Participation by other conference attenders is welcome as long as their activity does not interfere with student learning.

TUESDAY MORNING, 16 NOVEMBER 2004 ROYAL PALM SALONS 3 & 4, 8:30 A.M. TO 12:00 NOON

Session 2aMU

Musical Acoustics: Pipe Organs

Thomas D. Rossing, Chair
Physics Department, Northern Illinois Univeristy, DeKalb, Illinois 60115

Invited Papers

8:30

2aMU1. On the acoustical design of the ears of flue organ pipes. Yumiko Sakamoto, Shigeru Yoshikawa (Dept. of Acoust. Design, Grad. School of Kyushu Univ., Fukuoka, Japan, yumiko@rms.kyushu-id.ac.jp), and Judit M. Angster (Fraunhofer-Inst. Bauphysik, Stuttgart, Germany)

Application of the ears to the flue organ pipe is one of the important voicing techniques. Ears are the projections on both sides of the pipe mouth. Organ builders say the ears make not only the sound lower and darker, but also the buildup of tone smoother and quicker. The aim of this research is to confirm their recognitions and make the causes clear. For that purpose, we made acoustical and flow measurements (measurement of the velocity profiles at the mouth) with model pipes. As a result, we could confirm the recognition of the organ builders. In addition, our experiments indicate a slight increase in the blowing pressure in the foot and an increase in the inharmonicity of the pipe eigenmodes. The ear reduces the maximum jet velocity but keeps the characteristic profiles. In some cases, the profiles move as a whole more inside of the pipe. Recent acoustic measurements (eigenmodes of the pipe resonator and of the mouth tone, attack transient and stationary spectrum) on real organ pipes with ears of different heights will be also reported at the meeting.

9:00

2aMU2. CFD simulation of deflection of a jet emerging from organ pipe flue. Seiji Adachi (ATR Human Info. Sci. Labs, Keihanna Sci. City, Kyoto 619-0288 Japan, sadachi@atr.jp)

The phenomenon of an air jet deflected by sound is the crucial element in the sounding principle of air-jet driven instruments. Currently, the jet deflection model proposed by Fletcher is most widely accepted. This model successfully predicts basic properties of the instrument such as overblowing behavior with a change in blowing pressure. However, a gap still exists between the actual behavior and that predicted by the model. This gap is probably due to the model's conceptual approximations, such as an inviscid fluid and a linear response of the deflection amplitude to the magnitude of the sound field. This research aims to develop a jet deflection model directly from the Navier–Stokes equations, which govern all flow-related phenomena. Therefore, computational fluid dynamical (CFD) simulation of a jet emerging from a flue slit and deflected by the external sound field was carried out. The results were analyzed to obtain the velocity profile of the jet, the phase delay, and the magnitude of the jet deflection amplitude. The overblowing behavior estimated from the analysis is discussed by comparing it with that actually observed.

9:30

2aMU3. The influence of pipe scaling parameters on the sound of flue organ pipes. Judit M. Angster, Tilo Wik, Christian Taesch, Yumiko Sakamoto, and Andras Miklos (Fraunhofer Inst. Bauphysik, Nobelstr. 12., D-70569 Stuttgart, Germany, rata@ibp.fhg.de)

When basic phenomena of the physics of flue organ pipes is studied, experiments on models are acceptable. But these models often differ considerably from real organ pipes. For this reason the fine details of pipe sounds should be investigated on real pipes. The sound quality of an organ pipe is mainly influenced by the attack transients. This onset is first dominated by the edge tone, while later the pipe resonator will play a more important role. To understand the physics of a flue organ pipe it is necessary to measure the acoustic properties of the pipe resonator to analyze the edge tone, the attack transient, and the stationary sound of the pipe. Several special pipes with the same pitch have been investigated: pipes with different diameters; a pipe of which the cut up and a pipe of which the length is adjustable. By the evaluation all physical effects contributing to the production of sound were taken into account. The results together with the results of subjective listening tests will be used for developing a scaling method for dimensioning labial organ pipes and a software for designing organ pipe dimensions of the most important ranks. [Work supported by the European Commission.]

10:00

2aMU4. Attack transient for combinations of flue pipes. A. W. Nolle (Dept. of Phys., Univ. of Texas, Austin, TX 78712, nolle@mail.utexas.edu.)

The tonal attack due to two or more flue pipes sounded simultaneously is compared to that for a single pipe. Attack duration is measured by the times required to progress from 1% to 50% and 90% of final amplitude. Duration can be reduced by the presence of the pipe(s) of higher pitch, because the number of periods in the sound buildup for pipes of a given design does not change greatly with the pitch. Experimental results are from sound records taken in close proximity to the pipes of a Casavant organ having wind chests of Pittman type. The positive division provides stops intended for use with a closed-end 8-foot rank. The great division provides ranks intended for use with an 8-foot (open) principal. A special situation is found in principal pipes near 200 Hz. Sounding one of these with another stop at unison produces an initial sound suggesting a celeste combination. This is found to be the result of a prolonged burst of second harmonic, locked to the growing fundamental, not of frequency difference. This possibility appears in an analytical mode-locking model by Fletcher [N. H. Fletcher, J. Acoust. Soc. Am. **64**, 1566–1569 (1978)].

10:30

2aMU5. Experiments on redirection of organ pipe sound by coupling. Mendel Kleiner (Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180, kleiner@rpi.edu), Matthias Scholz (Chalmers Univ. of Technol., Gothenburg, Sweden), and Munetaka Yokota (Gothenburg Univ., Gothenburg, Sweden)

It is well known among organ builders that the sound of an organ pipe may be influenced by the addition of closely tuned pipes in the vicinity of the pipe. Rather than being a result of absorption of sound it is hypothesized that the effect is due to the sound being redirected due to coupling between pipes, similar to the coupling between elements in a Yagi antenna array. Measurement results will be shown and discussed. [Work supported by RPI and Chalmers University of Technology.]

11:00

2aMU6. Acoustics of organ reed pipes. Eric Cox and Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115)

In most lingual organ pipes, the reed vibrates against a fixed shallot, and modulates the flow of air passing the shallot into the resonator. We have measured vibrations of plucked and blown reeds of lingual pipes with and without the resonators. We discuss the effects of the acoustic field on the reed vibration and especially the interaction between the reed and the tuned resonator.

11:30

2aMU7. A comparison of different expression devices in pipe organs. Jonas Braasch (CIRMMT, Faculty of Music, McGill Univ., Montreal, QC H3A IE3, Canada, jb@music.mgill.ca) and Thomas D. Rossing (Northern Illinois Univ., DeKalb, IL)

After the introduction of the orchestra crescendo at the end of the 18th century by the ''Mannheim school,'' the ability to play the organ expressively like an orchestra was one of the organ builders' greatest concerns. Soon, both the wind swell and door swell came into fashion, and later another swell system, the crescendo wheel, was introduced. While both door swell and crescendo wheel can be successfully applied to all types of organ stops, the wind swell only works well with free reeds for tuning reasons. In this investigation, all three swell systems were measured on various instruments and compared to each other. In addition, two free-reed pipes were measured at the Northern Illinois University using a laser vibrometer. The crescendo wheel was found to be most effective, and for frequencies around 2 kHz the increase in sound-pressure level could be up to 50 dB between the softest and the loudest adjustment. The maximum dynamic range for the wind and the door swells is approximately 10 dB in the same frequency range. While the dynamic range is lower for the wind swell and the door swell, their advantage is the continuous variability.

TUESDAY MORNING, 16 NOVEMBER 2004 ROYAL PALM SALON 1, 9:00 TO 11:35 A.M.

Session 2aNS

Noise: Propulsion/Airframe Aeroacoustics I

Joe W. Posey, Chair
NASA Langley Research Center, Hampton, Virginia 23681

Chair's Introduction—9:00

Invited Papers

9:05

2aNS1. NASA's propulsion airframe aeroacoustics research. Russell H. Thomas (NASA, MS 166, NASA Langley Res. Ctr., Hampton, VA 23681-2199)

The integration of propulsion and airframe is a fundamental consideration in the design of an aircraft system. Many considerations influence the integration, such as structural, aerodynamic, and maintenance factors. In the future, a focus on the aerodynamic and acoustic interaction effects of installation, propulsion airframe aeroacoustics will become more important as noise reduction targets become more difficult to achieve. In addition to continued fundamental component reduction efforts, a system level approach that includes propulsion airframe aeroacoustics will be required in order to achieve the 20-dB noise reductions envisioned by the aggressive NASA goals. This emphasis on the aeroacoustics of propulsion airframe integration is a new part of NASA's ongoing acoustics research. The presentation will review current efforts and highlight technical challenges and approaches.

9:35

2aNS2. Diagnostics and reduction of propulsion airframe aeroacoustic interactions. Vinod G. Mengle, Robert W. Stoker, and Ronen Elkoby (The Boeing Co., P.O. Box 3707, MC 67-ML, Seattle, WA 98124)

In the past, significant reduction in aircraft noise has been achieved by studying engine noise and airframe noise in isolation as two separate components. However, when an engine is installed on an aircraft the flow and acoustic interactions between them produce a total noise signature which is often different than the sum of the component noise signatures. While take-off conditions are typically dominated by engine noise, approach conditions are characterized by both engine and airframe noise. Nevertheless, propulsion airframe aeroacoustic (PAA) interactions are present under all conditions and gain in importance when engine or airframe noise is further reduced. This paper focuses on the diagnostics and reduction of the flow-acoustic interaction effects between an engine exhaust and the airframe, especially the high-lift system on wings. Flow and acoustic results are presented for an isolated scale-model nozzle in a free jet, and also when it is installed under a wing at take-off and approach conditions. The difference between these two noise signatures, in the far field and at the source, captures the PAA effect. In particular, the PAA effect due to changes in nozzles, pylons, and flaperons is presented and the overall noise reducing mechanisms are postulated.

10:05

2aNS3. Propulsion airframe aeroacoustics practices at Honeywell. Donald S. Weir (Honeywell, P.O. Box 52181, MS 503-333, Phoenix, AZ 85072-2181)

Honeywell has been developing and applying acoustic models of propulsion airframe aeroacoustic phenomena for over 20 years. The initial application of a wing-shielding model was developed for the NASA General Aviation Synthesis Program in 1982. Since that time, more sophisticated models of wing shielding and reflection have been developed with internal and NASA funding. Recent work has involved models of wing shielding for aft mounted engines and wing reflection for wing mounted engines. These methods

are described in the presentation. Comparisons with the Raynoise Code by LMS and measured aircraft fly over noise data are made to show the effectiveness of the model. The attenuation of the inlet noise by the wing of an aft mounted engine and the amplification of the noise by wing mounted engines are evaluated.

10:25–10:40 Break

10:40

2aNS4. Distributed exhaust for jet noise reduction. David B. Schein (Northrop Grumman Corp., 9S26/W6, One Hornet Way, El Segundo, CA 90245-2804, david.schein@ngc.com)

Acoustic detection of low flying, jet-powered military air vehicles and takeoff noise levels from commercial jets are often driven by jet mixing noise radiated from the exhaust. Cueing provided by the noise signature results in increased opportunity for a ground-based visual observer against such a target. Significant reductions in detectability and takeoff noise levels can be achieved through reductions in jet noise. Research has been conducted over the past several years to develop innovative, quiet, distributed exhaust nozzle (DEN) concepts, which attempt to achieve revolutionary reductions in jet mixing noise while minimizing propulsion penalties. The DEN's benefit relies on discharging the exhaust flow through many miniature nozzles rather than one or two large nozzles. Noise suppression from the DEN concept results from a favorable shift in frequency content compared to conventional jets. Significant increases in atmospheric attenuation and decreases in the ear's sensitivity at these higher frequencies result in much reduced detection ranges and perceived noise levels. This technology is applicable to both subsonic and supersonic aircraft, and particularly to low flying fixed-wing unmanned air vehicles, special operations forces transports, and future commercial transports with especially stringent noise reduction requirements. [Work supported by NASA LaRC.]

11:00

2aNS5. Lateral attenuation and airport noise modeling. Kevin P. Shepherd (NASA Langley Res. Ctr., M.S. 463, Hampton, VA 23681, k.p.shepherd@nasa.gov)

Prediction of airport noise is generally accomplished using semiempirical methods such as those contained within the F.A.A.'s Integrated Noise Model or the U.S.A.F.'s Noisemap. One component of these models addresses the prediction of noise for observers located at lateral, or sideline, positions for which the elevation angle to the source is small. Under these conditions the effect of the presence of the ground surface on sound propagation is important, as is the source directivity. Recent efforts in the United States and Europe have been aimed at separating these two phenomena with the result that acoustical directivity characteristics have been quantified for a range of commercial aircraft. Differences between aircraft with wing- and tail-mounted engines are significant. These recent studies will be reviewed and attempts made to identify the physical mechanisms responsible for the observed source directivity characteristics.

Contributed Paper

11:20

2aNS6. Further developments in aircraft flyover noise synthesis and propagation. Brenda M. Sullivan and Stephen A. Rizzi (Structural Acoust. Branch, NASA Langley Res. Ctr., Hampton, VA 23681)

Subjective assessments of the noise from aircraft flight operations require time histories of acoustic pressure at listener positions. Synthesized sound has an advantage over recordings by allowing the examination of proposed aircraft, flight procedures, and other conditions or configurations for which recordings are unavailable. A two-stage process for synthesizing flyover noise at listener positions on the ground was previously developed, enabling the creation of an immersive test environment. The first stage entails synthesizing time histories at the flying source. Rizzi and Sullivan [J. Acoust. Soc. Am. **113**, 2245 (2003); **114**, 23 (2003)] presented an approach for synthesizing sound from broadband sources (e.g., jet noise) based on predicted $\frac{1}{3}$-octave band source spectra, with the inclusion of temporal fluctuations based on empirical data. Reported here are further developments in the synthesis of tone-dominated source (e.g., fan noise). The second stage entails propagation of the synthesized sound from the flying source to the listener. Improvements in the atmospheric attenuation part of the second stage are also presented.

Session 2aPA

Physical Acoustics: Topics in Atmospheric Sound Propagation

Roger M. Waxler, Chair
National Center for Physical Acoustics, University of Mississippi, 1 Coliseum Drive, University, Mississippi 38677

Contributed Papers

7:45

2aPA1. The acoustical equivalence principle for ray tracing. Edward R. Floyd (10 Jamaica Village Rd., Coronado, CA 92118-3208)

While the acoustical wave functions for various acoustical problems can be mapped into each other by a coordinate transformation (the acoustical equivalence principle), the corresponding ray tracings cannot be so mapped. Applying the acoustical equivalence principle [A. E. Faraggi and M. Matone, Int. J. Mod. Phys. A **15**, 1869–2017 (2000) for the analogous quantum equivalence principle] to the underlying acoustical Hamilton–Jacobi equation for ray tracing renders a modified acoustical Hamilton–Jacobi equation which does obey the quantum equivalence principle. Solutions of the modified acoustical Hamilton-Jacobi equation, which are either acoustical Hamilton's principal function or the acoustical reduced action (acoustical Hamilton's characteristic function), are the generators of motion for rigorous ray tracing.

8:00

2aPA2. Application of the Green's function parabolic equation (GFPE) method to realistic outdoor sound propagation scenarios. Jennifer L. Cooper (ENSCO, Inc., 5400 Port Royal Rd., Springfield, VA 22151, cooper.jennifer@ensco.com)

The Green's function parabolic equation (GFPE) is a powerful method for outdoor sound propagation prediction. Given appropriate inputs, GFPE predictions can include the effects of terrain, ground cover, sound speed profile, and turbulence on atmospheric propagation. In many situations, however, access to such input information can be limited. The input data resolution and accuracy can have a significant impact on model output accuracy. In turn, variability of terrain and sound speed profile and statistical aspects of the turbulence impact the model parameter requirements. For example, in the presence of turbulence, lower limits on computational grid height are determined by turbulence correlation scales as well as acoustic wavelength. Input data resolution requirements and some techniques that may be employed to account for insufficient input data will be discussed. Model results for realistic situations will be presented.

8:15

2aPA3. A parabolic equation solution for advected acousto-gravity waves. Jeremy Bruch, Michael D. Collins, Dalcio K. Dacol, Joseph F. Lingevich (Naval Res. Lab., Washington, DC 20375), and William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180)

It is relatively difficult to account for advection in parabolic equation solutions. An effective solution was recently derived for the acoustic case [J. Acoust. Soc. Am. **111**, 729–734 (2002)]. This work is currently being extended to include buoyancy effects. In the initial study, emphasis is placed on wave numbers near the Lamb wave, where buoyancy and compressibility both have a significant effect. An approximate parabolic solution is derived and tested for range-indendent cases using a spectral solution. The limiting cases of small ambient flow and/or no gravity are compared to the extended result for consistency. [Work sponsored by the ONR.]

8:30

2aPA4. The theory of the generation of atmospheric microbaroms. Roger Waxler and Kenneth E. Gilbert (NCPA, 1 Coliseum Dr., P.O. Box 1848, University, MS 38677, rwax@olemiss.edu)

It is well known that the standing wave components of large ocean wave systems, such as those produced by storms, radiate infrasound (microbaroms) into both the ocean and the atmosphere. The radiated sound is found in a narrow band (from 0.2 to 0.3 Hz wide) centered near 0.2 Hz. It is also well known that the radiation mechanism is nonlinear since ocean-wave wavelengths are too short to directly radiate. A systematic theoretical study of the radiation mechanism has been undertaken. The model considered is the two-fluid system consisting of atmosphere (a very rare fluid) over ocean (a very dense fluid). The interface between the fluids is allowed to undulate. The equations of fluid mechanics are then solved to second order in Mach number. It is found that the physical mechanism for (as well as the mathematical form of) the radiation into the atmosphere is entirely different from that of the radiation into the ocean. In particular, the compressibility of the atmosphere plays a crucial role.

8:45

2aPA5. Fluctuating pressure distributions around spherical foam windscreens. Jeremy Webster, Richard Raspet, and Kevin Dillion (Natl. Ctr. for Physical Acoust, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, jwebster@olemiss.edu)

The low-frequency wind noise reduction produced by spherical foam windscreens outdoors can be estimated by measuring the steady-state flow pressure distribution around the windscreen and then area averaging to produce a pressure response at the center of the sphere [Zheng (2003)]. We have constructed a foam windscreen with a distribution of four probe microphones near the surface to investigate the validity of this hypothesis at low frequencies and to measure correlations between the pressure fluctuations at different positions on the surface and at the center. The aim of this research is to develop methods of wind noise reduction which combine the effectiveness of passive foam windscreens with active multiple microphone processing techniques. In this paper we describe the construction of the test windscreen and initial measurements made outdoors.

9:00

2aPA6. A simple normal-mode model for nighttime traffic noise. Kenneth E. Gilbert, Roger Waxler, and Carrick L. Talmadge (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, Coliseum Dr., University, MS 38677, kgilbert@olemiss.edu)

At night, traffic noise propagates to long distance via normal modes trapped near the ground in the sound duct created by surface cooling. Consequently, at distances of 500 m or more, traffic noise at any given location is the sum of contributions from many individual vehicles. To predict the band-averaged and time-averaged levels in such situations, one can use an incoherent sum of normal modes. Using an incoherent modal sum, we derive a simple analytic expression for the noise due to traffic on an infinitely long highway. For long highways (10–20 km), it is shown that the controlling factor in the noise propagation is not distance alone,

but, instead, the product of the mode attenuation coefficients and the distance from the highway. The predictions for an infinitely long highway are compared with those for a highway of finite length and error bounds are given.

9:15

2aPA7. The near ground structure of the nocturnal sound field: The existence of a quiet height in the frequency domain. Roger Waxler, Carrick L. Talmadge, Kenneth E. Gilbert, and Shantharam Dravida (NCPA, 1 Coliseum Dr., University, MS 38677, rwax@olemiss.edu)

The sound field generated by a point source in the downward refracting atmosphere typical of the nocturnal boundary layer is considered. Theoretical arguments are given that, at a fixed frequency, at ranges greater than a few hundred meters from the source the magnitude of the sound field necessarily has a deep minimum a few meters above the ground. The precise height of this minimum depends on the sound speed profile as well as the frequency, decreasing with increasing frequency. It is shown that a recently developed modal description of the nocturnal sound field yields a simple explanation for the existence of this quiet height and its emergence with increasing range. Experimental results verifying the existence and frequency dependence of the quiet height are presented.

9:30

2aPA8. Forest effects on acoustic pulse propagation. Donald G. Albert, Frank E. Perron, Jr., and Stephen N. Decato (USA ERDC Cold Regions Reasearch and Eng. Lab., 72 Lyme Rd., Hanover, NH 03755)

A series of short-range outdoor measurements was conducted to investigate forest effects on acoustic pulse propagation. The measurements investigated seven different forest stands with a variety of different tree species including deciduous, evergreen, and mixed. A 0.45-caliber blank pistol shot was used as the source of the acoustic pulses, and the signatures were recorded 30 to 60 m away using a digital seismograph. An identical measurement was conducted in an open field for comparison. The recorded waveforms generally show the elongation characteristic of pulse propagation over a highly porous ground surface, with high-frequency reverberation arrivals superimposed on the basic waveform shape. These measurements can provide parameters useful for theoretical predictions of acoustic propagation within forests, and also illustrate some of the natural variability encountered in these environments. [Work funded by U.S. Army.]

9:45

2aPA9. Contributions of a thin attenuating layer in forest sound propagation. Michelle Swearingen and Michael White (Engineer Res. and Development Ctr., Construction Eng. Res. Lab., 2902 Farber Dr., Champaign, IL 61822)

The GFPE has been modified to incorporate a forest by replacing the local wavenumber with the bulk wavenumber of Twersky's multiple scattering theory. For modeling the effect of the forest canopy, this corresponds to an attenuating layer at some height above the ground surface. Because of propagation paths above and beneath the canopy, the attenuation per unit distance is generally less than that of the canopy. In this presentation, the attenuation contribution of this layer will be examined by comparing propagation predictions with and without the attenuation layer, with refracting and non-refracting sound speed profiles, and with rigid and realistic ground surfaces. It is interesting that some propagating modes in the forest are suppressed by the attenuation, removing a source of interference and leading to an apparent decrease in attenuation. An analysis of the sensitivity of the GFPE to attenuation layers of different strengths will be performed for receivers near the ground, taking both frequency and distance into account.

10:00

2aPA10. Nonlinearity in outdoor propagation of periodic signals: Measurement results. J. Micah Downing, Michael M. James, Christopher M. Hobbs (Wyle Labs., 2001 Jefferson Davis Hwy. Ste. 701, Arlington, VA 22202), Kent L. Gee, Victor W. Sparrow (The Pennslyvania State Univ., University Park, PA 16802), and Sally A. McInerny (The Univ. of Alabama, Tuscaloosa, AL 35487)

As a continuation of a joint program among Wyle Laboratories, Penn State, and The University of Alabama to develop better noise models for military aircraft, field measurements of high-amplitude periodic and broadband waveforms were conducted at Blossom Point Range, MD. The source of the acoustic signals was the Army Research Laboratory's Mobile Acoustic Source (MOAS) pneumatic speaker, which was designed to produce high-amplitude, low-frequency sounds. The results of the measurements demonstrate nonlinear effects in the higher harmonics when compared to linear models and provide a relative comparison of nonlinear effects with other propagation effects such as atmospheric and ground. The results also provide a simple data set for evaluation of proposed propagations models without introducing some of the complexities produced by real jet noise sources. One such comparison is provided in a companion paper (Kent L. Gee *et al.* [J. Acoust. Soc. Am 116, 2517 (2004)]). [Work is supported by the Strategic Environmental Research and Develop Program.]

10:15–10:30 Break

10:30

2aPA11. Nonlinearity in outdoor propagation of periodic signals: Prediction model development. Kent L. Gee, Victor W. Sparrow (Grad. Program in Acoust., The Penn State Univ., 217 Appl. Sci. Bldg., University Park, PA 16802), Michael M. James, and J. Micah Downing (Wyle Labs., Arlington, VA 22202)

Propagation measurements made of high-amplitude periodic signals generated by the Army Research Laboratory's Mobile Acoustic Source (MOAS) demonstrate greater energy at high harmonics relative to linear predictions, suggesting the possible influence of nonlinear effects (see Downing *et al.* [J. Acoust. Soc. Am. 116, 2517 (2004)]). An arbitrary-waveform version of an Anderson-type algorithm has been developed in order to compare numerical predictions with the measured MOAS spectra. In general, results demonstrate good agreement between predicted and measured spectra out to 375 m. However, comparisons at greater distances (approximately 1 km) and also for measurements made later in the afternoon exhibit less agreement. For these cases, the Anderson calculations generally overpredict the amount of high-frequency energy present in the measurements. Probable causes for these discrepancies include the effects of propagation at grazing incidence over a finite-impedance ground as well as increased atmospheric turbulence. [Work supported by the Strategic Environmental Research and Development Program.]

10:45

2aPA12. Comparison of computer codes for propagation of high-frequency energy from blast waves. Alexandra Loubeau and Victor W. Sparrow (Grad. Program in Acoust., The Penn State Univ., 202 Appl. Sci. Bldg., University Park, PA 16802, aloubeau@psu.edu)

Current environmental regulation requires that the Department of Defense assess the potential impact of noise from military training on endangered wildlife. One concern is the effect of noise from Army weapons on the hearing of bats. High frequencies are generated from nonlinear propagation of finite-amplitude shock waves created by explosions. These frequencies may be harmful to bats because their auditory systems are sensitive to high-frequency information that they use for flight navigation, communication, and hunting. Determining the spatial extent of high-frequency propagation requires an understanding of the short rise times associated with blast waves. A comparison of computer codes is performed for the most sensitive frequency range of bat hearing, 10–100 kHz. Earlier published results by the authors [A. Loubeau and V. W. Spar-

row, Proc. NOISE-CON 2004, 193–201 (2004)] are compared to numerical solutions of the generalized Burgers equation with molecular relaxation included. The approaches considered are a time-domain method by Cleveland and a hybrid time–frequency-domain algorithm by Anderson. The results are also compared to recent experimental explosion data obtained by the US Army. [Work supported by US Army Engineer Research and Development Center CERL.]

11:00

2aPA13. Height-of-burst influence on spectra emitted by small explosions. Michael J. White (U.S. Army ERDC/CERL, P.O. Box 9005, Champaign, IL 61826)

Explosions in air nearby to the ground produce high-pressure waves that obey nonlinear superposition. The ground-reflected wave, when superimposed on the positive pressure phase of the direct wave, advances to merge in a mach stem. ANSI Standard S2.20 offers a correction factor applied to the real charge weight, to account for the increase in peak overpressure. Over a hard surface, the factor ranges from 2 near the ground to a maximum value of 5.6 at a scaled height (ratio of height to cube root of charge mass) equal to 3.9 m/$[(\mathrm{kg})^{(\frac{1}{3})}]$. We performed a set of experiments to measure the height-of-burst effect from explosions of composition C-4 over grass-covered ground at horizontal distances between 5 and 174 m. We offer an empirical analysis of the resulting spectra, and extend the height-of-burst analysis to far-field spectral sound power.

11:15

2aPA14. Propagation of spark generated N waves through turbulent media. Philippe Blanc-Benon and Sébastien Ollivier (Ctr. Acoustique, LMFA UMR CNRS 5509, Ecole Centrale de Lyon, 69134 Ecully Cedex, France)

It is generally accepted that turbulence plays a role in the propagation and the distortion of sonic booms generated by supersonic flights. This paper describes laboratory experiments in which acoustic shocks from electric sparks have been used to model the propagation of sonic booms [Ollivier *et al.*, 10th AIAA/CEAS Conference, AIAA 2004-2921]. In order to study separately the influence of random fluctuations of temperature and velocity, the N-waves are propagated through a turbulent plane jet or over a heated grid. The recording of 1000 snapshots allows statistical analysis of the variation of the parameters used to describe N-waves, including the maximum and the minimum peak pressures, the rise time, the arrival times and the duration of the wave. In particular, the data show that the increase of the average rise time and the decrease of the average peak pressure with the increase of the rms value of the fluctuations or the increase of the propagation distance could be linked to the probability of occurence of random caustics as observed in numerical simulations [Blanc-Benon *et al.*, J. Acoust. Soc. Am. **111**, 487–498 (2002)]. [This work is partly supported by the European Community SOBER project, Contract No. G4RD-CT-2000-00398 and by the French Ministère de la Recherche et des Nouvelles Technologies.]

11:30

2aPA15. Exploitability of fluctuations to enhance signal processor performance. Kenneth E. Gilbert and Ronald A. Wagstaff (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, rwagstaf@olemiss.edu)

Acoustic literature abounds with experimental and theoretical studies of fluctuations. Fluctuations are often the cause of a signal processors failure to satisfy its objectives, e.g., not achieving enough signal-to-noise ratio (SNR) gain to detect a signal of interest. Fortunately, the spectral time-histories of fluctuating parameters, such as amplitude, are encoded with signal present or signal absent information. Overcoming and exploiting the fluctuations to achieve various forms of gain requires understanding the changes in the fluctuations that are caused by the signal being present and then choosing the appropriate fluctuation-based signal processing algorithm to exploit those differences. First, detrimental influences of fluctuations on the general performance of signal processors will be discussed. Next, differences in the fluctuations for signal present and for signal absent will be addressed, with their corresponding implications for degrading signal processor performance. It will be illustrated that when knowledge of the differences in the fluctuations is applied to signal processing algorithms, higher levels of signal processor performance can be achieved. Improvements include increases in SNR, spatial and spectral resolution, temporal coherence, and unalerted auto-detection of signals of interest. [Work was sponsored by the U.S. Army under Contract No. DASG60-00-C-0061.]

11:45

2aPA16. Three generic signal processing techniques in fluctuation-based processing for increasing gain. Ronald A. Wagstaff and Kenneth E. Gilbert (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, rwagstaf@olemiss.edu)

Fluctation-based processors (FBPs) utilize an approach to signal processing that recognizes there is signal present or signal absent information imbedded in the fluctuations of the spectral time-histories. FBP algorithms are designed to focus on the changes in the fluctuations that are caused by a signal being present. There are three generic classes of FBP. Each class is defined by the fluctuating parameter and the method that the signal processing algorithm uses to exploit it. Typical gains include increases in signal-to-noise ratio, spectral and spatial resolution, and auto-identification of signals. The signal processing versatility of two of the classes provide additional gains by increasing temporal coherence. This versatility also led to development of processors that self-adapt to become more robust and provide higher gains. This includes adaptive processors that utilize the fluctuations to self-tune their governing equations. Thus they are able to preserve desired signals, while eliminating noise, transients, and undesirable signals, e.g., clutter. The operational manner of these three classes of FBP will be described. Key equations and algorithms will be explained. Results from single sensors and arrays for undersea and atmosphere acoustic data will be presented. [Work sponsored by the U.S. Army under Contract No. DASG60-00-C-0061.]

Session 2aSA

Structural Acoustics and Vibration: Analysis Methods: Statistical, Numerical and Analytical

Courtney B. Burroughs, Chair
Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804-0030

Contributed Papers

8:30

2aSA1. A multiscale computational approach for structural and acoustic medium-frequency vibrations. Herve Riou and Pierre Ladeveze (LMT Cachan, 61 Ave. du President Wilson, 94235 Cachan Cedex, France, riou@lmt.ens-cachan.fr)

Today, all major numerical modeling techniques for the analysis of medium-frequency vibrations are based on finite element or boundary element approaches. In order to account for small-wavelength phenomena, these techniques require huge numbers of degrees of freedom. High-frequency approaches, such as statistical energy analysis or any of its improved variations, do not appear to be suitable for medium-frequency vibrations: the resulting vibrational behavior is too smooth and, in general, the coupling loss factor cannot be calculated in a predictive way. The variational theory of complex rays (VTCR) is a predictive computational tool for dealing with medium-frequency vibration problems. This strategy is based on a multiscale treatment of the shape functions (propagative and evanescent waves) chosen to represent the solution. All the waves are taken into account and only the amplitudes of the waves (the slowly varying scale of the solution) are discretized. The boundary and interface conditions between substructures are weakly enforced using an *ad hoc* variational formulation. The performance of the VTCR for an acoustic problem will be assessed.

8:45

2aSA2. The (energy) loss factors in a dynamic system composed of a master dynamic system and an adjunct dynamic system that are coupled. G. Maidanik (Carderock Div., Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817)

The (energy) loss factor (η) of a dynamic system is defined such that together with the stored energy (E) it yields the power Π dissipated in that dynamic system; $(\omega\eta)E=\Pi$, where ω is the frequency and η, E, and Π are functions of ω. An externally driven master dynamic system is defined by the loss factor (η_0) and the modal density (ν_0). The relevant quantities in the absence of coupling are the external input power (Π_e^0), the stored energy (E_0^0), and the power (Π_0^0) dissipated. The conservation of energy demands $\Pi_e^0=\Pi_0^0$. In the presence of coupling Π_e^0 E_0^0 and Π_0^0 become the quantities Π_e, E_0, and Π_0, respectively. The externally driven master dynamic system is coupled to an adjunct dynamic system defined by the loss factor (η_s) and the modal density (ν_s). The relevant additional quantities are the net power (Π_s) transferred from the master dynamic system to the adjunct dynamic system and the stored energy (E_s) and the power (Π_s) dissipated in the adjunct dynamic system. The conservation of energy demands $\Pi_e=\Pi_0+\Pi_s$. A number of (energy) loss factors may be defined to describe the energetics of these coupled dynamic systems. A few asymptotic relationships among these loss factors are cited. [Work supported by ONR.]

9:00

2aSA3. Variance of energy and energy density in statistical energy analysis of complex systems. Vincent Cotoni (ESI R&D, 12555 High Bluff Dr., San Diego, CA 92130) and Robin Langley (Univ. of Cambridge, Cambridge, CB2 1PZ UK)

Standard statistical energy analysis (SEA) predicts the mean energy level in each subsystem of a complex system at high frequencies. The resulting energy is to be seen as averaged over an ensemble of similar systems (like nominally identical products coming off an assemble line). This prediction of the mean has been recently extended to the ensemble variance of the energy. The variance gives indications of how far from the ensemble mean can be the response of one particular member of the ensemble. Using few additional assumptions, the variance can be used to predict the confidence intervals around the SEA mean prediction, and can be extended to the variance of the response at a point (i.e., the energy density). Some numerical validations for several types of vibro-acoustic systems are presented.

9:15

2aSA4. A 3-D hierarchic finite-element tool with infinite elements for structural acoustic scatter modeling. Mario Zampolli, David S. Burnett, Alessandra Tesei (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy), John B. BlottmanIII (Naval Undersea Warfare Ctr., Newport, RI 02841), and Timothy A. Westermann (2219 Parkland Cove, Round Rock, TX 78681)

Fully three dimensional prolate spheroidal acoustic infinite elements have been added to a steady state finite-element structural acoustics tool (FESTA), which is based on hierarchic polynomial shape functions. The infinite elements are used as an alternative to the Bayliss–Turkel approximate radiation condition to ensure outward propagation of the scattered and/or radiated acoustic fields surrounding one or more objects. The main advantages of the infinite elements compared to the Bayliss–Turkel condition are the reduction of the distance required between the object(s) and the boundary of the finite element computational domain, and meshes which can be used over a broader frequency band. Furthermore, the infinite element formulation adopted, which is based on the prolate spheroidal version of the Atkinson–Wilcox expansion, is ideal for modeling the scattering and/or radiation from elongated structures of particular interest to the underwater acoustics community. The performance of the tool is assessed by comparison to an axisymmetric thin-shell FE/virtual source code, and to a hierarchic axisymmetric FE code with infinite elements. Furthermore, the tradeoffs between the infinite elements and the prolate spheroidal Bayliss–Turkel conditions are assessed.

9:30

2aSA5. Solving for the vibro-acoustic interaction between a fluid and a structure with discontinuities by the boundary element method using local/global homogenization. Pavel Danilov and Donald Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708, pvd2@duke.edu)

The boundary element method (BEM) was used to determine the radiation properties of finite structures with periodically attached impedances. Due to the discontinuities and fine-scale structural wave motion, BEM would generally require small size elements. Local/global homogenization (LGH) was used in order to eliminate the discontinuities and short waves from the problem. LGH provides the global smooth problem, free of discontinuities, but containing all the information needed to describe the radiation. The global problem results in modification of BEM kernels in order to implement coupling with the local part. Instead of the exact solution, BEM used with LGH requires only the smooth global part, allowing fewer elements. Model problems show good accuracy of the method along with a significant reduction of computational costs.

9:45

2aSA6. Reducing the dispersion error in acoustic wave modeling with modified integration rules. Bin Yue and Murthy N. Guddati (Dept. of Civil Eng., North Carolina State Univ., Raleigh, NC 27695, byue@ncsu.edu)

Numerical analysis of transient acoustic wave propagation is often performed using finite element or finite difference methods along with central difference time stepping. Such analysis results in wave velocities that are different from exact ones. For conventional algorithms, this dispersion error is second order with respect to discretization parameters. In this paper, modified integration rules are combined with a modified central difference scheme to reduce this error to fourth order. Essentially, the locations of integration points used for evaluating the system matrices are selected such that the wave velocity error is minimized. It was found that the optimal integration points are square roots of 2/3 for the stiffness matrix, which eliminates the anisotropy in the second-order error. The optimal integration points for mass matrix, which depend on the wave velocity, the mesh size, and the time-step size, are also identified to completely eliminate the second-order error. The resulting dispersion error is thus reduced to fourth order. Numerical experiments illustrate that the proposed method has superior performance over existing methods, not only for structured square meshes, but also for unstructured meshes.

10:00–10:15 Break

10:15

2aSA7. Sound radiation from rectangular plates with elastic boundary restraints. Wen Li (Dept. of Mech. Eng., Mississippi State Univ., Mississippi State, MS 39762, li@me.msstate.edu)

Sound radiation from rectangular plates is of considerable interest to both researchers and engineers. Different methods and techniques have been developed for determining the sound radiated from plates with various complicating effects. However, the previous investigations are mostly focused on the simply supported plates, although it is widely known that the radiation efficiency of a plate will be strongly dependent upon its boundary condition. In this study, a general analytical method is presented for the calculation of the sound power radiation from a rectangular plate that is elastically restrained along its edges. The vibration of the elastically restrained plates is expressed in terms of a generalized Fourier series expansion. Thus, many of the existing solutions for the sound radiation from the simply supported plates can be directly applied to the plates with general elastic boundary restraints. The radiation characteristics are compared to the plates under different boundary conditions.

10:30

2aSA8. Acoustic radiation by a submerged source in shallow water—Influence of the boundaries. Emmanuel Charmes, Pierre-Philippe Beaujean (Dept. of Ocean Eng., Florida Atlantic Univ., Seatech, 101 North Beach Rd., Dania Beach, FL 33004), and Joseph M. Cuschieri (Perry Technologies, Riviera Beach, FL 33404)

The acoustic radiation and scattering characteristics of an acoustic source in an unbounded deep water environment are different from those in a shallow water environment, where one has to take into account the influence of the boundaries on the source radiation impedance. In this paper the influence of the waveguide (surface and bottom of the ocean in a shallow water environment) on the radiation from an elastic noise source is investigated by means of a superposition method. The radiated sound is initially determined from the scattering of an incident wave from the elastic target. Using the superposition method [A. Sarkissian, J. Acoust. Soc. Am. (1994)], the scattered field, including multiple scattering effects, is computed and represented by a series of point sources positioned inside the scatterer. A number of calibration points located at the surface of the scatterer are selected to determine the source strength of each point source using a least square approximation. Having obtained the source strength of each point source, the combined propagation field in the waveguide from all point sources and the influence back on the source are determined using normal mode or ray tracing solutions common in underwater sound propagation. [Work supported by ONR.]

10:45

2aSA9. Axisymmetric acoustic radiation from submerged prolate spheroidal shells. Jeffrey E. Boisvert (NAVSEA Newport, Newport, RI 02841) and Sabih I. Hayek (Dept. of Eng. Sci. and Mech., Penn State Univ., University Park, PA 16802)

The equations of motion for nonaxisymmetric vibration of prolate spheroidal shells of constant thickness were derived using Hamiltons principle [S. I. Hayek and J. E. Boisvert, J. Acoust. Soc. Am. **114**, 2799–2811 (2003)]. The shell theory used in this derivation includes shear deformations and rotatory inertias. The shell displacements and rotations were expanded in infinite series of comparison functions. These include associated Legendre functions in terms of the prolate spheroidal angular coordinate and circular functions in the azimuthal angle coordinate. For axisymmetric vibration of a submerged shell, the external (heavy) fluid loading impedance was computed using expansions of prolate spheroidal wavefunctions. The shell was excited by axisymmetric normal surface forces, including a point load at the shell apex and ring load at other locations. Far-field radiated pressure spectra are presented for several shell thickness-to-half-length ratios ranging from 0.005 to 0.1, and for various shape parameters, a, ranging from an elongated spheroidal shell ($a = 1.01$) to a spherical shell ($a \sim 100$). The far-field directivity of acoustic radiation is presented at selected frequencies. [Work supported by the NUWC ILIR Program and the ONR/ASEE Summer Faculty Research Program.]

11:00

2aSA10. Local/global decomposition for fluid-loaded periodic structures. Pavel Danilov and Donald Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708, pvd2@duke.edu)

The problem of interaction between a fluid and a structure with periodic impedance attachments is considered. The full solution is divided into two coupled parts: the global part, providing overall structural motion and acoustic radiation, and the local part, containing small structural oscillations and evanescent pressure modes. Fluid loading in the local part results in additional structural inertia, which is conveniently accounted for in the global problem. A separate stand-alone equation is constructed for the global solution. The local part can be recovered afterwards. Specific conditions are imposed on the local and global parts, e.g., the global solution with finite wavenumber support is used to recover the radiating part of the

spectrum and the local solution is made to vanish at the impedances. Thus, the global solution describes radiating properties and structural response. Different conditions on the local/global decomposition are considered in order to enhance the accuracy of the method. Application to structures of finite extent is discussed. Model problems showed good accuracy of radiated pressure as well as structural response.

11:15

2aSA11. Analysis of wave dispersion in cylindrical shells with anisotropic layers by Rayleigh–Ritz expansions. Elizabeth A. Magliula (Dept. of Aero. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215), J. Gregory McDaniel, and Charles N. Corrado (Appl. Physical Sci. Corp., New London, CT 06320)

Recent advances in composite materials offer the potential to dramatically affect the dispersions of helical waves that propagate in layered cylindrical shells. Effective choices and configurations of these materials require rapid dispersion predictions that take into account general anisotropy and variations of elastic variables through the thickness of each layer. This presentation describes a solution to this problem that is based on a previously developed theory [J. G. McDaniel and J. H. Ginsberg, J. Appl. Mech. **60**, 463–469 (1993)] for the vibrations of cylindrical shells with isotropic layers. The approach uses propagating wave representations in the axial and circumferential coordinates. A series expansion with polynomial basis functions represents dependences on the radial coordinate and equations for the series coefficients are derived using the Rayleigh–Ritz method. These equations yield a dispersion relation that is solved for the complex-valued axial wavenumber at each frequency and circumferential harmonic. Examples illustrate the interface of this approach with optimization algorithms that identify layer designs possessing specified dispersion properties. [Work supported by ONR.]

11:30

2aSA12. Characterizing close fitting enclosures for fluid-loaded cylindrical sources. Joseph M. Cuschieri (Lockheed Martin MS2, Perry Technologies, 100 East 17th St., Riviera Beach, FL 33404, joe.cuschieri2lmco.com)

The influence of a close fitting enclosure on the radiated noise from a submerged (in water) acoustic source has been shown to be beyond the scope of simple close fitting enclosure expressions found in some textbooks, especially when dealing with relatively soft intermediate layers [Cuschieri, J . Acoust. Soc. Am. **115**, 2537 (2004)]. A more detailed analysis is thus warranted. Using a multilayer shell approach, with the innermost layer representing the source with uniform radial velocity, the middle layer representing the intermediate isolation layer and the outside layer representing the exterior cladding, with water fluid-loading on the outside of the exterior cladding, the insertion loss (IL) of the multiple layers is investigated for different intermediate layer materials, different exterior cladding layer characteristics and source size. IL results are presented as a function of frequency and layer/cladding characteristics. From the formulation the behavior of the intermediate layers is also obtained which help explain the IL results. [Work sponsored by ONR.]

TUESDAY MORNING, 16 NOVEMBER 2004 CALIFORNIA ROOM, 9:00 A.M. TO 12:00 NOON

Session 2aSC

Speech Communication: Clinical and Developmental Perspectives on Speech (Poster Session)

Peggy B. Nelson, Chair

Communication Disorders, University of Minnesota, 164 Pillsbury Drive, SE, Minneapolis, Minnesota 55455

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon.

2aSC1. An intonational analysis of disfluency patterns in chronically stuttered speech. Timothy R. Arbisi-Kelm (UCLA, Los Angeles, CA)

While previous stuttering research has successfully revealed areas vulnerable to disfluency at the word level in stuttering, identifying the specific factors responsible for this instability has proved difficult; moreover, inconsistent results are complicated by a failure to control for the effects of phrasal prosody, which govern such word-level factors as lexical stress. The present experiment tested the hypothesis that disfluencies in stuttering are directly proportional to the prominence-level of a given production. Three stuttering subjects participated in an oral sentence-reading task testing a variety of sentence types while manipulating intonational factors such as pitch accent type and location. It was anticipated that pitch-accented syllables, representing a higher degree of stress in an intonation phrase than stressed but non-pitch-accented syllables, would be most prone to triggering disfluency, since they bear a greater level of prominence in the utterance. The results of the study confirmed the major hypothesis: in all of the comparisons between pitch-accented and non-pitch-accented positions of stress, the former attracted the highest rate of disfluent speech productions. This supports the principal hypothesis that intonationally prominent domains, not simply lexically stressed syllables, are a better indicator of unstable positions in stuttered speech.

2aSC2. Effects of a prosodic control approach for patients with motor speech disorders. Noriko Kobayashi, Hajime Hirose, Minako Koike, Yuki Hara (Sch. Allied Health Sci., Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Japan 228-8555), Hiroki Mori, and Hideki Kasuya (Utsunomiya Univ., 7-1-2 Yoto, Utsunomiya, Japan 321-8585)

For patients with motor speech disorders, the Lee Silverman method (Ramig, 1997) has been known as an effective voice therapy method. In our clinical experiences, however, some Japanese patients could not easily produce very loud voice required in the method, presumably due to the cultural background. Therefore, in this study, a prosodic control approach called ''the intonation emphasis therapy'' was used as well as the Lee Silverman method for three types of patients. The disorder types were

amyotrophic lateral sclerosis (ALS), Parkinson's disease (PK), and olivo-ponto-cerebellar atrophy (OPCA). Acoustic analyses revealed wider $F0$ ranges after the intonation emphasis therapy than the Lee Silverman method for three types of the patients. In the perceptual judgments by speech therapists regarding articulation, voice quality, intonation, and abnormal impression of speech, better ratings were obtained for the intonation emphasis therapy than for the pretherapy speech and the Lee Silverman method in the patients with ALS and PK. However, the listeners perceived inappropriate intonation and increased abnormality for the OPCA patient after the intonation emphasis therapy. It was suggested that our therapy method was effective for patients with motor speech disorders unless the disorders were associated with poor muscular coordination such as OPCA.

2aSC3. Vowels and final consonant production by adults with Down's syndrome. Mary Fung, Rizzah Decopain, and Nancy McGarr (St. John's Univ., Jamaica, NY 11439, mcgarrn@stjohns.edu)

Speech production studies have suggested a typical voice quality as well as vowel and consonant errors in speakers with Down's syndrome. In this study, the production of bVt or bVd words produced by ten adults with Down's syndrome and age and gender matched controls was examined for influence of final consonant voicing status. Acoustic measures of vowel duration and consonant voicing were made. Listener judgments were also obtained. Contrasts between vowels pairs were maintained by most subjects, e.g., duration for tense vowels was greater than for lax, although duration for the test subjects productions was greater overall than for the controls. Similarly, consonant voicing contrasts were maintained, e.g., vowel duration preceding voiced consonants was longer than for voiceless, but again overall duration was longer for the Down's syndrome adults than the controls. The results of the acoustic analyses suggest that while the test subjects demonstrate knowledge for producing final voicing contrasts, aspects of fine coordination are not achieved. Intelligibility of final consonants was high and listener judgments generally agreed with acoustic measures. [Work supported by McNair Scholars' Program to St. Johns University.]

2aSC4. Articulatory characteristics of speakers with apraxia of speech during sentence production at different speaking rates. Michiko Hashi (Health Sci., Univ. of Hokkaido, Sapporo Hokkaido, Japan), Katharine Odell, H (Meriter Hospital, Madison WI USA), Ryoko Hayashi (Kobe Univ., Kobe, Japan), and Takeshi Nakayama, R (Health Sci., Univ. of Hokkaido, Sapporo Hokkaido, Japan)

The present study developed quantitative descriptions of articulatory movements in three speakers with apraxia of speech during sentences produced at different speaking rates. Point-parameterized articulatory data were obtained using the x-ray microbeam (XRMB) technique, and were compared with similar materials from 24 normal young-adult speakers of American English drawn from an existing XRMB database. Speed histories of markers attached to the lips and jaw during production of the test sentence ''The other one is too big.'' spoken at slow, habitual and fast rates were analyzed to determine mean peak speed, the number of speed peaks, and the timing of particular speed peaks for each sentence replicate by each talker. An attempt was made to derive speed histories for the lip marker expressed relative to concurrent movements of the jaw, to evaluate the relative contribution of jaw speed to lip speed in both talker groups. A description and discussion of results from the analysis will emphasize speaking rate effects on articulator speed and inter-articulator timing.

2aSC5. Motherese and Chinese: Evidence of acoustic changes in speech directed at infants and foreigners. Monja Knoll and Maria Uther (Dept. of Psych., Univ. of Portsmouth, King Henry Bldg., King Henry 1 St., Portsmouth, PO1 2DY, UK)

Infant-directed speech (IDS) is characterized by hyperarticulation, increased pitch, and high emotional affect, which is in turn thought to reflect a linguistic and emotional role for IDS. If the linguistic role is an independent contributor of the changes in IDS, then similar hyperarticulation should also occur in foreigner-directed speech (FDS) but without positive affective changes. To test this, nine mothers were recorded talking to their infants, and British- and foreign-adult confederates. Mothers were provided with toys to elicit the target words shark, sheep, and shoe, containing the corner vowels /a/, /i/, and /u/. Speech samples were analyzed to determine mean pitch and formant 1 and 2 ($F1/F2$) values of target vowels. Low-pass filtered speech samples of the mother's interactions in each condition were rated on positive and negative vocal affect by 24 independent raters. Results showed that mothers hyperarticulated their vowels (indexed by $F1/F2$ values) in both IDS and FDS relative to ADS. Furthermore, mean pitch was highest in IDS compared to FDS and ADS. Positive affect was highest in IDS and lowest in FDS compared to ADS. These data suggest a linguistic function for prosody modifications in FDS and IDS that is independent of affective changes.

2aSC6. Strategies used to increase speech clarity by normal-hearing children. Dana L. Ide-Helvie, Elizabeth A. McCleary, Sarah C. Sullivan, Andrew J. Lotto (Boys Town Natl. Res. Hospital, 425 North 30th St., Omaha, NE 68131, lottoa@boystown.org), and Maureen B. Higgins (Boys Town Natl. Res. Hospital, Omaha, NE 68131)

It is common among clinicians to ask children to produce their best speech during intervention. However, it is unclear whether children know how to make their speech clearer. The strategies used by children with and without hearing loss have implications for maximizing intelligibility and for understanding the development of communication competency. As a first step toward this understanding, children (7 to 12 years of age) with normal hearing were asked to read ten simple sentences. They were told that they were testing a new computer program designed to recognize speech. There was, in fact, no recognition program and each child received the same output feedback. After providing normal speech to allow the program to get used to their voices, they subsequently produced their best and then very best, very clearest speech in order to see how accurate the recognition program could be. Acoustic analyses (intensity, fundamental and first two formant frequencies for all vowels, as well as sentence, vowel, and VOT durations) were performed on recorded waveforms from each repetition in order to determine what the children were varying to comply with best speech instructions. The results demonstrate large individual differences in strategies and persistent gender differences. [Work supported by NIH.]

2aSC7. Speech perception in noise among language-impaired individuals. Ratree Wayland and Linda Lombardino (Program in Linguist., Univ. of Florida, Dept. of Commun. Sci. and Disord., Univ. of Florida, Gainesville, FL 32611-5454)

This study compared and contrasted the ability to process speech signals between language-impaired individuals and their age-matched normal controls. Participants were administered a speech categorization task and a sentence processing task. A synthetic /ba–da/ continuum varying in second formant onset frequencies was used in the speech categorization task and 64 short (4–5-word-long) English sentences were used in the sentence processing task. The stimuli were presented with and without background noise to participants for categorization and identification. The /ba–da/ categorical boundaries and the number of target words correctly identified by

two groups of participants under the two presentation conditions were compared. A connection between phonological processing and language impairment was discussed.

2aSC8. An echological theory of language learning. Francisco Lacerda and Ulla Sundberg (Dept. of Linguist., Stockholm Univ., Stockholm, Sweden)

An ecological theory of language acquisition (ETLA), attempting to account for the linguistic development during about the first 18 months of age, is proposed in this contribution. The theory does not assume that the infant is endowed with specialized linguistic devices or strategies at the onset of life. ETLA considers instead the multi-sensory aspects of the adult-infant interaction and the typical ecological setting of that interaction. Rather than focusing on the speech signal *per se*, ETLA considers the infant's multi-sensory exposure to the phonetic, prosodic, syntactic and semantic characteristics along with visual, tactile, olfative and other sensory information as a key to the spontaneous emergence of linguistic structure early in life—a structure implicit in the adults use of the language, and that is partly simplified by the infant's limited production, perception and representation capacities. A functional model illustrating how such general-purpose mechanisms interacting with the typical sensory input available to the infant may lay the ground for linguistic structure will be presented for discussion. [Research supported by grants from the Swedish Research Council and the Bank of Sweden Tercentenary Foundation.]

2aSC9. Children's reliance on spectral information for understanding speech in noise. Peggy Nelson and Yingjiu Nie (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, nelso477@umn.edu)

Children have significantly greater difficulty understanding in noisy situations than do adults. This has serious implications for the design and structure of classrooms and for communicating effectively with young children. The reasons for children's susceptibility to noise are not well understood. One factor may be children's reliance on frequency cues for speech understanding. This project investigated children's perception of speech in noise by systematically evaluating the perceptual importance of speech frequency information. Typically developing children (ages 7 to 13 years) and adults were tested using full spectrum speech as well as 4-, 8-, and 12-band simulations of cochlear implant processing (after Shannon *et al.*, 1995). Stimuli were color and number words selected from the coordinate response materials (Brungart, 2001), and were presented in quiet and in pink noise at +5-dB signal-to-noise ratio. Preliminary results support the hypothesis that children rely more on the frequency content of speech than do adults for understanding speech in noise, and that children need fuller frequency representations of speech than do adults to segregate speech from background noise. [Work supported by the Graduate School of the University of Minnesota.]

2aSC10. Phonemic contrasts in a mutiple talker task with hearing impaired listeners. Mark A. Ericson (Air Force Res. Lab., Wright–Patterson AFB, OH) and Pamela J. Mishler (Dept. of Veteran Affairs, Dayton, OH)

Word identification ability of mild to moderately hearing impaired listeners was measured in both spatial and nonspatial multiple talker tasks. A digital recording of the modified rhyme test was made for three male talkers. The beginning of the test word in each phrase was synchronized using Cool Edit and equalized to the same average rms power. The phrases were processed for binaural presentation using nonindividualized HRTFs with Tucker Davis Technology System 3 equipment and presented over Sennheisser 600 headphones. The phrases were presented in random order at a self-paced manner. Five mild and five moderately hearing impaired subjects, ranging in age from 24 to 54 years, participated in this study. Results were analyzed by six attributes: place, voicing, nasality, frication, duration, and stops. Phoneme identification performance was ranked based on most salient (most information transferred) to least salient (least information transferred). Spatial separation had no effect on the relative importance of the speech attributes. Duration was the most salient cue for hearing impaired listeners, followed by voicing and frication. Stops and nasality were weak cues and place was by far the least important cue.

2aSC11. Perceptual overshoot in listeners with cochlear implants. Radhika Aravamudhan and Andrew J. Lotto (Boys Town Natl. Res. Hospital, Omaha, NE 68131)

Perceptual overshoot (PO), a phenomenon in which the boundary for perceived vowel categories shifts as a result of preceding formant transitions, has been demonstrated previously with synthetic vowels in CV contexts in normal-hearing individuals. In one of the previous studies by the author, the same phenomenon was tested with sinewave analogs that mimicked the synthetic vowels. The results demonstrated that PO could be elicited for sinewave analogs after training on categorization of sinewave steady states. These findings suggest that PO may be partly mediated by general processes in the auditory system. In the current study, subjects with cochlear implants were presented with synthetic vowel and sinewave continua with and without transitions. It was predicted that PO would be greatly diminished or absent for this population because of the degraded nature of the spectral input. Implications of the results for theories of speech perception will be presented.

2aSC12. Fundamental frequency and the intelligibility of competing sentences with cochlear implant processing. Ginger S. Stickney (Dept. of Otolaryngol., Univ. of California, Irvine, 364 Med Surg II, Irvine, CA 92697), Janice Chang (Univ. of California, Berkeley, CA 94720-1762), Peter F. Assmann (Univ. of Texas at Dallas, Richardson, TX 85083-0688), and Fan-Gang Zeng (Univ. of California, Irvine, CA 92697)

Speech perception in the presence of another competing voice is one of the most challenging tasks for cochlear implant users. Several studies have shown (1) that the fundamental frequency ($F0$) is a useful cue for segregating competing speech sounds and (2) the $F0$ is better represented by the temporal fine structure than the temporal envelope. However, current cochlear implant speech processing algorithms emphasize the temporal envelope information and discard the temporal fine structure. In this study, speech recognition was measured as a function of the $F0$ separation of the target and competing sentence in normal-hearing and cochlear implant listeners. For the normal-hearing listeners, the combined sentences were processed through either a traditional cochlear implant simulation or a new algorithm which additionally extracts a slowed-down version of the temporal fine structure. The results showed no benefit of increasing $F0$ separation for the cochlear implant or traditional simulation groups. However, as found with unprocessed sentences presented to normal-hearing listeners, the new algorithm resulted in gradual improvements with increasing $F0$ separations. These results demonstrate that inadequate coding of the temporal fine structure in current speech processing algorithms may complicate the segregation of competing speech sounds. [Work supported by NIDCD F32DC05900 and R01DC02267.]

2aSC13. Individual differences in auditory discrimination of spectral shape and speech-identification performance among elderly listeners. Mini N. Shrivastav[a)], Larry E. Humes, and Diane Kewley-Port (Dept. of Speech and Hearing Sci., Indiana Univ., 200 South Jordan Ave., Bloomington, IN 47405, mnarendr@csd.ufl.edu)

Speech-understanding difficulties observed in elderly hearing-impaired listeners are clearly associated with recognition and discrimination of consonants, particularly within consonants that share the same manner of articulation. Spectral shape is an important acoustic cue that serves to distinguish such consonants. The present study examined whether indi-

vidual differences in speech understanding among the elderly could be explained by individual differences in spectral-shape discrimination ability. This study included a group of 20 elderly hearing-impaired listeners, with a group of young normal-hearing adults also included for comparison purposes. All subjects were tested on speech-identification tasks, with natural and computer-synthesized speech stimuli, and on a series of spectral-shape discrimination tasks. The young normal-hearing adults performed better than the elderly listeners on many of the identification tasks and on all but two discrimination tasks. Regression analyses on data from the elderly listeners revealed moderate predictive relationships between some of the spectral-shape discrimination thresholds and speech-identification performance. The results indicated that when all stimuli were at least minimally audible, some of the individual differences in the identification of natural and synthetic speech tokens by elderly hearing-impaired listeners could be attributed to the differences in their spectral-shape discrimination abilities for similar sounds. [a] Currently at Dept. of Commun. Sci. and Disord., 336 Dauer Hall, Univ. of Florida, Gainesville, FL 32611.

2aSC14. Interactions between the frequency allocation, stimulation mode, and pitch perception on speech recognition by cochlear implant listeners. John GalvinIII and Qian-Jie Fu (House Ear Inst., 2100 W. 3rd. St., Los Angeles, CA 90057)

In cochlear implants, speech recognition depends strongly on the acoustic frequency-to-electrode assignment. In a previous experiment, the pitch of an electrode was shown to vary with stimulation mode, especially for widely spaced electrode pairs. In the current study, Nucleus-22 users' speech recognition was measured as functions of the stimulation mode and frequency allocation. Two stimulation modes were tested: BP+1 and ''mixed-mode'' [the active electrode was fixed and the return electrode varied, i.e., (2,22), (2,21), etc.]. A range of frequency allocations was tested to measure the effect of spectral mismatch between the acoustic input and the electrode configurations. Preliminary results showed a differential effect of frequency allocation, depending on the stimulation mode used in the speech processor. For the BP+1 configuration, multi-talker vowel recognition decreased significantly as the frequency allocation was shifted from Table 9 (150–10 823 Hz) to Table 3 (85–6184 Hz). However, for the mixed-mode processor, performance improved as the allocation was shifted from Table 3 to Table 9. Because Table 3 may have produced less spectral mismatch (given the generally lower pitch percepts for widely spaced electrode pairs), mixed mode processors may be used to extend the pitch range and thereby map more low-frequency spectral content to apical electrodes.

2aSC15. Cortical networks underlying audio-visual speech perception in normal-hearing and hearing impaired individuals. Julie J. Yoo (Speech and Hearing Biosci. and Technol. Program, Harvard-MIT Div. of Health Sci. & Technol., Cambridge, MA 02139), Frank H. Guenther (Boston Univ., Boston, MA 02215), and Joseph S. Perkell (MIT, Cambridge, MA 02139)

Functional magnetic resonance imaging (fMRI) was used to investigate the brain activity underlying audio-visual speech perception in normal-hearing and congenitally deaf individuals. Data were collected while subjects experienced three different types of speech stimuli: audio stimuli without visual input, video of a speaking face without audio input, and video of a speaking face with audio input. A control condition consisted of viewing a blank screen. The stimuli were vowels or CVCV syllables, presented in different blocks. Active brain regions for normal-hearing subjects during the visual-only condition included visual cortex, angular gyrus, fusiform gyrus, and auditory cortex, as well as premotor areas in the frontal cortex. The pattern of activation found for deaf subjects while viewing visual-only stimuli was similar to that of normal hearing subjects, but showed distinctly more activity in the right hemisphere (for both vowels and CVCVs), and far less activity in premotor and parietal cortex. Interestingly, the pattern of activity for deaf subjects in the visual-only case was found to be similar to normal-hearing subject activation in the audio-visual case. Finally, effective connectivity analyses were done to investigate connectivity between brain regions in the different conditions for both groups of subjects. [Research supported by NIDCD.]

2aSC16. Can context diminish the effects of rapid speech recognition in older listeners? Izumi Furukawa, Nancy Vaughan, and Daniel Storzbach (Natl. Ctr. for Rehabilitative Auditory Res., Portland VA Medical Ctr., 3710 SW US Veterans Hospital Rd., Portland, OR 97207)

This is a part of a larger study investigating the effects of cognitive slowing on speech perception among older adults. In addition to neurocognitive tests (working memory, attention, and speed of processing), time-compressed speech was used to evaluate the effects of age on rapid speech recognition. Two different speech materials (with and without contextual cues) were prerecorded and time-compressed to four different rates (40%, 50%, 60%, 65%) in order to increase the demand on processing speed. All sentences were presented monaurally through MedRx Otowizard via ER3 insert phones to older adults (50 to 75 year olds) with normal hearing and with mild to moderated hearing loss. For the listeners with hearing loss, the master hearing aid function with NAL-NP target was used. At the slower time-compression rate (40%), there was no significant difference between the two groups. However, at faster rates, the hearing loss group performed significantly worse than the normal hearing group regardless of contextual cues. These results suggest that neither of the older groups was able to benefit from the use of context in these rapid speech materials. Implications of these speech recognition test results will be discussed with regard to the neurocognitive findings.

Session 2aSP

Signal Processing in Acoustics and Underwater Acoustics: Signal Processing Arrays with Many Elements in Novel Configurations or Novel Environments Part I

Jens M. Meyer, Cochair
20 River Terrace, New York, New York 10282

David I. Havelock, Cochair
National Research Council, IMS/ASP, Montreal Road, Ottawa, Ontario, K1A 0R6, Canada

Invited Papers

8:30

2aSP1. Processing microphone arrays for spatial selectivity in three dimensions. J. L. Flanagan (Ctr. for Adv. Information Processing, Rutgers Univ., 96 Frelinghuysen Rd., Piscataway, NJ 08854, jlf@caip.rutgers.edu)

Reverberation degrades the performance of conventional beamformers. Matched-filter processing combats multi-path distortion and provides improved quality by utilizing signal information arriving from different paths. Implicit is spatial selectivity in three dimensions—which enables an array to ''reach over'' intervening sources of interference. The cost is arithmetic, storage, and *a priori* characterization of the enclosure. Source-to-receiver transmissions are described by impulse responses, which are employed in causal time-reversed form to accomplish matched filtering. Work conducted earlier for the National Science Foundation is summarized and demonstrated, along with a description of algorithms for automatic tracking of moving talkers. Array configuration and placement remain challenges in geometric optimization.

9:00

2aSP2. A scalable spherical microphone array for spatial sound capture. Gary W. Elko (Avaya Labs, 233 Mt. Airy Rd., Basking Ridge, NJ 07920, gwe@ieee.org) and Jens Meyer (mh acoustics, New York, NY 10282)

With recent popular consumer acceptance of surround-sound audio, there is the need to develop new technologies that will allow for more accurate sound-field recording commensurate with or exceeding current multichannel audio playback systems. Real-world playback systems can have widely varying geometries and number of channels due to space and aesthetic requirements. Thus, there is a need to have a general recording scheme that allows for computationally simple modifications of the signals to optimize the playback for any geometry and number of playback speakers. One would also like to have a recording system that is scalable for future advances in spatial audio playback technology. A spherical microphone array has been proposed as a solution to these requirements. A spherical array has the desired property of enabling a beam pattern that is steering direction invariant with a relatively simple beamformer structure. The spherical array topology enables scalability and an elegant and mathematically straightforward beamformer design for efficient sampling and processing of the acoustic field. This talk will describe how the spherical array beamformer decomposes the sound field into compact multichannel representations that allow computationally efficient transformations to uniquely tailor the playback signals to any specific playback geometry.

9:30

2aSP3. Two-dimensional wave field decomposition using circular microphone arrays and its application to acoustic source localization. Heinz Teutsch and Walter Kellermann (Multimedia Commun. and Signal Process., Univ. of Erlangen-Nuremberg, Cauerstr. 7, 91058 Erlangen, Germany, teutsch@lnt.de)

Two-dimensional wave fields can be used as reasonable models for propagating acoustic sound fields in closed rooms where ceiling and floor reflections are sufficiently attenuated. A natural way of analyzing a 2D wave field is to decompose it into an orthogonal set of eigensolutions to the acoustic wave equation in cylindrical coordinates, i.e., the cylindrical harmonics. These harmonics can be shown to be the coefficients of a Fourier series expansion applied along a circular aperture that is located within the 2D wave field under observation. Circular microphone arrays and, in particular, microphones mounted in a rigid cylindrical baffle that perform the cylindrical harmonics decomposition will be presented and its advantages and limitations will be discussed. As a specific example for utilizing the wave field decomposition approach, its application to acoustic source localization will be considered. By exploiting structural similarities between the response of linear microphone arrays and the decomposed response of circular microphone arrays, most well-known subspace tracking algorithms can be applied to this problem with only minor modifications. It will be shown that acoustic source localization based on wave field decomposition has the potential to track multiple simultaneously active sources in the array's full 360° field of view.

10:00–10:30 Break

10:30

2aSP4. Sectorized beam-base adaptation for long line arrays. Henry Cox (LM Orincon Defense, 4350 North Fairfax Dr., Ste. 470, Arlington, VA 22203)

Long line arrays of many elements present special problems for adaptive beamforming. The long travel time for an endfire signal to traverse the array requires the use of narrow frequency bins to avoid a significant phase change within a frequency bin. The narrow beamwidth near broadside makes the array sensitive to small bearing charges that occur within the coherent integration time so that adaptation must take place rapidly. These two effects combine to cause a snapshot-starved situation. A beam-based adaptation approach is suggested in which the first step is to form shaded conventional far-field beams. This is followed by a sectorized adaptive combination of beams in a sector to provide adaptive near-field and far-field beams. For a sector near broadside where the beams are very narrow and sensitive to motion but the travel time across the array is short, less temporal averaging and more frequency averaging is used. Near endfire, where the beams are wide but the travel is long, narrow frequency beams are used with longer temporal averaging. The approach is applied to the case of strong far-field interference masking a weak near-field signal. The approach is computationally efficient as well as providing improved snapshot support for adaptation.

11:00

2aSP5. Time-reversal arrays in underwater acoustics. Philippe Roux, Tuncay Akal, W. S. Hodgkiss, W. A. Kuperman, Hee Chun Song (Marine Physical Lab., Scripps Inst. of Oceanogr., UC San Diego, La Jolla, CA, 92093), and Mark Stevenson (NATO Undersea Res. Ctr., La Spezia, Italy)

Since 1996, several time-reversal experiments have been performed in shallow water acoustics involving source and receive arrays made of many elements. This equipment allowed us to reach different goals among which were long-range time-reversal focusing, passive and active multiple input multiple output communications, bottom refocusing from reverberation, acoustic barrier detection, comparison between reciprocal and nonreciprocal time-reversal and shallow water tomography. This presentation will review this work with emphasis on array signal processing.

Contributed Papers

11:30

2aSP6. Experimental demonstration of time reversed reverberation focusing in an oceanic waveguide. Karim Sabra, Philippe Roux, Hee-Chun Song, William Hodgkiss, William Kuperman, Tuncay Akal (MPL-SIO, UCSD. 9500 Gilman Dr., La Jolla, CA 92093-0238), and Mark Stevenson (NATO Undersea Res. Ctr., La Spezia, Italy)

The robust focusing and pulse compression provided by time reversal techniques may be exploited for active sonar. Backscattering from the rough water–bottom interface can be used as a surrogate probe source [Lingevitch *et al.*, J. Acoust. Soc. Am. **111**, 2609–2614]. This method uses a selected time-gated portion of the reverberation signals (or backscattered signals) to provide a transfer function between a time-reversing array (TRA) and a corresponding range interval on the bottom. The potential of a TRA for studying the application of shallow water bottom scattering is investigated experimentally using both sea-data collected in July 2004 in the Mediterranean sea off the Italian coast, and ultrasonic water tank data.

11:45

2aSP7. Processing of reverberation data from 400–3500 Hz using line arrays with left/right discrimination. John Preston (Appl. Res. Lab., The Penn State Univ., State College, PA, 16804, preston@ciao.arl.psu.edu)

The author recently participated in two experiments using directional towed arrays. One was the 2004 boundary characterization experiment near the Malta Plateau. That experiment was led by the NATO Undersea Research Centre (NURC). The area is rich in clutter objects like wrecks and mudvolcanoes and has some sub-bottom features that may be important. Sources were monostatic coherent pulses and SUS. The receiver was the NURC cardioid array. The other experiment was ONR's 2003 Geoclutter effort to study shallow-water bottom reverberation and clutter in the STRATAFORM off New Jersey. That experiment was led by M.I.T. Sources were bistatic coherent pulses. The receiver was the five-octave research array (FORA). The STRATAFORM is known to have benign surface morphology but strong clutter is observed. Some highlights of the reverberant returns are discussed that include the correlation of returns with manmade targets and probable fish schools. The purpose of this work is to assess the directional characterization of the observed clutter and reverberation. The cardioid arrays should yield good directional estimates of reverberation sources above ~1600–1800 Hz. Examples from the data analysis are presented using a cardioid beamforming algorithm developed by NURC. [Work supported by ONR Code 32, Grant N00014-03-1-0113.]

Session 2aUW

Underwater Acoustics: Reverberation, Scattering and Boundary Interaction

Kevin D. LePage, Chair
Naval Research Laboratory, Code 7144, 4555 Overlook Avenue, SW, Washington, DC 20375

Contributed Papers

8:30

2aUW1. Effect of multipath propagation on the statistics of reverberation in shallow-water environments. Kevin D. LePage (Naval Res. Lab., Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375)

The envelope of direct-path scattered energy has been shown theoretically to deviate from Rayleigh statistics for small ensembles of scatterers with non-Gaussian amplitude distributions [Abraham and Lyons, IEEE J. Ocean. Eng. (2002)]. Here, we investigate the effects of multipath propagation, scatterer correlation length scale, scatterer amplitude distribution, and sonar source and receiver characteristics on the distributional properties of reverberation envelopes in shallow-water environments. For this study we employ a broadband version of the coherent time-domain R-SNAP reverberation model. Model predictions indicate that the non-Rayleighness of shallow-water reverberation caused by distributions of homogeneously distributed diffuse scatterers is time dependent, and sensitive to the number of resolved multipath in the waveguide, the source bandwidth, and the source/receiver beamwidths of the sonar system under consideration. [Work supported by ONR.]

8:45

2aUW2. Unified acoustic model for simulating propagation and scattering effects in oceanic waveguides. Ildar M. Tamendarov and Natalia A. Sidorovskaia (Phys. Dept., Univ. of Louisiana at Lafayette, UL BOX 44210, Lafayette, LA 70504)

The paper addresses the recent developments of Shallow Water Acoustic Modal Propagation (SWAMP) model (http://www.ucs.louisiana.edu/~nxs7560/swamp.html) to account for the scattering events along the acoustic signal propagation path. The theoretical and numerical aspects of SWAMP extension to model the acoustic pulse scattering by a spherical elastic shell in an inhomogeneous oceanic waveguide within the T-matrix approach are discussed. The algorithm utilizes the incident modal functions obtained by SWAMP for an empty waveguide in the spherical representation and the free-field T-matrix. The theoretical foundation for the approach is based on the work by Hackman and Sammelmann [J. Acoust. Soc. Am. **80**, 1447–1458 (1986)]. The T-matrix calculations are implemented in MATLAB software. The interpretation of the scattering event as the modal transformation through the T-matrix is attempted. The numerical studies based on the new model are presented. [Work supported by the Louisiana Board of Regents Support Fund, Contract No. LEQSF(2001-04)-RD-A-38.]

9:00

2aUW3. Unified model for 3-D scattering and forward propagation in a stratified ocean waveguide with random seabed inhomogeneities. Purnima Ratilal and Nicholas C. Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

The field forward propagated through a waveguide containing random volume inhomogeneities in the seabed or random seafloor roughness is modeled using a modal formulation that analytically expresses the effects of dispersion, attenuation and redistribution of modal energy in the forward direction [Ratilal and Makris, J. Acoust. Soc. Am. **114**, 2428 (2003)]. The scatter function density of the random medium is modeled using the Rayleigh–Born approximation. The model is used to determine whether scattering from seabed inhomogeneities can have a noticeable effect on the forward propagated field in the New Jersey Strataform area. Relevant statistical properties of the seabed are obtained from geophysical surveys [Goff, Marine Geol. (2004)] and are further constrained by fitting modeled bottom backscatter with data obtained during the Main Acoustic Clutter Experiment of 2003. The results indicate that scattering from seabed volume inhomogenieties has a negligible effect on the forward propagated field in the 300-Hz to 2-kHz frequency range, but rough surface scattering may be non-negligible.

9:15

2aUW4. Scintillation of short duration records. Steven Lutz, R. Lee Culver, and David Bradley (Appl. Res. Lab, Penn State, P.O. Box 30, State College, PA 16804)

We have investigated propagation of sound in the ocean using the theory of wave propagation in random media. The theory classifies specific source/receiver geometries as either unsaturated, partially saturated, or saturated. In the unsaturated regime the variance of the intensity fluctuations will generally be quite small such that the scintillation index is less than unity. The partially saturated regime can give rise to scintillation indices larger than unity. In the fully saturated regime the scintillation index asymptotically approaches unity. We have conducted an experiment in which the two-dimensional sound-speed field was measured directly using a towed CTD chain, and relatively short acoustic records were captured using a number of source/receiver pairs. The length scales and index of refraction variation place the experiment in the saturated regime. However, measured scintillation indices are much less than 1. We provide an explanation for this difference and utilize local current measurements to bring the theory into agreement with the data. Our results are applicable to operational sonar systems, where short duration records are ordinarily utilized to transmit information or detect targets. [Work sponsored by ONR Code 321US.]

9:30

2aUW5. Analysis of digital side-scan sonar survey data in support of KauaiEx. Jerald W. Caruthers (Univ. of Southern MS, Stennis Space Ctr., MS 39529) and The KauaiEx Group

A side-scan sonar (SSS) survey was conducted off the northwest coast of Kauai, HaI, in support of the high-frequency, channel-characterization experiment (KauaiEx). The SSS used in this survey was a modified system that operates alternately at 150 and 300 kHz and produces digital data as well as standard tiff images of the seafloor. This paper discusses analyses of the digital backscattering data. Previous reports discussed the system characteristics and analyses of standard tiff images produced by the SSS as originally configured. The digital data are high-resolution in space and in dynamic range and offer improved interpretations of the character of the bottom in the KauaiEx range. The data in the range yield interpretations of sand ripples parallel to the depth contours with wavelengths of about 1 m and heights of a few centimeters. These ripples appear to cover the entire KauaiEx range. Data taken at a shallow-water site nearby provide additional information on bottom characteristics that include lava flows, mud, and sand. [Work supported by the Ocean Acoustics Program of ONR.]

9:45

2aUW6. Broadband match-field processing applied in a highly reverberant environment. Natasha A. Chang and David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, 1231 Beal Rd., 2019 Lay Auto Lab., Ann Arbor, MI 48109)

Application of matched-field processing (MFP) techniques to locate and identify broadband acoustic sources is of interest for studying low event rate cavitation and other hydroacoustic noise sources in water tunnels and other reverberant environments. Here, the main challenge lies in fully exploiting the potentially hundreds of kHz of bandwidth of the cavitation signal to refine the source location estimate. This presentation reports on the preliminary application of MFP in a reverberant enclosure that geometrically mimics the test section of a laboratory water tunnel. For this effort a sound projector and one to four hydrophones are used to make narrow-band and broadband sound field measurements within the enclosure. The requisite field model was initially based on a sum of modes assuming pressure release boundaries, but was empirically modified to accommodate finite impedance boundaries. The model modification allowed the enclosure's impulse response to be accurately modeled from 3 to 11 kHz and the incoherent Bartlett processor to locate the source when it emitted a series of continuous single tone signals from 3 to 12 kHz. Extension of this effort to signal pulses, a ray-based field model, and water-tunnel testing and sea trials will be discussed as well. [Work sponsored by ONR.]

10:00

2aUW7. Gaussian beam summation formulation for rough surface scattering in complex configurations. Goren Gordon, Ehud Heyman (School of Elec. Eng., Tel Aviv Univ., Tel Aviv 69978, Israel, heyman@eng.tau.ac.il), and Reuven Mazar (Ben-Gurion Univ. of the Negev, Beer-Sheva 84105, Israel)

A Gaussian beam summation (GBS) representation for wave propagation problems in the presence of rough surfaces is introduced. In this formulation, the coherent and incoherent scattered fields are decomposed into a discrete phase space summation of Gaussian beams (GBs) that emanate from a discrete set of points and directions on the rough surface. The formulation therefore involves stochastic GB-to-GB scattering matrices for the coherent and incoherent fields, and deterministic GB propagators. It benefits from the simplicity and accuracy of the latter, and can be used in applications involving propagation in complex environments and in inverse imaging. The coherent and noncoherent GB2GB scattering matrices are calculated from the statistical moments of the rough surface scattering amplitude, which is given either analytically or empirically. As an example, closed form expressions are calculated within the small perturbation regime. The GB2GB matrices are then used to analyze the propagation along rough-surface waveguides and to estimate the bi-static reverberations. The results clearly demonstrate and explain the phase-space footprints of the stochastic multiple scattering events and of the deterministic propagation between the surfaces.

10:15–10:30 Break

10:30

2aUW8. Acoustic scattering by axisymmetric finite-length bodies. D. Benjamin Reeder (Naval Postgrad. School, Monterey, CA 93943) and Timothy K. Stanton (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

A general scattering formulation is presented for predicting the far-field scattered pressure from irregular, axisymmetric, finite-length bodies for three boundary conditions—soft, rigid, and fluid. The formulation is an extension of a two-dimensional conformal mapping approach [D. T. DiPerna and T. K. Stanton, J. Acoust. Soc. Am. **96**, 3064–3079 (1994)] to scattering by three-dimensional finite-length bodies. This extended formulation, which is inherently numerically efficient to evaluate, involves conformally mapping the surface of an irregular, finite-length body to a new, orthogonal coordinate system in which the separation of variables method may be used to solve the Helmholtz equation and satisfy the boundary conditions. Extensive comparisons with previously published results using other formulations are presented. This formulation is shown to be very accurate in the prediction of scattering from smooth, symmetric bodies for a wide range of frequencies (Rayleigh through geometric scattering region), scattering angles (monostatic and bistatic), aspect ratios, and for each of the three boundary conditions listed above. Reasonable agreement has also been demonstrated for irregular, realistic shapes with soft boundary conditions. [Work supported by ONR.]

10:45

2aUW9. Efficient computation of acoustical scattering from *N* spheres via the fast multipole method accelerated flexible generalized minimal residual method. Nail A. Gumerov and Ramani Duraiswami (Inst. for Adv. Comput. Studies, Univ. of Maryland, A.V. Williams Bldg., College Park, MD 20742, gumerov@umiacs.umd.edu)

Many problems require computation of acoustic fields in systems consisting of a large number of scatterers, which can be modeled as spheres (or enclosed by them). These spheres can have different sizes, can be arbitrarily distributed in three dimensional space, and can have different surface impedance. Solution of this problem via direct T-matrix approach [Gumerov and Duraiswami, J. Acoust. Soc. Am., **112**, 2688–2701 (2002)] is practical only for relatively low number of scatterers, N, since its computational complexity grows as $O(N^3)$. We developed and implemented an efficient computational technique, based on an iterative solver employing a flexible generalized minimal residual method with a right preconditioner. Matrix-vector multiplications involving a large system matrix and the preconditioner are sped up with the aid of the multilevel fast multipole method. We tested the accuracy, convergence and complexity of the method on example problems with $N \sim 10^4$ (millions of unknowns). These tests showed that the method is accurate for a range of frequencies, and experimentally scales as $O(N^{1.25})$. The method has substantial advantages in speed and convergence compared to the reflection method reported earlier [Gumerov and Duraiswami, J. Acoust. Soc. Am., **113**, 2334 (2002)]. [Work supported NSF Awards 0086075 and 0219681, which are gratefully acknowledged.]

11:00

2aUW10. Acoustic intensity measurements involving forward scatter from prolate spheroids. Brian R. Rapids and Gerald C. Lauchle (Grad. Prog. in Acoust. and Appl. Res. Lab., Pennsylvania State Univ., State College, PA 16804)

Underwater acoustic measurements made by scalar pressure sensor are only able to provide an estimate of the magnitude of the total intensity associated with an equivalent plane wave. This equivalence is based upon an assumption that the relative phase between pressure and particle velocity is identically zero. True intensity sensors measure simultaneously the acoustic pressure and components of particle velocity (or related quantity such as acceleration, displacement, or pressure gradient) at a single coordinate in space. The measurement of both pressure and velocity provides the magnitude of acoustic intensity as well as the relative phase between acoustic pressure and velocity. Numerical computations of a steady-state acoustic field perturbed by the presence of a rigid spheroid indicate that the plane wave equivalence may not be a perfect assumption in certain bistatic geometries and that additional information regarding the total acoustic field can be observed only with true intensity sensors. A varnished prolate spheroid constructed from red oak was employed as part of an underwater scattering experiment. A dual axis p-a probe was employed at frequencies near 10 kHz to test the theoretical hypotheses. Theoretical and experimental results will be discussed. [Work supported by ONR, Code 321MS under Grant No. N00014-01-1-0108.]

11:15

2aUW11. Experimental and theoretical target strength of fluid filled spheres. David M. Deveau (PSC 1012, Box 701 FPO, AA 34058)

Investigation into the scattering nature of surfaces or other physical objects often requires the use of measurement systems which cannot always be well controlled. This lack of control can be compensated for by calibrating the resulting measurements against a known target. While these targets can be any object, the goal is to use a target that has a stable scattering response and is independent of angle. The ideal shape is that of a sphere but even this can be improved with the addition of internal fluids that focus and temperature stabilize the scattering response. A scatter response model of a sphere has been developed and used to design four thin walled spheres, each with a different diameter and filled with a focusing fill fluid (Fluorolube). One pair of spheres was measured in an ocean environment while the second pair was tested in a controlled test tank from 5 to 50 kHz but using shorter continuous wave pulses. While the ocean measured spheres closely matched the model, the test tank measurements showed a marked difference from the model. Changes to the model will be explored to determine if theoretical minimums for pulse length are insufficient for targets of this density or focusing capability.

11:30

2aUW12. Interpretation of the ground wave arrival as a head wave sequence. Jee Woong Choi and Peter H. Dahl (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698)

The ground wave arrival (sometimes referred to as a precursor arrival) is interpreted in this work as a train of the head waves, defined here as the head wave sequence. For a Pekeris waveguide, the spectrum of the ground wave arrival is found to be that corresponding to the spectrum of a discrete (time-separated) head wave arrival, evaluated at the modal cutoff frequencies, and the spectrum of a discrete head wave arrival is known to be proportional to the source spectrum divided by frequency. These properties are illustrated for a Pekeris waveguide via Fourier synthesis of a narrow-band complex parabolic equation (PE) field, using the RAM PE code. Although the ground wave arrival in a more realistic waveguide is complicated by several effects, e.g., layering, gradients in the sediment sound speed, some of the above ideas carry over. The introduction of a sound speed gradient in the sediment also produces a key difference in the spectrum of a ground wave arrival (or head wave sequence); this effect is demonstrated through PE simulation, and is also related to data collected in the Yellow Sea. [Research supported by ONR Ocean Acoustics.]

11:45

2aUW13. Frame bulk moduli of air- and water-saturated granular marine sediment models. Masao Kimura (Dept. of Geo-Environ. Technol., Tokai Univ., mkimura@scc.u-tokai.ac.jp)

The frame bulk modulus is important in analyzing the acoustic wave propagation in granular marine sediments. The frame bulk modulus is related to the longitudinal and shear wave velocities. We have reported that the derived values of the frame bulk moduli of air- and water-saturated glass beads are different. In this study, the longitudinal and shear wave velocities in air- and water-saturated beach sands, and in same media in vacuum, are measured. Then the frame bulk moduli are derived from the values of these velocities. The results show that the frame bulk moduli in water-saturated beach sands are about ten times greater than that in air-saturated beach sands, and the frame bulk moduli have dependence on the grain size. These phenomena are investigated using the idea of the gap stiffness between the grains.

TUESDAY AFTERNOON, 16 NOVEMBER 2004 PACIFIC SALON 2, 1:30 TO 4:45 P.M.

Session 2pAA

Architectural Acoustics and Musical Acoustics: Integration of Synthesis Techniques and "Acoustical" Music

K. Anthony Hoover, Chair

Cavanaugh Tocci Associates, Inc., 327F Boston Post Road, Sudbury, Massachusetts 01776

Chair's Introduction—1:30

Invited Papers

1:35

2pAA1. Musical instrument sound levels. K. Anthony Hoover, Andrew Carballeira (Cavanaugh Tocci Assoc. Inc., 327 F Boston Post Rd., Sudbury, MA 01776, thoover@cavtocci.com), Pam Harght, and Sam Ortallono (Berklee College of Music, Boston, MA 02215)

Sound-pressure levels produced by different types of musical instrument can vary considerably. Convincing auralization and appropriate sound-isolation design can be greatly affected by this variation. This paper will present A- and C-weighted statistical sound levels produced by a variety of musical instruments and musical genres, measured at Berklee College of Music. A simple comparison between measured sound isolations using a broadband noise source and a reproduced-music source will also be discussed.

1:55

2pAA2. Conquering space and time—for art's sake. Alex U. Case (Dept. of Music, Univ. of Massachusetts Lowell, 35 Wilder St., Lowell, MA 01854, alex_case@uml.edu)

While sound within a performance space must adhere to the laws of physics governing its propagation, the music synthesized for loudspeaker playback may take some liberties. Recording artists—with the help of the signal processing tools available in the recording studio—often fabricate multiple, physically unrealizable, audio spaces for simultaneous presentation in a single music

performance. Additionally, these musicians are free to slow, accelerate, repeat, and reverse time for musical benefit. This paper isolates and demonstrates some of the unusual space- and time-focused synthesis techniques in popular recorded music, evaluating their technical basis and artistic merit.

2:15

2pAA3. Loudspeaker array as a musical composition genre. David Moulton (Moulton Labs., 39 Ames Rd., Groton, MA 01450)

A fixed array of loudspeakers can be used as a proprietary format for electronic music composition, leading to a fixed genre of work. The author will describe his experiences using a 6-channel array of full-range loudspeakers, describing the compositional principles he has employed, including issues of source localization, ambient and reverberant fields, available bandwidth, and sound-pressure levels. Spatial relationships, musical issues and problems, and related concerns will all be discussed.

2:35

2pAA4. Spectrum-based analysis and synthesis of percussion sounds. James W. Beauchamp (School of Music and Dept. of Elec. and Computer Eng., Univ. of Illinois at Urbana–Champaign, Urbana, IL 61801)

From the standpoint of the listener, convenient ways to catagorize percussion sounds are in terms of mode frequency positions, mode frequency spacing, and decay rates. Some sounds, like those of bells, chimes, and other struck bars, have modes which are widely spaced. Other sounds, such as those of timpani, tam-tam, and cymbals, have many more closely spaced modes. Measurements of average modal spacing, time-varying spectral shift, spectral incoherence, attack time, and decay rate are presented for a variety of percussion instrument sounds. Examples of percussion synthesis using this information will be presented.

2:55–3:10 Break

3:10

2pAA5. The influence of synthesized music on the expansion of trombone performance techniques. Thomas Plsek (Berklee College of Music, 1140 Boyslton St., Boston, MA 02215)

Much of the development of synthesized music has been involved with the modeling of traditional acoustic sources such as strings, brass, woodwinds, and percussion. It has, of course, also been concerned with the creation of new, yet unheard, sounds and the modifications and manipulations of existing sources through signal processing. The author has been involved with these processes for many years as a trombone explorer, and has had to interact and integrate wth them. Much was discovered and learned about how trombone performance can be expanded to acoustically simulate these developments in electronic music. This paper demonstrates some of the techniques such as the real-time control of sound source location and the manipulation of timbre. In many instances the acoustical solution proved to be as, or even more, interesting as the synthetic process.

3:30

2pAA6. Measurements of room effects on digital simulations of a concert hall sound field. Patrick R. McAtee and Lily M. Wang (Architectural Eng. Program, Univ. of Nebraska - Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, pmcatee@unlnotes.unl.edu)

Digital signal processing is becoming more and more commonplace in the field of architectural acoustics. One way it has been utilized is to simulate the acoustics of a space within another one through convolution of that space's digitized acoustic characteristics with a source signal. In the consumer market, Yamaha pioneered and leads this area, producing a full range of amplifiers with several multichannel auralizations built in. It is not possible to ensure that all users will tune their listening space to allow the auralization to perform its best. This raises questions about how well the auralization can perform in various acoustic environments, and what indicators there might be as to the quality of that performance. A Yamaha system and two sets of speakers were taken to more than a dozen spaces, including offices, classrooms, band rooms, and larger performance spaces. Measurements of each room's acoustical characteristics were recorded and analyzed to see how the acoustic parameters changed when the digital simulation of a concert hall sound field was running. Comparisons of the frequency responses, reverberation times, and clarity indices will be presented. [Work supported by UNL College of Engineering Summer Undergraduate Engineering Research Experience Grant.]

3:50

2pAA7. Design challenges for the integration of virtual acoustics in music practice rooms. Ron Freiheit (Wenger Corp., 555 Park Dr., Owatonna, MN 55050, ron.freiheit@wengercorp.com)

The use of virtual acoustics has provided a new level of practice experience for the musician. By integrating the sound isolation of music practice rooms with the signal processing of an active acoustic system (with time-variant gain before feedback), musicians can now benefit from the experience of practicing in multiple acoustic environments. The variability of the acoustic environment allows the musician to clearly hear their intonation and articulation, which may be difficult to discern in a small practice room. For the effective implementation of virtual acoustics in small spaces (typically 15 square meters) a number of technical issues had to be resolved including sound-field immersion, reducing input level due to proximity of the source to the microphone transducers, and a simplified system control allowing musicians to focus on the art of practice and not the technology. This paper deals with these issues and the subsequent design decisions to address them.

4:10

2pAA8. Composition and performance as outgrowth of synthesis techniques. Richard Boulanger and Greg Thompson (Berklee College of Music, 150 Massachusetts Ave., Boston, MA 02215)

New methods of controlling and interacting with synthesizers are increasingly available to the composer. Advances in input devices enable a large range of gestural parameters to be translated into musically-relevant data, which can then be used in synthesis environments like Csound and Max/MSP. This paper will examine such input devices, including the radio baton, and show how this technology may be combined with existing synthesis techniques to produce exciting new sounds, algorithms, and compositional works.

Contributed Paper

4:30

2pAA9. Violin acoustic radiation synthesis: A source model for direct sound enhancement in musical acoustic environments. Jacob Waxman and Mark Bocko (Dept. of Elec. and Comput. Eng., Univ. of Rochester, Rochester, NY 14627, jw001j@mail.rochester.edu)

Within the context of immersive acoustic environments (real or virtual) for purposes of musical performance, it is useful to recreate the acoustic field radiated by real musical instruments. Given the popularity of wave field synthesis for direct sound enhancement and other methods of holophonic sound imaging, source modeling is a desirable development for such musical applications. There exist careful directivity measurements of the sound radiation of the violin in the frequency range from 1 to 5 kHz, over which the far field directivity changes rapidly as a function of frequency [L. M. Wang, ''Radiation Mechanisms from Bowed Violins,'' Ph.D. thesis, Pennsylvania State University, 1999]. Source models based on these measurements using a cylindrical harmonic decomposition are presented. With this preliminary approach to violin acoustic radiation modeling, issues regarding further auralization of this synthesized field are subsequently addressed.

TUESDAY AFTERNOON, 16 NOVEMBER 2004 PACIFIC SALON 1, 1:25 TO 5:10 P.M.

Session 2pAB

Animal Bioacoustics: Marine Mammal Acoustics: Session in Honor of Ron Schusterman II

Colleen Reichmuth Kastak, Chair
Long Marine Laboratory, University of California, Santa Cruz, 100 Shaffer Road, Santa Cruz, California 95060

Chair's Introduction—1:25

Invited Papers

1:30

2pAB1. Temporal integration in a California sea lion and a harbor seal: Estimates of aerial auditory sensitivity as a function of signal duration. Marla M. Holt (UC Santa Cruz Long Marine Lab., 100 Shaffer Rd., Santa Cruz, CA 95060), Brandon L. Southall (NOAA Fisheries Acoust. Program, Silver Springs, MD 20910), David Kastak, Ronald J. Schusterman, and Colleen Reichmuth Kastak (UC Santa Cruz Long Marine Lab., Santa Cruz, CA 95060)

Stimulus durations shorter than some critical value result in elevated signal-detection thresholds due to temporal integration (or temporal summation) properties of the auditory system. These properties are important from a theoretical perspective in terms of the trade-offs of stimulus duration and intensity on sensitivity. From a methodological perspective, temporal integration is important because absolute detection thresholds measured using signal durations shorter than the temporal integration period may underestimate hearing sensitivity. In this study, aerial sound-detection thresholds were estimated at 2500 and 3530 Hz in a California sea lion (*Zalophus californianus*) and a harbor seal (*Phoca vitulina*). Thresholds were measured for each frequency at seven stimulus durations ranging from 100 to 500 ms using behavioral psychophysics in a hemianechoic chamber. In general, thresholds increased as tone duration decreased for durations shorter than approximately 300 ms. For tone durations longer than 300 ms, thresholds were not different from those measured with the longest duration tested. These results suggest temporal integration times of approximately 300 ms for these species, which are consistent with data collected on other mammals. Our findings suggest that tone durations longer than 300 ms should be used in estimating pinniped auditory sensitivity.

1:50

2pAB2. Noise-induced temporary threshold shifts in pinnipeds: Effects of noise energy. David Kastak (UCSC Long Marine Lab., 100 Shaffer Rd., Santa Cruz, CA 95060), Brandon Southall (UCSC Long Marine Lab and NOAA Fisheries Acoust. Program, Silver Spring, MD 20910), Marla Holt, Colleen Reichmuth Kastak, and Ronald Schusterman (UCSC Long Marine Lab., Santa Cruz, CA 95060)

Auditory pure-tone thresholds were obtained in air and in water from three pinnipeds before and immediately after exposure to octave-band noise. Noise exposure durations were 1.5, 12, 22, 25, or 50 min, and noise levels were 65, 80, or 95 dB referenced to each subjects pure-tone threshold. In air and in water, pre- and postnoise thresholds were obtained at the center frequency of the octave band. In water, thresholds were also obtained at a frequency octave higher than the octave-band center frequencies. Maximum

threshold shifts for each species were about 15 dB in air and in water. Under all exposure conditions hearing sensitivity recovered within 24 h. In both media, a curvilinear function best predicted the magnitude of threshold shift from noise energy flux density. At TTS magnitudes greater than 10 dB, this function predicts a 2-dB increase in threshold shift for a 1-dB increase in noise energy, agreeing well with data collected from other mammals.

2:10

2pAB3. Objective measures of steady-state auditory evoked potentials in cetaceans. James J. Finneran (US Navy Marine Mammal Program, SPAWARSYSCEN San Diego, Code 2351, 53560 Hull St., San Diego, CA 92152) and Dorian S. Houser (BIOMIMETICA)

Although behavioral methods provide the most direct means of assessing hearing capability, the time, access, and cost required to train individual marine mammal subjects has limited large-scale application of this technique. As an alternative to behavioral testing, auditory evoked potentials (AEPs) have been measured in a number of marine mammal species. AEP measurements using steady-state amplitude modulated tones allow rapid estimates of hearing threshold without extensive subject training. These stimuli result in sinusoidal AEPs, allowing frequency domain measures of AEP amplitude. Unfortunately, AEPs from near-threshold stimuli possess very low signal-to-noise ratios, making discrimination of AEPs from noise difficult. The most common procedures used for AEP threshold estimates in marine mammals have featured a subjective component, requiring an experienced observer to assess the presence or absence of the AEP. In this talk, the use of objective techniques to determine the presence of an AEP will be discussed. The focus of the talk will be on parametric and nonparametric methods within the frequency domain and the importance of AEP phase information. Steady-state AEPs measured in bottlenose dolphins will be used to evaluate the different techniques and compare to behavioral response measures. [Work supported by ONR and SSC San Diego ILIR.]

2:30

2pAB4. Hearing thresholds of a stranded infant Rissos dolphin. Paul E. Nachtigall, Michelle M. Yuen, T. Aran Mooney, and Kristen A. Taylor (Marine Mammal Res. Program, Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734, nachtiga@hawaii.edu)

The underwater hearing of an infant male Rissos dolphin that stranded off the coast of southern Portugal was measured using evoked auditory potentials (AEPs). Hearing thresholds were measured from envelope following responses to amplitude modulated pure tones ranging from 4 to 150 kHz. Acoustic signals were presented within the calibrated rehabilitation pool. Evoked responses were passively gathered from human EEG electrode sensors imbedded within rubber suction cups that were gently attached to the animal with slight suction and conductor gel. One sensor was placed behind the blowhole and the other reference sensor on the back. Tones were presented 1 m directly in front of the animal under water while the animal was held in a neutrally buoyant position. Unlike the previously published audiogram for an older Rissos dolphin, the audiogram, obtained by presenting 18 different frequencies, showed that this very young animal heard tones up to 150 kHz in a manner similar to other odontocetes. These data, collected within 4 days, are the first measure of the hearing of a neonate marine mammal, verify that very young dolphins likely hear better than older animals, and show the value of using AEPs to obtain hearing thresholds of stranded animals undergoing rehabilitation.

2:50

2pAB5. *In vivo* imaging correlated with otoacoustic emissions as a metric for ear disease in seals. D. R. Ketten (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543; Harvard Med. School, Boston, MA 02114, dketten@whoi.edu), W. F. Dolphin (Boston Univ., Boston, MA 02115), R. William (Woods Hole Aquarium, Boston, MA 02110), J. Arruda, and J. O'Malley (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Otoacoustic emissions (OAEs) coupled with auditory brainstem responses (ABR) can help differentiate central, sensorineural, and conductive hearing losses. Pinnipeds with moderate bore canals are good OAE candidates, but OAE utility for assessing marine mammal inner ear health is not known. We examined three juvenile harbour seals (*Phoca vitulina*) with ear disease with computerized tomography (CT), OAE, and ABR. Combining CT, OAE, and ABR allowed simultaneous ear pathology documentation, quantification, and intracanal probe microphone position determination. Hearing was tested bilaterally from 500 Hz to 15 kHz. CT/OAE/ABR results were assessed independently. In two animals, CT showed middle ears occluded with fluid but normal auditory nerve anatomy, suggesting short-term circumscribed infection with no retrograde neuronal loss. OAE found moderately elevated response levels consistent with conductive hearing loss. ABR confirmed normal brainstem functioning. In the third animal, no OAEs or ABRs were obtainable up to 70 dB re 1 μPa, suggesting retrograde loss through brainstem level. CT for this animal showed inner, middle, and external ear occlusions consistent with aggressive, long-term disease. These data show volume and site of auditory pathologies are strongly correlated with OAE results in pinnipeds. [Work supported by Seaver Institute and ONR N00014-93-1-0940; N00014-94-1-1081.]

3:10–3:25 Break

3:25

2pAB6. Noise exposure metrics for auditory and nonauditory damage in aquatic animals. Mardi C. Hastings (Office of Naval Res., Code 341, 800 N. Quincy St., Arlington, VA 22217, mardi_hastings@onr.navy.mil)

The total acoustic energy flux is often used to correlate impacts of different types of sounds having various durations on marine animals. To calculate this metric, one must know both the acoustic pressure and particle velocity. But, in practice, the acoustic pressure is measured usually at just one point and the particle velocity is unknown. The total energy flux is then estimated by assuming the stimulus is an ideal plane wave and calculating the sound exposure level. The inner ears of many aquatic animals, however, respond directly to acoustic particle motion. In addition, because aquatic animals are acoustically coupled to the surrounding water, acoustic particle motion is also critical in estimating sound exposures that may cause damage to tissues outside the auditory system. Thus in noise impact studies, both pressure and local particle velocity must be measured to accurately estimate the total acoustic energy dose received by an aquatic animal. Several case studies taken from the literature will be presented to demonstrate how to estimate acoustic particle velocity from pressure gradient measurements in common aquatic animal testing environments and calculate the total acoustic energy flux.

3:40

2pAB7. Ocean bioacoustics, human generated noise and ocean policy. Michael Stocker (Seaflow, Inc., 1062 Fort Cronchite, Sausalito, CA 94965)

The recent release of the U.S. Commission on Ocean Policy (USCOP) report, just a year on the heels of the Pew Oceans Commission report, has alerted policymakers and the public about the precarious biological health of our seas. While the reports discuss ecosystem based management, ocean bioacoustics is given short treatment in both reports. The ocean is not a visual-dominant environment, rather it is an acoustic environment. Most animals in the sea depend on sound, but we know next to nothing about how living organisms use it. We do know from recent studies that ocean habitats are being seriously compromised by human generated noise in evidence through stranded whales and, more recently, high fish mortality and low productivity in fishing areas due to seismic exploration and civil engineering. Due to the ubiquity of sounds and noises in all of our ocean enterprises, legislating anthropogenic sound promises to be a Byzantine endeavor. This paper examines some of the known challenges to crafting ocean noise policy.

3:55

2pAB8. A potential explanation for marine mammal strandings. L. Crum, S. Kargl, and T. Matula (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

The stranding of whales following naval exercises has been a topic of increasing interest. A recent article [P. D. Jepson *et al.*, ''Gas-bubble lesions in stranded cetaceans,'' Nature (London) **425**, 575–576 (2003)] has suggested that some version of decompression sickness may be involved. Considerable additional evidence has been presented that indicates both acute and chronic effects of these bubbles. We suggested some time ago [L. A. Crum and Y. Mao, ''Acoustically enhanced bubble growth at low frequencies and its implications for human diver and marine mammal safety,'' J. Acoust. Soc. Am. **99**, 2898–2907 (1996)] that the acoustic fields produced by naval sonars could induce bubble formation under certain conditions, particularly if there were a high level of nitrogen supersaturation in these animals. Recent evidence of high supersaturation levels [D. S. Houser, R. Howard, and S. Ridgway, ''Can diving-induced tissue nitrogen supersaturation increase the chance of acoustically driven bubble growth in marine mammals?,'' J. Theor. Biol. **213**, 183–195 (2001)] has stimulated us to perform a series of experiments that suggests that even modest acoustic intensities can trigger bubble formation under supersaturated conditions. Once bubbles are nucleated, these local supersaturation levels then result in bubble growth to macroscopic sizes, and their potential deleterious bioeffects. The results of these preliminary experiments will be presented. [Work supported by APL IR&D funds.]

4:10

2pAB9. High frequency components in cetacean echlocation signals. Thomas G. Muir, Ronald W. Toland, Steven R. Baker (U.S. Naval Postgrad. School, Monterey, CA 93943), Diane J. Blackwood, Lew A. Thompson, and Preston S. Wilson (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713)

The rich literature of high-resolution biosonar capability is sometimes baffling as to attainable sonar resolution. This has led us to measurements in San Diego Bay, on captive research dolphins, utilizing an extremely wide band piezo-composite hydrophone, with a frequency response extending to 2 MHz. The results indicate that cetacean echolocation signals contain frequency components, above ambient noise, that can extend to the neighborhood of 500 kHz. The study was conducted on two bottlenosed dolphins and one beluga whale. Attempts were made to determine if these animals were actually using these high frequency components in echolcation, but they were not completely successful. However, measurements with rho-c matched, acoustic, low pass filter panels, placed between the blindfolded subjects (on a bite bar) and their test targets, showed that as the detection tasks became more difficult, the animals increased the intensity of the high frequency content of their transmissions (to ~500 kHz) and their pulse repetition frequencies also increased. [Work supported by the U.S. Navy ONR.]

4:25

2pAB10. The modulation rate transfer function of a Rissos dolphin, *Grampus griseus*, using the envelope following response.. T. Aran Mooney, Paul E. Nachtigall, and Michelle M. Yuen (Marine Mammal Res. Program, Hawaii Inst. of Marine Biol., 46-007 Lilipuna Rd., Kaneohe, HI 96744, mooneyt@hawaii.edu)

Auditory evoked potentials (AEPs) were used to obtain a modulation rate transfer function (MRTF) of a young, stranded Rissos dolphin (*Grampus griseus*). The animal's neurological response to sound was stimulated by triangle-shaped broadband clicks, $\frac{1}{2}$ ms in duration, played at presentation rates from 100 to 2000 Hz. These stimuli were played from a transducer 1 m in front of the subject, within a calibrated rehabilitation pool. The AEP responses were recorded using human EEG sensors attached by two soft suction-cups, placed just behind the blowhole and in front of the dorsal fin, respectively. By playing the stimuli at different rates, from low to high frequency, it was possible to determine the maximum and most efficient modulation rate, to be used in an AEP audiogram. The MRTF was similar to other cetaceans as it is low-pass in shape. Corner frequencies were somewhat lower than other published marine mammal MRTFs, dropping steeply after 1000 Hz. At this frequency, the evoked potential was highest, 230 nV, compared to noise levels in the tens of nV. Thus, 1000 Hz was determined to be the modulation rate used in determining the animals AEP audiogram.

4:40

2pAB11. Directional sensitivity in bottlenose dolphins: Evoked potential study. Vladimir Popov and Alexander Supin (Inst. of Ecology and Evolution, 33 Leninsky Prosp., 119071 Moscow, Russia)

ABR threshold-versus-azimuth function was measured in two bottlenose dolphins. The measurements were done in a circular pool 6 m in diameter and 40 cm deep. The animal was supported by a stretcher, the head in the center of the pool. The transducer was moved around the animal's head at the radius of 1.2 m. ABRs were recorded using electrodes mounted in suction caps fastened at the head surface. ABR thresholds were measured at azimuth of the sound source of $\pm 90°$ from the midline and frequencies from 8 to 128 kHz. Both acuteness of the directional diagram and the lowest-threshold azimuth value depended on stimulus frequency. Acuteness increased with the frequency increase. The lowest-threshold azimuth was near midline at high frequencies and deviated laterally up to 45–60° with lowering the frequency. The maximum azimuth-dependent thresholds shift (20–35 dB) was at high frequencies (90–128 kHz); at lower frequencies (8–16 kHz), the azimuth-dependent threshold shift decreased down to 10–15 dB. [Work supported by the Russian Foundation for Basic Research.]

4:55

2pAB12. Comparison of behavioral and auditory evoked potential (AEP) audiograms of a false killer whale (*Pseudorca crassidens*). Michelle Yuen, Paul E. Nachtigall, Marlee Breese (Hawaii Inst. of Marine Biol., Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734, myuen@hawaii.edu), and Alexander Ya. Supin (Severtsov Inst. of Ecol. and Evolution, 119071 Moscow, Russia)

Behavioral and auditory evoked potential (AEP) audiograms of a false killer whale were measured using the same subject and experimental conditions from 2001 to 2004. The objective was to compare and assess the validity of auditory thresholds collected by psychometric and electrophysiological techniques. Behavioral audiograms used 3-s pure-tone stimuli from 4 to 45 kHz. AEP audiograms used 20-ms sinusoidally amplitude-modulated (SAM) tone bursts from 4 to 45 kHz. Electrophysiological responses were received through gold disk electrodes mounted in rubber suction cups placed on the animals dorsal surface. Psychometric data were reliable and repeatable with the region of best sensitivity for the behavioral audiograms between 16 and 24 kHz, and with peak sensitivity at 20 kHz. The AEP measures produced thresholds that were consistent over time, with ranges of best sensitivity from 16 to 22.5 kHz and peak sensitivity at 22.5 kHz. Behavioral thresholds were lower than AEP thresholds. Signal type and duration differences may explain the discrepancies between thresholds obtained under the two conditions. These data indicated that behavioral and AEP techniques can be used successfully and interchangeably to measure cetacean hearing sensitivity. [Work supported by ONR.]

TUESDAY AFTERNOON, 16 NOVEMBER 2004 SHEFFIELD ROOM, 1:30 TO 3:30 P.M.

Session 2pAO

Acoustical Oceanography and Underwater Acoustics: Acoustic Sensing of Internal Waves II

James F. Lynch, Chair

Woods Hole Oceanographic Institute, 203 Bigelow Building, Woods Hole, Massachusetts 02543

Contributed Papers

1:30

2pAO1. Fluctuations of high-frequency sound field in shallow water in the presence of internal waves. Mohsen Badiey (Univ. of Delaware, Newark, DE 19716), Boris Katsnelson, and Serguey Pereselkov (Voronezh Univ., Voronezh, 394006 Russia)

The structure of space-time fluctuations of a high-frequency (several kHz) sound field in a shallow water waveguide is considered. The sound field fluctuations due to internal and surface waves show different temporal variability. The perturbation theory within the framework of ray description is applied in order to derive some analytical estimation of fluctuations of arrival time and vertical angle distribution of high frequency broadband acoustics signals. The parabolic equation is used for numerical modeling of space-time fluctuations and vertical sound field intensity distribution. The experimental measurements of ocean dynamics from recent acoustic experiments are used as input data in numerical simulation of sound field fluctuations. [Work was supported by ONR and CRDF].

1:45

2pAO2. Internal- and surface-wave-induced fluctuations and frequency spreading in shallow water acoustic propagation over short ranges. Stephen D. Lynch, Gerald L. D'Spain, and Eric Terrill (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093)

During a 2001 ocean acoustics experiment approximately 40 miles west of San Diego, a moored, underwater source at 30-m depth transmitted a set of narrow-band tones in the 100–1000-Hz band to a 32-element, 992-m-long bottom hydrophone array deployed from R/P FLIP. Also deployed about FLIP were multiple, fixed vertical thermistor strings with ambient pressure sensors, and CTD casts. These data provide observations of the effects of a fluctuating ocean environment on acoustics in shallow water (180 m) over short ranges (2.5 km). A series of very long fast Fourier transforms of the received signals reveal time dependence in the frequency spreading about the transmitted frequency. Time series of spatial and temporal coherence estimates show temporal variations in the relatedness of the received signals. The environmental data are used along with physics-based models of ocean surface and internal waves to predict the signal fluctuation and frequency spreading characteristics. Results of

comparison of acoustic observations with oceanographic-data/model-based predictions under various assumptions will be presented. [Work supported by ONR, Code 321(US).]

2:00

2pAO3. Vertical line array beamforming of signal and noise in shallow-water regions. Tim Duda, James Lynch (AOPE Dept., M.S. 11, Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Phil Abbot (Ocean Acoust. Services and Instrumentation Systems, Inc., Lexington, MA 02421), and Ruey-Chang Wei (Natl. Sun Yat-sen Univ., Kaohsiung City, Taiwan)

The familiar mode-stripping and coupling processes are prevalent in shallow water, shaping the vertical directional content of noise from distributed sources and of signals from discrete sources. Directional patterns can help in the determination of the physical processes, including internal waves, controlling acoustic behavior in a given region. VLA beamformer output from the New England PRIMER and the South China Sea ASIAEX studies are shown. Beamformed output from moored source signals and of noise vary rapidly for ASIAEX, less so for PRIMER. Fluctuations of signal and noise directional content have limited coherence. Correlated fluctuations of these will not alter the signal/noise ratio for a given beam, whereas uncorrelated fluctuations will. Such correlation depends upon the distance between the mode-coupling structure and the receiver. Predictions of pattern variability are complex because many environmental degrees of freedom are involved. For example, mode coupling of distant sources will spread the pattern from near horizontal (low mode), but not always for nearby sources; mode coupling of noise will energize low angle beams, although most effectively near noise sources. Implications of waveguide characterization based on beam energy variability will be discussed and compared with direct environmental waveguide measurements and with other signal properties. [Work supported by ONR.]

2:15

2pAO4. Acoustic mode coupling effects from propagation through nonlinear internal waves. Laurel K. Reilly-Raska (Rennselaer Polytech. Inst., Troy, NY 12180, reilll@rpi.edu), James F. Lynch, John A. Colosi (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), and William L. Siegmann (Rennselaer Polytech. Inst., Troy, NY 12180)

Existing literature demonstrates that costal and shelf break regions frequently contain significant nonlinear internal wave (NIW) activity and fronts. Our previous work, which incorporated stochastic modeling techniques in conjunction with azimuthally varying NIWs in an adiabatic setting, showed significant variations in intensity. However, NIWs can strongly affect energy coupling between acoustic modes. We seek to understand the evolution of coupled mode propagation through a field of randomly varying NIWs. We will examine the effects of coupling due to propagation through the intersection of two or more NIWs crossing at oblique angles. We use a fully 3D parabolic equation method and other tools to examine the ensemble of effects. [Work supported by ONR.]

2:30

2pAO5. The relative effects of sound-speed variability in the water column versus the seabed on normal-mode propagation in shallow water. George V. Frisk (Dept. of Ocean Eng., Florida Atlantic Univ., Dania Beach, FL 33004, gfrisk@seatech.fau.edu), Cynthia J. Sellers (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), and Luiz L. Souza (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Normal-mode propagation in shallow water is influenced by variations in the acoustic properties of both the water column and the seabed. Specifically, the sound speed in the water may fluctuate due to the passage of an internal wave, while the sound speed in the bottom may change due to variable geological features. Assessing the relative effects of the water column versus the seabed on the characteristics of modal propagation is critical to the understanding of both the forward and inverse problems in shallow-water acoustics. In this paper, perturbation theory is combined with a Pekeris waveguide model to provide an analytic formalism for evaluating the delicate interplay between the water column and the seabed in shallow water propagation. In particular, the relative contributions of sound-speed fluctuations in the water and the bottom to variations in the modal eigenvalues are determined. The results are affected by both the strengths of the fluctuations and the magnitudes of the background modal eigenfunctions in the water and the seabed. The complexity of the problem is illustrated with synthetic and experimental data. [Work supported by ONR.]

2:45

2pAO6. Coupled mode theory for sound propagation through random internal wave fields. John Colosi (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

In the late 1970's Dozier and Tappert were able to formulate a theory for the range evolution of deep-water acoustic normal-mode energies utilizing (1) a perturbation method up to second-order mode transitions; (2) a phase-randomized internal wave field; (3) the Markov approximation; and (4) zero cross-mode coherence. Since 1989, vertical receiver array observations of broadband signals at 75 and 250 Hz have shown distinct time-resolved wavefronts at multimegameter ranges, thereby invalidating the zero cross-mode coherence assumption. In this work, the theory is modified to remove the zero cross-mode coherence assumption, thereby yielding range evolution equations for both the mode energies and the cross-mode coherences. Using a Lorentzian internal wave spectrum which fits the Garrett–Munk spectrum very well, analytic expressions for the coefficients of the evolution equations are obtained. Predictions using the theory agree with Monte Carlo numerical simulations at 75 Hz at roughly the 1–2-dB level. It is found that cross-mode coherences decay very slowly with range, and that the modal energy equations nearly decouple from the cross-mode coherence equations. The theory may offer a means to acoustically measure internal wave spectra.

3:00

2pAO7. Random horizontal refraction at long-range sound propagation in the ocean. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA/Environ. Technol. Lab., 325 Broadway, Boulder, CO 80305), Alexander G. Voronovich, and Valery U. Zavorotny (NOAA/Environ. Technol. Lab., Boulder, CO 80305)

Internal gravity waves (IWs) are a major source of sound speed fluctuations in the ocean. For refracted waves in deep water, 3-D and 4-D acoustic effects are caused by cross-range gradients and time dependence of the sound speed. In this paper, IW-induced 3-D and 4-D acoustic effects are studied assuming a continuum of random IWs with the Garrett–Munk spectrum. Fluctuations in azimuthal arrival angles and frequency of tonal signals as well as travel time biases due to horizontal refraction are quantified with a theory which allows one to calculate directly statistical moments of acoustic observables as a weighted integral of an appropriate statistical moment of environmental perturbations. Predictions of the theory are compared to results of Monte Carlo simulations of 3-D and 4-D acoustic effects in the ray approximation. Magnitude of the effects associated with random horizontal refraction at megameter ranges is found to be sensitive to a choice between forms of the Garrett–Munk spectrum which are often viewed as equivalent. Implications of this observation on acoustic characterization of IW fields in deep water are discussed. [Work supported by ONR.]

3:15

2pAO8. Space-frequency sound field distribution in the neighborhood of a temperature front. Boris Katsnelson, Alexander Tshoidze (Voronezh Univ., 1, Universitetskaya sq., Voronezh, 394006, Russia), James Lynch, and Ta-Wei Wang (Woods Hole Oceanogr. Inst, Woods Hole, MA 02543)

In this paper, the properties of the sound field in the area close to a temperature front (TF) in shallow water are studied. It is shown that for a realistic temperature front providing the variation of the sound speed profile, interesting effects due to horizontal refraction can be observed if the

source/receiver are placed at distances of several hundred meters (up to 1 km) from the TF. It is shown that strong interferences in the sound field take place both in the range (horizontal plane) and frequency domains (resonance-like spectra of received signals). Effects of mode separation due to horizontal refraction can be experimentally observed (in principle) as well as the aforementioned intensity gain. These effects are similar to observed horizontal refraction phenomena in the presence in internal solitons. Numerical modeling is carried out, and an experimental setup is discussed for typical conditions in shallow water areas. [Work supported by RFBR (03-05-64568) and CRDF (REC 010-0)].

TUESDAY AFTERNOON, 16 NOVEMBER 2004 ROYAL PALM SALON 6, 1:25 TO 4:15 P.M.

Session 2pEA

Engineering Acoustics, Signal Processing in Acoustics, Psychological and Physiological Acoustics and Committee on Standards: Hearing Aids

Gary W. Elko, Cochair
Avaya Labs, 233 Mt. Airy Road, Basking Ridge, New Jersey 07920

James G. Ryan, Cochair
Gennum Corporation, 232 Herzberg Road, Kanata, Ontario K2K 2A1, Canada

Chair's Introduction—1:25

Invited Papers

1:30

2pEA1. Integrated circuit technology in hearing aids. Steve Armstrong (Gennum Corp., 970 Fraser Dr., Burlington, ON L7L 5P5, Canada)

In recent years, digital technology has enabled the provision of many new hearing aid features such as noise reduction, feedback suppression and environment classification. As silicon technology evolves, more and more features will be possible within the existing package size and power budget. This paper provides an overview of integrated circuit technology and its application to hearing aid design. The architecture of a digital hearing aid circuit is described in light of the system constraints. The talk will conclude with a brief discussion of expected technological advances and their impact on hearing aid design.

2:00

2pEA2. Noise reduction for digital hearing aids. Volker Hohmann and Birger Kollmeier (Universitaet Oldenburg, Medizinische Physik, D-26111 Oldenburg, Germany)

The normal-hearing system extracts monaural and binaural features from the signals at the left and right ears in order to separate and classify sound sources. Robustness of source extraction is achieved by exploiting redundancies in the source signals (auditory scene analysis). ASA is closely related to the ''Cocktail Party Effect,'' i.e., the ability of normal-hearing listeners to perceive speech in adverse conditions at low signal-to-noise ratios. Hearing-impaired people show a reduced ability to understand speech in noisy environments, stressing the necessity to incorporate noise reduction schemes into hearing aids. Several algorithms for monaural, binaural and multichannel noise reduction have been proposed, which aim at increasing speech intelligibility in adverse conditions. A summary of recent algorithms including directional microphones, beamformers, monaural noise reduction and perceptual model-based binaural schemes will be given. In practice, these schemes were shown to be much less efficient than the normal-hearing system in acoustically complex environments characterized by diffuse noise and reverberation. One reason might be that redundancies in the source signals exploited by the hearing system are not used so far by noise reduction algorithms. Novel multidimensional statistical filtering algorithms are introduced that might fill this gap in the future. [Work supported by BMBF 01EZ0212.]

2:30

2pEA3. A metric for evaluating speech intelligibility and quality in hearing aids. James M. Kates (GN ReSound, 3215 Marine St., Rm. W161, Boulder, CO 80309, jkates@gnresound.dk) and Kathryn H. Arehart (Univ. of Colorado, Boulder, CO 80309)

Noise and distortion reduce speech intelligibility and quality in hearing aids, but there are no metrics that encompass both noise and distortion. This presentation introduces new intelligibility and sound-quality calculation procedures based on the Speech Intelligibility Index [ANSI S3.5-1997]. The SII involves measuring the signal-to-noise ratio (SNR) in separate frequency bands, modifying the estimated noise levels to include auditory masking, and computing a weighted sum across frequency of the modified SNR values. In the new procedure, the estimated SNR is replaced by a signal-to-distortion ratio (SDR) computed from the coherence between the input and output signals of the system under test. The SDR replaces the system noise estimate with the combined noise and nonlinear

distortion. For the procedure, the signal is divided into three regions comprising the low-, mid-, and high-level signal segments. The SII is then calculated for each region using the corresponding coherence. The three coherence SII values are then combined to predict the intelligibility and the sound quality for the device under test. The coherence SII procedure is shown to be accurate for both normal-hearing and hearing-impaired listeners for additive noise, peak-clipping distortion, and center-clipping distortion. [Work supported by GN ReSound (JMK) and the Whitaker Foundation (KHA).]

3:00–3:15 Break

3:15

2pEA4. Hearing aids: real world outcomes of the engineering feats. Ruth Bentler (Dept. of Speech Pathol. and Audiol., Univ. of Iowa, 250 Hawkins Dr., Iowa City, IA 52242, ruth-bentler@uiowa.edu)

It is often the case that software and hardware designs result in profound changes in the way hearing aids function. However, designing engineers must rely on feedback from others' disciplines as to how these devices are evaluated and used in the "real world." Directional microphone systems, noise reduction algorithms, and feedback cancellation schemes are all designed to eliminate unwanted interference. How those functions work in the real world, either independently or concurrently, will be explored.

Contributed Papers

3:45

2pEA5. Audio scene analysis for hearing aids. Marie Roch, Tong Huang (Dept. of Computer Sci., San Diego State Univ., 5500 Campanile Dr., San Diego, CA 92182-7720), and Richard R. Hurtig (The Univ. of Iowa, Iowa City, IA 52242)

It is well known that simple amplification cannot help many hearing-impaired listeners, and numerous signal enhancement algorithms have been proposed for digital hearing aids. In many cases, algorithms are most effective in specific environmental or source conditions. If one can properly detect components of the auditory scene, it is possible to dynamically apply enhancement algorithms which are appropriate for a given situation. This work illustrates this principle by describing a cohort detection scheme which serves as a control system for a frequency-domain compression algorithm. The compression algorithm preserves formant ratios and thus enhances speech understanding for individuals with severe sensorineural hearing loss in the 2–3-kHz range. By detecting speakers from broad cohorts (e.g., male, female), it is possible to adjust the compression ratio dynamically based upon characteristics of the auditory scene, resulting in a more appropriate enhancement than one based upon a static ratio. Cohort decisions are derived from the likelihood scores of a Gaussian mixture model classifier.

4:00

2pEA6. A new psychoacoustic fitting algorithm for digital hearing aids. Jeong-Guon Ih, Kyung-Hoon Park (Dept. of Mech. Eng., KAIST, Sci. Town, Taejon 305-701, Korea), and Dong-Gu Yoo (Sammi Sound Tech., Cheongju, Chungbuk 360-813, Korea)

Existing fitting methods, which employ a pure tone stimulus, such as the Fig6 method or the POGO2 method, yield the same target gains when individual hearing thresholds are identical. Therefore, the loudness perception of an individual is hardly considered. Also, the fitting procedure becomes an empirical one, which is very time-consuming for detailed adjustment. Fitting methods using fractional octave-band tone stimulus often result in excessive gains at low frequencies and too many measurements for loudness level setting are required. In this study, subjects with normal hearing are tested and the loudness perception to a certain level of band-limited white noise at the modified 14 critical bands is classified by five categories. A standard database, as the target value, is constructed by statistically processing the results. The same procedures are applied to a hearing impaired patient and the response data are used for estimating individual hearing characteristics. Acquired hearing loss data are compared with the standard database of normal hearing and then the target gains for five loudness categories are obtained for compensation. The proposed fitting method was compared with the existing methods for some patients and the results suggest that the auditory performance is better than the existing algorithms or empirical readjusting fitting.

Session 2pED

Education in Acoustics: Take 5's

Uwe J. Hansen, Chair
Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Do you have a novel demonstration, a new laboratory experiment, a favorite video, a recorded sound example, or a new idea for teaching acoustics? Share it with your colleagues. A sign-up board will be provided for scheduling presentations. No abstracts are printed. Presenters are encouraged to have handouts for distribution. Multiple presenations are acceptable (not consecutively). Presentations are limited to 5 minutes. Keep them short! Keep them fun!

Session 2pNS

Noise: Propulsion/Airframe Aeroacoustics II

Kevin P. Shepherd, Chair
NASA Langley Research Center, Hampton, Virginia 23681

Chair's Introduction—1:30

Invited Papers

1:35

2pNS1. Installed jet noise prediction with Jet3D. Craig Hunter (NASA Langley, MS 499, Hampton, VA 23681, craig.hunter@nasa.gov)

Much of the aircraft noise reduction research in recent decades has been concentrated at the component level in understanding the physics of noise generation, formulating noise prediction methods, and developing noise reduction strategies. A new strategy in NASA's noise reduction program is a focus on the noise effects specifically attributable to installation. This focus will also extend to developing noise reduction strategies that take advantage of installation effects. An essential requirement for an installed jet noise prediction method is that it must be able to predict noise from complex three-dimensional flows. Having this physics based capability is also desirable in order to develop a flexible noise prediction method applicable to the investigation of advanced concepts and revolutionary configurations. A jet noise prediction tool satisfying these objectives is currently under development at NASA Langley Research Center. The Jet3D methodology is based on Lighthills Acoustic Analogy and uses Reynolds-averaged Navier–Stokes (RANS) computational fluid dynamics (CFD) simulations from the PAB3D flow solver, with temperature-corrected two-equation turbulence closure and anisotropic Reynolds stress modeling. This talk describes development of the Jet3D noise prediction method and its application to installed jet configurations.

1:55

2pNS2. Engine configurations for the Silent Aircraft. Cesare Hall (Dept. of Eng., Univ. of Cambridge, Trumpington St., Cambridge CB2 1PZ, UK, cah1003@cam.ac.uk)

The Silent Aircraft Initiative is a Cambridge-MIT Institute research project aimed at reducing aircraft noise to the point where it would be unnoticeable in the urban areas around airports. This talk investigates viable engine and installation options that can meet the Silent Aircraft noise requirements. To reduce the transmitted jet noise sufficiently, it is found that a very large exhaust nozzle area is required. If this is achieved with a conventional turbofan it leads to a very high engine diameter and low fan pressure ratio. The large diameter results in greater engine weight and unacceptably high installation drag at cruise. The low fan pressure ratio makes the engine more susceptible to operability problems such as fan stall and vibration, which lead to a further increase in weight. This presentation demonstrates how these problems can be mitigated with a propulsion system that has a variable cycle combined with a novel installation that is embedded into the airframe. Several suitable engine concepts are considered in detail and each configuration is assessed in terms of its relative performance, weight, noise emission, and operability.

2:15

2pNS3. Acoustic shielding of engine noise by the Silent Aircraft airframe. Anurag Agarwal (Dept. of Eng., Univ. of Cambridge, Trumpington St., Cambridge CB2 1PZ, UK)

The ''Silent Aircraft Initiative'' is a project funded by Cambridge-MIT Institute (CMI). Its aim is to discover ways of reducing aircraft noise to the point where it would be virtually imperceptible to people outside the airport perimeter in a typical urban environment. The present design of the Silent Aircraft is in the form of a flying wing with a large wing planform and a propulsion system that is embedded in the rear of the airframe. Thus, a large part of the forward-propagating noise from the intake duct of the engines is expected to be shielded from observers on the ground by the large upper surface of the wing. In this talk, the use of boundary element methods to predict the attenuation of sound due to shielding by the wing is discussed. The reciprocity principle is invoked to study the noise contributions from several engine-intake locations at given observer locations on the ground. This procedure yields the optimum location of the engine intake for maximum noise shielding. Finally, a technique to incorporate the effect of background mean flows on sound propagation is briefly discussed.

2:35–2:50 Break

2:50

2pNS4. Innovative aircraft design for step changes in noise reduction. Zoltan S. Spakovszky (MIT Gas Turbine Lab., 77 Massachusetts Ave, Cambridge, MA 02139, zolti@mit.edu)

Low-noise integrated propulsion system concepts are proposed for functionally silent aircraft with the goal to reduce airframe and propulsion system noise by as much as 30 dB. Silent in this context means sufficiently quiet that the aircraft noise is less than that of the background noise in a typical well-populated environment. The theme of the technical approach for this multi-disciplinary problem is based on a systems view rather than an individual component view of the airframe and interacting propulsion system. Simple analytical modeling and existing semi-empirical noise prediction methods and scaling laws are used to predict the acoustic signature of low-noise airframe and propulsion system concepts envisioned for a functionally silent aircraft. The design study and acoustic analysis is based on an aerodynamically clean blended-wing-body-type airframe configuration. A distributed propulsion system is proposed to facilitate airframe boundary layer ingestion and to take advantage of shielding effects by embedding engines in the airframe. Ultra-high bypass ratio engines are necessary to reduce jet noise and multiple small engines or a multi-fan engine system is envisioned to enable integrating the propulsion system with the airframe. Results from the noise assessment studies are discussed and preliminary design implications for a functionally silent aircraft are given.

3:10

2pNS5. Systems analysis of advanced quiet aircraft. Erik Olson, Geoffrey Hill, Sherilyn Brown, Karl Geiselhart (NASA Langley Res. Ctr., MS 248, NASA-Langley Res. Ctr., Hampton, VA 23681-2199, erik.d.olson@nasa.gov), and Cecile Burg (Georgia Inst. of Technol., Hampton, VA 23681-2199)

To achieve future goals for the reduction of aircraft community noise, it may be necessary to use unconventional engine and wing arrangements. Prediction of the resulting noise benefits cannot be made without taking into account the effect of the configuration on the whole aircraft system, including the impact on the aerodynamics, weights, and systems integration. Systems studies have been conducted that quantify the impact of various configurations on both the performance and noise of the aircraft. In this paper, the results for several systems studies are presented. In one study, a large matrix was assembled for candidate configuration options for a blended wing–body (BWB) aircraft–including engine, inlet, and nozzle types and placement options—and ranked based on prioritization between noise benefits and performance penalties. In another study, a pair of concepts was developed and evaluated with the goal of simultaneously reducing the noise produced and reducing or eliminating the production of harmful chemical emissions. Finally, a BWB configuration was studied to determine the sensitivity of the overall noise to the levels of the various engine and airframe sources, quantify the airframe shielding benefits using experimental data, and explore the effects of wing trailing-edge extensions on the shielding.

Contributed Papers

3:30

2pNS6. Observations of aeroacoustic modifications of a supersonic jet by using a spherical reflector. Kunisato Seto and Md. Tawhidul Khan Islam (Dept. of Mech. Eng., Saga Univ., 1 Honjo-machi, Saga, 840-8502, Japan)

Jet screech was eliminated by placing a spherical reflector at the nozzle exit of an underexpanded supersonic jet. The placement of the reflector minimized the sound pressure and this muted sound did not excite the unstable disturbance at the exit of nozzle and the loop of feedback mechanism disappeared, thus the generation of jet screech was cancelled. The proposed method was very effective in reducing overall sound pressure in the upstream region of the nozzle exit. Therefore, it could be a promising countermeasure against acoustic fatigue to protect the fuselage of an aircraft by incorporating it into a streamline form. The flow characteristic of the jet for the proposed system was observed by the Schlieren apparatus with a high speed video camera. A slit path was cut to the reflector for visualizing the jet properly. The shock structure of the jet was slightly modified and became more stable. [The research was supported by a grant from JSPS of the government of Japan.]

3:45

2pNS7. Control of supersonic jet noise with a spherical reflector. Md. Tawhidul Khan Islam and Kunisato Seto (Dept. of Mech. Eng., Saga Univ., 1 Honjo-machi, Saga, 840-8502, Japan)

Jet screech is considered to be generated from a feedback cycle in an underexpanded supersonic jet. Experiments were carried out to control the screech tone by placing a spherical reflector at the nozzle exit. The reflector controlled the location of the sound image-source and minimized the sound pressure at the nozzle exit. The weak sound could not excite the unstable disturbances at the exit and the feedback mechanism was cancelled and finally screech tone was eliminated. The new technique sup-

pressed not only the screech tones but also the broadband noise and reduced the overall sound pressure of the jet. The performance of the proposed technique was checked with different sizes of reflectors by placing the reflectors at different upstream positions of the nozzle exit for different pressure ratios and the result was very good. [The research was supported by a grant from JSPS of the government of Japan.]

TUESDAY AFTERNOON, 16 NOVEMBER 2004 PACIFIC SALONS 4 & 5, 1:00 TO 5:15 P.M.

Session 2pPA

Physical Acoustics and Biomedical Ultrasound/Bioresponse to Vibration: Sonoluminescence, Sonochemistry and Sonofusion

Thomas J. Matula, Chair
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698

Invited Papers

1:00

2pPA1. What are the limits of energy focusing in sonoluminescence? Seth Putterman, C. Camara, B. Kappus, C. K. Su, E. Kirilov, and A. Chakravarty (Phys. Dept., UCLA, Los Angeles, CA 90095)

Sonoluminescence (''SL'') is an amazing marker for the extraordinary degree of acoustic energy focusing achieved in a cavitating bubble. Local energy dissipation exceeds Kirchhoff's law by 10^{15} and the ambient acoustic energy density concentrates by 12 orders of magnitude to create picosecond flashes of broadband ultraviolet light. For single bubbles driven at 30 kHz, SL is nature's smallest blackbody. Therefore the bubble's interior is such a dense plasma that the photon-matter mean-free path is shorter than the wavelength of light. Excitation of a vertical column of fluid ($\sim$50 Hz), so as to create a water hammer, upscales flash energy by a factor over one million, achieving peak powers approach 1 W. At 1 MHz the spectrum resembles Bremsstrahlung from a transparant plasma with a temperature $\sim$1 MK. At 10 MHz the collapsed size of the SL bubble approaches 10 nm, which raises the possibility that the SL parameter space may extend to the domain of quantum mechanics. At 30 MHz experiments are under way to excite sonoluminescence with sound fields in excess of 3000 atm. The strongest cavitation collapses may be realized with Greenspan's ultrasonic resonators that reach fields in excess of 20 atm without cavitating. When bubbles are seeded with an external laser a massive cavitation event ensues.

1:30

2pPA2. Additional evidence of nuclear emissions during acoustic cavitation. Rusi Taleyarkhan, Colin West, JaeSeon Cho, and Richard Lahey, Jr. (Purdue Univ., W. Lafayette, IN 47907-1290, rusi@purdue.edu)

Time spectra of neutron and sonoluminescence emissions were measured in cavitation experiments with chilled deuterated acetone. Statistically significant neutron and gamma ray emissions (with more than 25 standard deviation accuracy) were measured with a calibrated liquid-scintillation detector. The neutron and sonoluminescence emissions were found to be time correlated over the time of significant bubble cluster dynamics. The neutron emission energy was at and below 2.45 MeV and the neutron emission rate was up to $\sim$400 000 n/s. Measurements of tritium production were also performed and these data implied a neutron emission rate due to D-D fusion which agreed with what was measured. In contrast, control experiments using normal acetone did not result in statistically significant tritium activity, or neutron or gamma ray emissions. The talk will highlight significant details of the acoustic chamber design, characterization and qualification along with the nuclear data obtained.

2:00

2pPA3. The temperatures of single-bubble sonoluminescence. Kenneth S. Suslick and David J. Flannigan (School of Chemical Sci., Univ. of Illinois at Urbana-Champaign, 600 S. Mathews Ave., Urbana, IL 61801)

We observe extraordinarily intense single-bubble sonoluminescence (SBSL) from concentrated sulfuric acid (H2SO4) containing noble gases. Strong atomic Ar emission and extensive vibronic progressions from sulfur monoxide (SO) are also present in the SBSL spectra. The Ar atom excited states observed are too high in energy to be thermally populated and must be excited by high energy particle impact, consistent with Ar atom SBSL from an emissive shell surrounding an optically opaque plasma core, just as in a star or thermonuclear explosion. From relative intensities of Ar lines, we find that the observed effective emission temperature during SBSL is 15 200$\pm$1900 K. SBSL emission temperatures can be systematically controlled over the range from $\sim$1500 to $\sim$20 000 K by changing the applied acoustic pressure or the thermal conductivity of the dissolved gas.

2:30

2pPA4. High-frequency, high-power ultrasonic chemical reactors. Michael Hoffmann, Timothy Lesko, and Agustin Colussi (California Inst. of Technol., Eng. & Appl. Sci. 138-78 CIT, Pasadena, CA 91125)

A novel high-frequency (612 kHz), high-power (4 kW), pilot-plant scale sonochemical reactor was developed and used to study the degradation of organic chemical pollutants in aqueous solutions. The degradation rates of trichloroethylene, dichloromethane, and phenol were found to exceed those of similar frequency, small-scale bench reactors by factors ranging from 2.5 to 7. In addition, there

is linear dependence between the observed sonolytic rate constants and the applied power density. The addition of ozone during sonication did not affect the first-order degradation rate constants for phenol degradation compared to the linear combination of sonication and ozonation. However, enhancement in the degradation rates of the total organic carbon (TOC) were observed. The enhanced reactivity of sonolysis coupled with ozonation is due to the sonolytic formation of hydrogen peroxide in water, which in turn reacts with ozone to form a highly reactive ozonide intermediate, dihydrogen trioxide, that reacts in a similar fashion to hydroxyl radical. However, its lifetime in aqueous solution is found to be substantially longer, and thus it is more likely to react with refractory organic compound fragments.

3:00

2pPA5. The effect of surfactants on inertial cavitation activity and sonoluminescence intensity in aqueous solutions. Muthupandian Ashokkumar, Judy Lee, Sandra Kentish, Franz Grieser (School of Chemistry, Univ. of Melbourne, VIC 3010, Australia), and Thomas J. Matula (Univ. of Washington, Seattle, Washington 98105)

The understanding and influencing of bubble clusters in an ultrasonic filed is important for many applications. We have investigated the various concentration-dependent effects of sodium dodecyl sulfate (SDS), an anionic surfactant, in aqueous solutions on bubble coalescence, stability of bubble nuclei and rate of bubble growth in an ultrasonic field. For example, the addition of low concentrations of SDS has been found [J. Phys. Chem. B, **101**, 10845 1997] to enhance the intensity of multibubble sonoluminescence (MBSL). Further investigations on the effect of SDS on MBSL intensity have revealed that the SL enhancement observed is highly dependent on the ultrasound pulse length. We have observed that there exists a critical (ultrasound) pulse length below which the addition of SDS in fact lowers the MBSL intensity relative to that observed from water. The inertial cavitation activity has been experimentally measured under similar conditions to that of the SL experiments and has been found to correlate with MBSL results. Both MBSL and inertial cavitation activity results suggest that the presence of SDS at higher concentrations could also enhance the rate of bubble growth by rectified diffusion. This presentation will focus on how specific surfactants influence bubble cluster dynamics and bubble coalescence.

3:30–3:45 Break

3:45

2pPA6. Acoustically assisted cleaning in microelectronics below 65 nanometers (nm): An urgent need to understand mechanisms and effects. Gary Ferrell (SEZ America, Inc., 2632 Bayshore Pkwy., Mountain View, CA 94043)

1000-kHz ultrasonics (megasonics) has been an important technology in controlling killer defects in modern microelectronics. In the construction of, say, a microprocessor, perhaps 10 percent of the 500 process steps may involve acoustically aided cleaning. Feature sizes are currently at 65 nm and will be at 32 nm by 2009. In our laboratory, we have been utilizing multiple-bubble sonoluminescence (MBSL) as a probe into cavitation processes in the hopes of learning to control bubble dynamics and growth. Strong evidence has been found which links MBSL with the removal of 100 nm and larger particles on single-crystal silicon substrates during wafer cleaning. Improved understanding of the bubble interaction with the surface features must be developed in order to maximize cleaning and eliminate damage to nanoscale features. Surface tension modification, overpressure, and nucleation control are some of the experimental controls being developed. A novel nanoscale optical damage probe is in development to further aid in understanding the forces impinging on surface features.

Contributed Papers

4:15

2pPA7. Temporal evolution of sigle bubble sonoluminescence spectra: From bubble creation to stabilization. Jeremy Young (NASA John Glynn Res. Ctr., 21000 Brookpark Rd., Cleveland, OH 44135) and Thomas Matula (Univ. of Washington, Seattle, WA 98105)

Single bubble sonoluminescence (SBSL) is generated by creating and levitating a bubble at the pressure antinode of a standing wave. Previously, it was determined that the SBSL total light intensity evolves during the stabilization process [T. J. Matula and L. A. Crum, Phys. Rev. Lett. **80**, 865–868 (1998)]. The evolution process was used to confirm a theoretical hypothesis that chemical diffusion effects were important for stable SBSL. In this talk we report how the emission spectrum evolves over time, from bubble creation to stabilization. SBSL was generated in a 100-ml spherical cell filled with degassed (air) filtered water. Boiling from a nichrome wire created bubbles which were then forced to the pressure antinode. Spectra were recorded with a multichannel PMT (Burle Industries) using 16 color channels in a 4×4 grid with 20-nm bandwidth interference filters covering the range from 250 to 625 nm. Continuous recording of the spectra (with a 192-ms integration time) was performed until the bubble stabilized. Spatial, crosstalk, filter, and tube response calibrations were performed. Temporal evolution measurements showed evidence of an early-stage emission band near 300 nm, suggesting that hydroxyl emission becomes swamped by the continuum during stabilization. [Research was funded by NASA.]

4:30

2pPA8. Stable multibubble sonoluminescence. Larry R. Greenwood, Gerald J. Posakony, Leonard J. Bond, Morris S. Good (Pacific Northwest Natl. Lab., P.O. Box 999, Richland, WA 99354), Salahuddin Ahmed, Michael D. Wojcik, Warren W. Harper, and Marino Morra (Pacific Northwest Natl. Lab., Richland, WA 99354)

A multibubble standing wave pattern can be generated from an acoustic wave reflected from a flat surface. By adding a second transducer at 90 deg from the transducer generating the standing wave, a three-dimensional volume of stable single bubbles can be established. Further, the addition of the second transducer operating at the same frequency stabilizes the bubble pattern so that individual bubbles may be studied. The size of the bubbles and the separation of the standing waves depend on the frequency of operation. Two transducers, operating at frequency of 630 kHz, provided the most consistent results for the configuration used in this study. The bubbles exhibit bright sonoluminescence. Spectral measurements are in progress. Effect of the shape of the configuration will be discussed

along with the standing wave patterns, spectral data, and pictorial results of separation of individual bubble sonoluminescence in a multibubble sonoluminescence environment.

4:45

2pPA9. Physical analysis of a stable cavitation bubble structure at high acoustic intensity. Bertrand Dubus, Alexei Moussatov, Christian Granger (IEMN dpt ISEN, UMR CNRS 8520, 41 bd Vauban, 59046 Lille Cedex, France, bertrand.dubus@isen.fr), Cleofe Campos-Pozuelo (CSIC, Madrid, Spain), Christian Vanhille (Universidad Rey Juan Carlos, Madrid, Spain), Robert Mettin, Topi Tervo, and Werner Lauterborn (Gottingen Univ., Germany)

A cavitation bubble structure stable at high acoustic intensity (from 1.8 to more than 8.2 W/cm^2) has been experimentally observed [A. Moussatov *et al.*, Ultrason. Sonochem. **10**, (2003)]. At the vincinity of an axisymmetrical radiating surface, big streamers of bubbles get ejected from the surface and build up a bubble structure of conical shape denoted CBS. In this paper, results on the observation and analysis of the CBS are reported for 20-kHz horn-type transducers with different sonotrode diameters. It is found that: (i) the CBS is always a zone of high chemical activity, even when the bubble structure is not observed due to high speed streaming; (ii) the geometry of the CBS is determined by nonlinear acoustic wave propagation. These results are supported by various experimental data: chemiluminescence measurements, high speed movies (2250 frames/s) under CW scattered light or LED flashes synchronized with driving signal and measurement of the acoustic pressure and of the time-averaged acoustic pressure. [Work supported by CNRS-CSIC cooperation project and European Union (FEDER 2).]

5:00

2pPA10. The influence of ultrasound power on multibubble sonoluminescence intensity from aqueous solutions containing surface active solutes. Devi Sunartio, Muthupandian Ashokkumar, and Franz Grieser (School of Chemistry, Univ. of Melbourne, VIC 3010, Australia)

Water soluble surface active solutes have been found to affect the intensity of multibubble sonoluminescence (MBSL) in aqueous solutions. For example, the presence of charged surfactants enhances the MBSL intensity relative to that observed from pure water, whereas the presence of volatile surface active solutes decreases the MBSL intensity in reference to pure water [J. Phys. Chem. B **101**, 10845 (1997)]. Further investigations on how ultrasound power influences the effect of surface active solutes on MBSL intensity have shown that ultrasound power plays an important role in governing the above mentioned effects. The relative enhancement in MBSL intensity by charged surfactants has been found to vary with changes in ultrasound power level; the relative enhancement decreases with an increase in the ultrasound power level. Also, the extent of MBSL quenching by alcohols has been found to increase with an increase in the applied ultrasound power level; no SL quenching at lower power levels and $>80\%$ SL quenching at higher power levels have been observed. The influence of ultrasound power on the population of ''active'' bubbles and the cavitation bubble temperature, in aqueous solutions containing surface active solutes, will be discussed in order to rationalize the observed experimental data.

TUESDAY AFTERNOON, 16 NOVEMBER 2004 ROYAL PALM SALON 5, 1:30 TO 5:00 P.M.

Session 2pSA

Structural Acoustics and Vibration: Structural Acoustics in MEMS

Karl Grosh, Chair

Mechanical Engineering Department, University of Michigan, 2350 Hayward Street, Ann Arbor, Michigan 48109-2125

Invited Papers

1:30

2pSA1. Viscothermal wave propagation, including acousto-elastic interaction. Willem Beltman (Intel, 2111 NE 25th Ave., M/S JF3-254, Hillsboro, OR 97124, willem.m.beltman@intel.com)

Standard wave equation models neglect the effects of viscosity and thermal conductivity on acoustic wave propagation. For wave propagation in narrow tubes or thin layers, like in MEMS devices, this might not be accurate. This presentation outlines models that take the effects of inertia, viscosity, thermal conductivity and compressibility into account. Three classes of models are described and characterized based on the use of dimensionless parameters. The most important parameter is the shear wave number, an unsteady Reynolds number that indicates the ratio between inertial and viscous effects. It is shown that for most applications the low reduced frequency model is sufficient and the most efficient. The wave propagation models are coupled to the structural models to capture the acousto-elastic interaction. For simple geometries, analytical solutions can be found for these coupled analysis cases. For more complex geometries, a finite element model was developed, based on the low reduced frequency model, in which viscothermal acoustic finite elements are coupled to structural elements. Examples of applications are presented.

2:00

2pSA2. Fluid–structure interaction in a physical model of the human cochlea. Michael J. Wittbrodt, Charles R. Steele, and Sunil Puria (Mech. and Computation Group, Mech. Eng., 262 Durand Bldg., Stanford Univ., Stanford, CA 94305, wittbrod@stanford.edu)

Understanding the mammalian cochlea requires an understanding of the complex fluid structure interaction of an elastic partition separating two fluid channels. Using a combination of microfabrication and macrofabrication technologies, a passive physical model was developed. Compared to most MEMS devices the elastic partition is large. The width is tapered from 100 to 500 μ over the 36-mm length. Orthotropic properties are achieved with 9000 discrete aluminum fibers supported by either 1 or 5 μ of a polyimide

thin film. Two fluid channels are macromachined from plastic and filled with saline. A magnet-coil system excites the fluid channel. The model demonstrates a traveling wave which peaks at a characteristic place. Normalized responses show gains of 2.5 and 3 at 29 and 23 mm from the basal end with phase lags of 3 pi and 5 pi for 8 and 18 kHz, respectively. Calculations using the WKB asymptotic approximation confirm the general character of the responses measured. The presence of fluid provides an efficient means of transporting the wave energy to a characteristic place on the elastic partition followed by a sharp roll-off in response. [Work supported in part by grants DC03085 and DC05454 from the NIDCD of NIH.]

2:30

2pSA3. Sensors at small scales for applications in fluid structure interaction problems. Vasundara V. Varadan (Eng. Sci. Elect. Eng., The Pennsylvania State Univ., University Park, PA 16804, vvvesm@engr.psu.edu)

This talk will present an overview of the state of the art in the development of sensors at the nano and micro scale that are particularly suited for the study of small scale effects in fluid structure interaction problems. The parameters of interest pertain to pressure, shear stress, thermal effects, material surface modifications, and sensors and devices that play a role in microfluidics for applications in materials analysis on a chip as well as biomedical uses. In several applications wireless telemetry of data is a necessity and special challenges arise if devices are immersed in a liquid. Power sources for such devices as well as the telemetry system is also a research issue. Passive sensors such as remotely excited rf surface acoustic wave sensors are also of great interest. Real time correlation of data from several sensors measuring either the same or different parameters is also needed to provide parametric data for adaptive computer models that start with incomplete system models but rely on real time experimental data to evolve the computation. Active collaboration is needed between researchers working on device design and development and those studying fluid–structure interaction problems either experimentally or numerically.

3:00

2pSA4. Aspects of acoustics in MEMS devices. Jeffrey Dohner, Mark Jenkins, and Timothy Walsh (Sandia Natl. Labs., P.O. Box 5800, Albuquerque, NM 87185)

In this talk we will present an overview of aspects of structural acoustics in MEMS devices at Sandia. Recent results concerning viscous wave motion in a rotational micromechanical system will be presented. In addition, experimental approaches for the characterization of acoustic effects in MEMS devices will be discussed. We will also present an overview of massively parallel numerical simulation capabilities for large-scale and small-scale (MEMS) structural acoustic analysis. Since large numbers of degrees of freedom are typically present in structural acoustic analysis, massively parallel computations are essential in solving practical application problems. Two formulations for the fluid will be considered: a standard linear velocity potential formulation, and a nonlinear wave formulation. The applications of the two formulations will be discussed. A parallelization scheme will be presented that allows for arbitrary decomposition of the wet interface. Finally, numerical results of application problems will be presented. [Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.]

3:30–3:45 Break

3:45

2pSA5. An all surface-machined MEMS microphone. Gary W. Elko (Avaya Labs, 233 Mt. Airy Rd., Basking Ridge, NJ 07920, gwe@ieee.org), Flavio Pardo, Daniel Lopez, and David Bishop (Bell Labs, Lucent Technologies)

An all-surface machined MEMS microphone for telephony applications was constructed a few years ago at Bell Labs. This microphone was built-up from multiple layers of standard polysilicon. After etching, the polysilicon layers were mechanically actuated in an origami-like fashion, to form a condenser microphone shaped in a tetrahedral structure. The overall size of the microphone was roughly 300 μ on a side. The enclosed back reference volume was designed in concert with a unique multiple cantilever beam design to realize a diaphragm resonance frequency of approximately 20 kHz. A discrete component preamplifier based on a modulated carrier detection circuit to reduce $1/f$ preamplifier noise was constructed. The predicted overall noise level of the microphone was approximately 45 dBA. The measured noise was significantly higher, due most likely to stray capacitance of the lead wires to the external discrete preamplifier. Some design issues related to mechanical-thermal noise will be discussed and some suggestions to mitigate self-noise will be given.

Contributed Papers

4:15

2pSA6. Fluid–structure waves in a micromachined variable impedance waveguide. Robert White, Niranjan Deo, and Karl Grosh (Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48109, grosh@umich.edu)

A micromachined fluid–structure system, intended to demonstrate a unique cochlear-like acoustic sensing scheme, has been fabricated and analyzed. The system consists of a 0.11-mm-deep Pyrex fluid chamber, anodically bonded to a silicon structure housing a tensioned membrane. The membrane varies in width from 0.14 to 1.82 mm. Both isotropic LPCVD silicon nitride membranes and orthotropic nitride/polyimide membranes have been fabricated. Silicone oils of either 5 or 20 cSt viscosity are used. Laser vibrometry measurements show strong fluid–structure traveling waves. Wave speeds are between 50 and 300 m/s in the 4- to 35-kHz band. These traveling fluid–structure waves exhibit maximum structural motion at a location determined by their frequency. Wave decay rate is influenced by the viscosity of the fluid and, after the waves become evanescent, by membrane orthotropy. Results from finite element computations of an orthotropic membrane coupled to a compressible viscous fluid are compared with experiment. The fluid is modeled using either a two-dimensional thin-film approximation or the three-dimensional linearized Navier–Stokes equation. Both models accurately predict the

observed fluid–structure response. Nondimensional parameters controlling system performance, and regions of applicability of the models, will be discussed. [Work supported by ONR, NIH, and NSF.]

4:30

2pSA7. A theory for flexural vibration of thin plates with thermoelastic damping. Andrew Norris (Mech. and Aerosp. Eng., Rutgers Univ., Piscataway, NJ 08854-8058, norris@rutgers.edu)

Thermoelastic damping in thin plates can be the dominant loss mechanism under certain circumstances, e.g., as demonstrated by recent measurements for MEMS paddle oscillators. However, modeling of thermoelastic damping in flexural vibration of plates has lagged, despite the classical theory of Zener for beams. This talk shows how the thinness of the plate permits a very accurate approximation to the fully coupled system of stress and temperature. In this asymptotically valid theory, the flexural plate equations are modified by damping terms of two types. The first are viscosity-like terms in the plate equation of motion. These depend on the local curvature, being zero at saddles. The damping also modifies the boundary moment and shear force. Both the bulk and boundary effects are important in estimating the Q for resonant modes. The theory is derived from a generalized form of Hamilton's principle, using Kirchhoff's kinematic assumption and asymptotic approximations for small thermal coupling valid for all relevant materials. Modal damping can be estimated simply for certain boundary conditions, including fixed edges. Plates with free or partially free edges present a more interesting challenge, and various alternative estimates of Q are presented and illustrated numerically for rectangular plates and MEMS structures.

4:45

2pSA8. Development of film bulk acoustic wave resonators based on piezoelectric aluminum nitride. Hisanori Matsumoto, Kengo Asai, and Mitsutaka Hikita (Hitachi, Ltd., Central Res. Lab., 1-280, Higashi-koigakubo, Kokubunji-shi, Tokyo 185-8601, Japan, h-matsu@crl.hitachi.co.jp)

Film bulk acoustic wave resonators (FBARs) have high quality factors (Q) and are smaller than conventional surface acoustic wave devices, so they are particularly suitable for constructing the bandpass filters of mobile phones. This report covers the simulation and fabrication of FBARs consisting of an aluminum nitride (AlN) piezoelectric layer sandwiched between molybdenum (Mo) electrode layers. The simulation was based on Mason's equivalent circuit model. Other than material Q, the AlN material constants for the FBAR model were those reported by Tsubouchi and Mikoshiba [K. Tsubouchi and N. Mikoshiba, IEEE Trans. Sonics Ultrason. **32**, 634–644 (1985)]. The material Q of AlN was assumed to be 1000. Thickness values for the AlN and Mo films were determined by simulation, after which the FBARs were fabricated. Reactive radio-frequency sputtering and direct-current sputtering were adopted for deposition of the AlN and Mo films, respectively. In terms of frequency properties, our FBARs achieved Q of 800 and an effective electromechanical coupling coefficient of 5.7%. The transmission loss was greater than that calculated through simulation. This result suggests that the material Q of AlN is less than 1000. This may be due to suboptimal film quality.

TUESDAY AFTERNOON, 16 NOVEMBER 2004 CALIFORNIA ROOM, 2:00 TO 5:00 P.M.

Session 2pSC

Speech Communication: Measuring and Modeling Voice and Talker Characteristics (Poster Session)

Roger W. Chan, Chair
University of Texas Southwestern Medical Center, Otolaryngology-Head Neck Surgery, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9035

Contributed Papers

All posters will be on display from 2:00 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 2:00 p.m. to 3:30 p.m. and contributors of even-numbered papers will be at their posters from 3:30 p.m. to 5:00 p.m.

2pSC1. A search for listener differences in the perception of talker identity. Robert E. Remez, Stephanie C. Wissig, Daria F. Ferro, Kate Liberman, and Claire Landau (Dept. of Psychol., Barnard College, 3009 Broadway, New York, NY 10027)

Talkers differ along several dimensions that listeners can resolve. These include acoustic effects of native variation in the scale and shape of the articulatory anatomy, and the effects of age and use on the tissues of the vocal tract. Linguistically, talkers differ in phonetic habits occasioned by dialect and idiolect, and differ paralinguistically in manner of affective expression conveyed vocally. However, listeners are themselves likely to vary in sensitivity to intertalker variation. In this study we aimed to identify differences in sensitivity to talker variation as a function of linguistic experience. Speech samples were produced by female talkers 15–17 years old drawn from two dialect groups, one from Brooklyn, NY, and one from Bloomington, IN. Each talker produced sentences in a list-reading task. Listeners in our tests were also native either to Brooklyn or to Bloomington, and each was far more familiar with one dialect than the other. Tests of apparent similarity of talkers in each set were conducted with listeners from the same and different dialect group. The results of similarity scaling analyses calibrate the contribution of sensitivity to idiolectal contrast within and across dialect in the perception of a talker's characteristics. [Research supported by NIDCD.]

2pSC2. Relative contribution of temporal cues and spectral profile to voice gender discrimination. Sherol Chinchilla-Rodriguez, Geri Nogaki, and Qian-Jie Fu (House Ear Inst., 2100 W. 3rd St., DAIP, Los Angeles, CA 90057, schinchilla@hei.org)

Despite limited access to spectral and temporal cues, cochlear implant (CI) users are somewhat able to discriminate voice gender. The present study explored the relative contributions of spectral and temporal cues to normal-hearing (NH) subjects' voice gender discrimination, while listening to multi-channel simulations of CI processing. The output spectrum

was either matched (relative to normal) or upwardly shifted to simulate the spectral shift associated with CIs; the envelope filter in each channel was varied to examine the contribution of temporal cues. Voice gender discrimination was tested with two talker sets, in which the mean fundamental frequency ($F0$) between male and female talkers was either widely or narrowly separated. Results showed that for both talker sets, 16 spectral channels were needed before subjects could use the spectral profile to identify voice gender; when the speech spectrum was shifted, 32 channels were needed. Given enough temporal cues, the spectral profile had a relatively small effect on discrimination when the $F0$ was widely separated between male and female talkers. When there was little $F0$ separation between male and female talkers, the spectral profile had a much stronger effect; however, 16 or more channels were needed before listeners could attend to the profile. [Work supported by NIDCD-RO1-DC004993.]

2pSC3. Temporal change in memory of human voice. Hiroshi Kido (Dept. of Commun. Eng., Tohoku Inst. of Technol., Sendai, Japan, kidoh@tohtech.ac.jp) and Hideki Kasuya (Utsunomiya Univ., Utsunomiya, Japan)

This study investigates the extent to which human subjects accurately retain memory of speaker individuality and voice quality of utterances without receiving instruction to memorize them. Experiments constituted three phases. In phase one, 27 subjects listened to two target utterances in a paired-comparison test that required only a similarity judgment of voice quality. The target utterances were one sentence read by two males with average and characteristic voice quality, respectively, who were selected from 109 males. In phase two, performed 21 days after phase one, the same subjects evaluated Japanese expressions of voice quality [Proc. ICSLP-98, No. 1005] of the two previous target utterances based on memory only and judged with a degree of confidence whether each of seven speech stimuli (two targets and five additional speakers' utterances) was identical to one of the two targets. In the final phase, conducted immediately after phase two, the subjects repeated the same phase-two task but after listening to the two target utterances. Data processing based on the signal detection theory revealed that temporal change in memory of speaker individuality and voice quality was relatively small over a period of 21 days. Experiments over a longer period of time are now underway.

2pSC4. Multiregression analysis of autoregressive with exogenous input speech synthesis parameters and voice qualities. Hideki Kasuya, Masayoshi Kawamata (Utsunomiya Univ., 7-1-2 Yoto, Utsunomiya, Japan, kasuya@klab.jp), and Hiroshi Kido (Tohoku Inst. Technol., Sendai, Japan)

This study investigates the relationship between acoustic parameters utilized in the formant-based ARX (autoregressive with exogenous input) speech synthesis model (J. Acoust. Soc. Jpn., **58**, 386–397) and perceived voice qualities of synthetic speech. The acoustic parameters manipulated were $F0$ baseline, $F0$ range, spectral tilt of glottal flow (TL), formant scaling parameter (FS), and speaking rate (SR). Japanese expressions associated with voice qualities were high-pitched/low-pitched, masculine/feminine, hoarse/clear, calm/excited, powerful/weak, youthful/elderly, thick/thin, and tense/lax (Proc. ICSLP-98, No. 1005). A sentence utterance of an average speaker selected from a database of 109 male speakers was analyzed using the ARX method. Each of the five acoustic parameters of the utterance was manipulated at three levels, producing 243 samples of synthetic speech ($3\times3\times3\times3\times3$). Ten subjects evaluated the voice qualities of each of the 243 synthetic stimuli with regard to the eight Japanese expressions. Multiregression analysis showed that $F0$ range, $F0$ baseline, and FS were primary acoustic correlates of high-pitched/low-pitched and masculine/feminine, SR and $F0$ range for calm/excited, and $F0$ range, SR and $F0$ baseline for thick/thin. Significant relations were not found for the remainder of the Japanese expressions, which was thought to be associated in part with irregularities of glottal flow.

2pSC5. Fibrous proteins and tensile elasticity of the human vocal fold. Roger W. Chan (Otolaryngol. Head and Neck Surgery, Grad. Program in Biomed. Eng., Univ. of Texas Southwestern Medical Ctr., Dallas, TX 75390-9035), Neeraj Tirunagari, and Min Fu (Univ. of Texas Southwestern Medical Ctr., Dallas, TX)

Viscoelastic response of the human vocal fold under tension has been reported previously, demonstrating nonlinearity and hysteresis of stress–strain curves. However, the importance of various matrix molecules for tensile elasticity of the vocal fold is not well understood. It is important to correlate the biomechanical behavior of the vocal fold with its histological microstructure, so as to examine the relative contributions of the matrix proteins such that cultured tissue constructs for repairing voice disorders may be engineered accordingly. This study attempted to test the hypothesis that collagen and elastin play a major role in determining the tensile elasticity of the vocal fold. Uniaxial tensile elastic properties of the vocal cover and vocal ligament were quantified using a servo-control lever system. The specimens were also studied histologically with elastin van Gieson stain. Digital image analysis of the vocal fold cover showed that a more profoundly nonlinear stress–strain curve and a higher elastic modulus (tangent Young's modulus) were associated with higher densities of collagen and elastin in the specimens. These findings suggested that both collagen and elastin fibers likely contribute significantly to elasticity of the vocal fold under tension, thereby regulating vocal fold length changes and fundamental frequency control. [Work supported by NIH.]

2pSC6. Source spectra for excised and latex phonatory models. Fariborz Alipour (Dept. of Speech Pathol. & Audiol., Univ. of Iowa, Iowa City, IA 52242) and Ronald C. Scherer (Bowling Green State Univ., Bowling Green, OH 43403)

An excised larynx model and a latex physical model were used to study acoustic spectra of the phonatory source as a function of subglottal pressure, glottal adduction, and vocal-fold length. A supraglottal vocal tract was not used, indicating that the acoustic signal corresponded to the output glottal flow. Each model was mounted over an -inch tracheal tube through which flowed pressurized, heated, and humidified air. The subglottal pressure and EGG signal (excised model) were recorded on a personal computer. The mean flow rate, mean subglottal pressure, and SPL were recorded manually, and adduction was specified by the use of interarytenoid shims. The output audio signals from the larynx models were recorded on a DAT recorder. Spectral information of the audio signal was obtained via FFT analysis (MATLAB). Preliminary data indicate that the spectral slope did not have a constant dB/octave rate, and spectral slope had a primary dependence on subglottal pressure and a secondary dependence on adduction. Other acoustic details and the differences between the two models will be discussed. [Work supported by NIDCD Grant Number DC03566.]

2pSC7. Vocal tract influence on medial surface dynamics of the vocal folds. Michael Doellinger, David Berry (The Laryngeal Dynam. Lab., UCLA School of Medicine, 31-24 Rehabilitation Ctr., 1000 Veteran Ave., Los Angeles, CA 90095-1794), and Douglas Montequin (Univ. of Wisconsin, Madison, WI 53792-7375)

In previous work, quantitative measurement of the medial surface dynamics of the vocal folds was reported using a hemilarynx methodology. The technique was applied to excised larynges from both humans and canines, as well as to *in vivo* canine larynges. In the present investigation, a vocal tract was attached to the excised hemilarynx preparation, and systematic changes were made in vocal tract shape to study its influence on the medial surface dynamics. In particular, the width of the epilarynx was varied across experiments. Previously, in a similar experiment, phonation threshold pressure was investigated as a function of epilarynx width. It has also been shown that epilarynx width has an influence on glottal volume velocity and the acoustic output [I. R. Titze and B. H. Story, J. Acoust. Soc. Am. **101**, 2234–2243]. However, the direct influence of the epilarynx width on vocal fold vibration has never been quan-

tified. The present investigation considered the impact of epilarynx width on the medial surface dynamics, including the underlying empirical eigenfunctions which make up the vibration patterns. Further, quantitative measures such as displacement and velocity were reported and compared. [Work supported by NIH/NIDCD Grant No. R01 DC03072.]

2pSC8. The application of chaotic dynamics, synchronization, and parameter estimation in an asymmetric vocal fold system. Y. Zhang and J. J. Jiang (Dept. of Surgery, Div. of Otolaryngol. Head and Neck Surgery, Univ. of Wisconsin Med. School, Madison, WI 53792-7375)

Methods from nonlinear dynamics, including Poincare map, Lyapunov exponent and dimension, are applied to describe the vibrations of a vocal fold model with tension, stiffness, and mass imbalances. Bifurcation diagrams illustrate the effects of these imbalance parameters. When tension, stiffness, and mass imbalance parameters deviate from the normal value of 1, chaotic vibrations with positive Lyapunov exponents may occur. Furthermore, the technique of feedback synchronization allows us to manipulate a simulator to approach the vibratory patterns of an original vocal fold system. The minimal glottal area is applied as a feedback variable connecting between two systems. The simulator and the original system are synchronized when their state differences asymptotically converge to zero. The effects of noise and parameter mismatches on synchronization are investigated. Finally, a parameter estimation scheme based on feedback synchronization is applied. Despite noise perturbations and large initial parameter mismatches, the original system parameters can be reproduced in the simulator with parameter controls, and two chaotic vocal fold systems can be synchronized. A parameter estimation scheme shows the potential application to extract asymmetric biomechanical parameters of the vocal folds from a time series of the glottal area. [Work supported by NIDCD Grant No. R01 DC006019-01.]

2pSC9. An acoustic and electroglottographic study of V[glottal stop]V in two indigenous American languages. Christina Esposito (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095) and Rebecca Scarborough (UCLA, 3125 Campbell Hall, Los Angeles, CA 90095)

Both Pima, a Uto-Aztecan language spoken in Arizona, and Santa Ana del Valle Zapotec (SADVZ), an Otomanguean language spoken in Oaxaca, Mexico, have sequences of two vowels separated by an intervening glottal stop. In both languages, this V?V sequence becomes reduced in certain occurrences, with the perceptual effect of the loss of /?/ in Pima and the loss of V2 in SADVZ. The purpose of this study is to provide an acoustic and electroglottographic (EGG) description of these sequences in both full and reduced forms, prompted by varying speech rate. Two acoustic measures of phonation type (H1–H2, H1–A3) and two EGG measures (OQ and peak closing velocity) were made for each vowel at the midpoints and adjacent to /?/. For Pima, an issue of interest is what properties of the /V?V/ sequences (when V1=V2) allow them to be distinguished from phonemic long vowels in the reduced forms where /?/ is lost. It is hypothesized that /?/ will be preserved as vowel glottalization. For SADVZ, an important question is what property of V?(V) (where V2 is devoiced), V?, and creaky vowels allow them to be distinguished from each other in reduced forms, given that they are all characterized by glottalization.

2pSC10. The feature [stiff] interacts with intonation to affect vocal-fold vibration characteristics. Helen M. Hanson (36-585, MIT RLE, 77 Massachusetts Ave., Cambridge, MA 02139, hanson@speech.mit.edu)

As part of a larger study of the effect of prosody on segmental cues, previous work has shown that in a high pitch environment, $F0$ is significantly increased relative to a baseline following voiceless obstruents, but $F0$ closely follows the baseline following voiced obstruents. When a syllable carries a low or no pitch accent, $F0$ is increased only slightly following all obstruents. It is suggested that this difference occurs because demands of the segmental and prosodic levels of speech production conflict. In particular, results support a theory that the primary feature of voiced or voiceless obstruents is [.-stiff] or [+stiff] vocal folds, respectively. Enhancing features such as [-.spread] or [+spread] vocal folds are secondary. In a high pitch environment, the feature [-.stiff] conflicts with the need to raise pitch. Because of enhancing gestures, the prosodic level can override the segmental. Likewise, in a low pitch environment, the feature [+stiff] conflicts, so the vocal folds will not be stiffened for voiceless obstruents. If so, one might expect that stop consonants in low pitch environments will have longer or stronger prevoicing than in high pitch environments, and preliminary data show that speakers do show this tendency. [Work supported by NIH Grant DC04331.]

2pSC11. Acoustic characteristics of whispered vowels. Michael Kiefte (School of Human Commun. Disord., Dalhousie Univ., Halifax, NS B2Y 1P9, Canada)

It is well known that whispered speech is able to convey information that is normally associated with pitch. For example, it is possible to whisper the question ''You are going today?'' without any syntactic information to distinguish this sentence from a simple declarative. It has been shown that pitch change in whispered speech is correlated with the simultaneous raising or lowering of several formants [e.g., Kallail and Emanuel, J. Speech Hear. Res. **27**, 245–251 (1984)]. Data will be presented from 81 native speakers of English from the Halifax region of Nova Scotia (35 men and 46 women) who were asked to phonate and whisper the vowels /i,ɪ,e,ɛ,æ,ɝ,o,ʊ,u,ʌ,ɚ,ɔɪ,aʊ,aɪ/ at three different pitches across a range of roughly a musical 5th. Formant frequency variability is much greater for whispered vowels with different intended pitches resulting in much greater between-category overlap. Listeners' categorizations of these stimuli will be reported as well as results from a discriminant analysis based on either static or dynamic spectral information. [Work supported by SSHRC.]

2pSC12. Perception of source spectral slope. Jody Kreiman and Bruce R. Gerratt (Head/Neck Surgery, UCLA School of Medicine, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794)

Researchers broadly agree that the spectral slope of the voice source is an important concomitant of voice quality, but it is unclear which specific aspects of spectral slope are perceptually important. To examine this issue, 50 voice samples were inverse filtered, and a large variety of measures of spectral slope were calculated for the resulting source pulses. Factor analysis was applied to determine which of these measures of slope are in fact independent. Synthetic copies of the voices were generated. Listeners used a method of adjustment task to manipulate these independent aspects of spectral slope in the synthesizer so that the synthetic copies matched the natural voice samples. Patterns of listener agreement and variability provide information about the perceptual relevance of different acoustic measures of source spectral slope. [Research supported by NIDCD.]

2pSC13. Effects of spectral tilt on the perception of naturally spoken vowels. James M. Hillenbrand (Speech and Hearing Ctr., MS5355, Western Michigan Univ., Kalamazoo, MI 49008)

Experiments using synthetic signals with static spectral patterns have shown that spectral tilt can have an influence on vowel identity that is independent of the formant-frequency pattern [e.g., M. Kiefte and K. Kluender, J. Acoust. Soc. Am. **109**, 2294–2295 (2001)]. The purpose of this experiment was to determine whether spectral tilt affects vowel identity for naturally spoken utterances. Test signals consisted of filtered and unfiltered versions of 300 /hVd/ utterances selected from a large, multi-talker database [Hillenbrand *et al.*, J. Acoust. Soc. Am. **97**, 3099–3111 (1995)]. Phonetically trained listeners ($N=24$) identified the unfiltered signals and the same signals conditioned by a +9 dB/oct filter and a −9 dB/oct filter, presented in a single random order. The effects of the two filters were clearly audible to listeners, but there was no effect of spectral tilt on vowel

recognition accuracy. These findings do not invalidate earlier results with synthetic speech; rather, they suggest that when natural cues to vowel identity are preserved, overall spectral tilt plays little if any role in vowel recognition (within the limits of the ±9 dB tilt manipulation). The findings would also seem to present a significant challenge to whole-spectrum models of vowel perception, though they are by no means incompatible with this approach. [Work supported by NIH.]

2pSC14. The Davis addendum to the Tomatis effect. Dorinne S. Davis-Kalugin (Davis Ctrs., Inc., 98 Rt. 46 W, Budd Lake, NJ 07828, davis@thedaviscenter.com)

The relationship of the ear and voice is defined in three laws, known as the Tomatis effect. Specifically, these laws state: (1) The voice only contains the harmonics that the ear can hear. (2) If you give the possibility to the ear to correctly hear the distorted frequencies of sound that are not well heard, these are immediately and unconsciously restored into the voice. (3) The imposed audition sufficiently maintained over time results in permanently modifying the audition and phonation. Using a time domain frequency analysis of the voice, the spontaneous otoacoustic emissions from the ear were evaluated and compared with a frequency analysis obtained through voice analysis. One hundred percent correlation between the stressed frequencies of the ear and voice was noted. An addendum to the Tomatis Effect describes these effects. First, the ear emits the same stressed frequencies that are emitted by the voice. Second, when complementary or supplementary frequencies of stressed frequencies are introduced via sound vibration to the ear, vocal patterns regain coherence.

2pSC15. A comparison of three algorithms for estimating aspiration noise in dysphonic voices. Mina Goor (Dept. of Elec. and Computer Eng., Univ. of Florida, Gainesville, FL 32611), Rahul Shrivastav (Univ. of Florida, Gainesville, FL 32611), and John G. Harris (Univ. of Florida, Gainesville, FL)

Dysphonic voices are frequently characterized by increased aspiration noise. Several algorithms have been proposed to quantify the noise present in such voices. Yet, an independent analysis of these algorithms has not been reported. These algorithms differ in a number of aspects such as the theory underlying these measurements, the procedures used for estimating the noise, and the nature of their output. Three different algorithms for estimating noise in dysphonic voices were implemented in MATLAB and their output for synthetic and natural voice samples compared. These algorithms include: (1) the pitch-predictive signal-to-noise ratio reported by Milenkovic [Workshop on Acoustic Voice Analysis: Proceedings (1994)], which analyzes signals in the time domain and provides a time waveform of the aspiration noise: (2) the harmonics-to-noise ratio reported by deKrom [J. Speech Hear Res., **36**, 254–266 (1993)], which performs a cepstral analysis to provide the spectrum of the aspiration noise; and (3) the glottal-noise excitation reported by Michaelis, Frohlich, and Strube [J. Acoust. Soc. Am. **103**, 1628–1639 (1998)], which measures the correlation of the Hilbert transform of different frequency bands. The result of the study will help identify the algorithm(s) most suitable in the prediction of the listener judgments of voice quality. [Research supported by NIH/R21DC006690-01.]

TUESDAY AFTERNOON, 16 NOVEMBER 2004 PACIFIC SALON 3, 1:00 TO 3:00 P.M.

Session 2pSP

Signal Processing in Acoustics and Underwater Acoustics: Signal Processing Arrays with Many Elements in Novel Configurations or Novel Environments Part II

David I. Havelock, Cochair
National Research Council, IMS/ASP, Montreal Road, Ottawa, Ontario, K1A 0R6, Canada

Jens M. Meyer, Cochair
20 River Terrace, New York, New York 10282

Contributed Papers

1:00

2pSP1. Simulation of synchronized animal calling with a distributed sensor network. Efosa Ojomo, Praveen Mudindi, Isaac Amundson, and Kenneth D. Frampton (Dept. of Mech. Eng., Vanderbilt Univ., Nashville, TN 37235)

A distributed sensor network has been programmed to simulate synchronized calling exhibited by many species of frogs and insects. The distributed sensor network consists of numerous sensor nodes consisting of a microprocessor, wireless communications and a sensor board containing a buzzer and microphone. The goal was to program the sensor array to mimic the synchronized calling behavior (inspired in no small part by the recent appearance of cicada brood X). To this end, a single leader node was programmed to begin calling (i.e., buzzing). The remainder of the node array was programmed with an algorithm which allows them to detect other buzzing nodes and to synchronize their own buzz to that of the group. After start-up transients, the entire array calls in synchrony. Of particular interest is the system's transient behavior system when perturbed or when a second leader node begins calling out of synch with the array. The hardware and software algorithms will be described. Furthermore, numerous nodes will be distributed throughout the audience and a demonstration of system behavior will be provided. While this is a rather whimsical application of distributed array processing, it does demonstrate the unique system behaviors that can arise in truly distributed processing.

1:15

2pSP2. Optimized multistatic sonobuoy array patterns. Donald R. DelBalzo (Neptune Sci., Inc., 40201 Hwy., 190 E, Slidell, LA 70461, delbalzo@neptunesci.com), David P. Kierstead (Daniel H. Wagner, Assoc., Vienna, VA 22180), and Erik R. Rike (Neptune Sci., Inc., Slidell, LA 70461)

Standard patterns for monostatic sonobuoy fields were developed during the Cold War for deep, uniform underwater environments, where a simple median detection range defined a fixed spacing between sonobuoys (usually along staggered lines). Oceanographic and acoustic conditions in littoral environments are so complex and dynamic that spatial and temporal variability of low-frequency signal and noise fields destroys the basic homogeneous assumption associated with standard tactical search concepts. Genetic algorithms (GAs) have been applied to this problem to produce near-optimal, nonstandard search tracks for monostatic mobile

sensors that maximize probability of detection in such inhomogeneous environments. For the present work, a new capability, SCOUT (sensor coordination for optimal utilization and tactics) was developed to simulate multistatic distributed-sensor geometries and to optimize the locations of multistatic active sonobuoys in a complex, littoral environment. Both standard and SCOUT-derived tactics were evaluated for cumulative detection probability and compared. The results show (a) that the standard pattern is not optimal even for a homogeneous environment and (b) that standard patterns are grossly ineffective in inhomogeneous environments. [Work was sponsored by NAVAIR.]

1:30

2pSP3. Performance analysis of angle-or-arrival techniques in shallow water bathymetry measurements. Wen Xu, Marc Parent, and Fran Rowe (RD Instruments, 9855 Businesspart Ave, San Diego, CA 92131, wxu@edinstruments.com)

Conventional bathymetric side scan system measures the phase difference between two parallel rows of transducers and converts it to bottom echo signal direction and then depth. A main concern is that the phase measurements are easily dispersed by noises and interferences, particularly in shallow water applications. There have been considerable efforts to deal with the problem, for example, by adding more transducers and implementing signal angle-of-arrival estimation. In this work, experimental data anlaysis for a multiple-row system is conducted, disclosing shallow water echo signal spatial structure. In addition to a stable bottom return, there often exist returns from surface scattering or multi-path reflecting between surface and bottom, more spatially-dispersed depending on the surface wave condition. By comparing the differential phase approach, which is blind to interferences, to the angle-of-arrival approach, which exploits the multi-path signal structure, it is concluded that the differential phase approach can yield systematic bias toward surface in bottom depth measurements. Performance of the angle-or-arrival technique is then investigated. The Cramer-Rao bound is derived and evaluated in the presence of two signal sources, followed by discussions on source separation (resolution) and individual source strengths.

1:45

2pSP4. Low-frequency passive performance of a towed volumetric array during maneuvers. Jerrold Dietz, James Edgerton, Bruce Newhall, Juan Arvelo, Jr., and Catherine Frazier (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099)

For a large towed volumetric array, array element deformations significantly degrade system performance both during tow-ship maneuvers as well as during straight tows. Using a least squares fit for the shape estimation sensors we can generate accurate array shapes that allow us to maintain signal gain through array maneuvers. Array signal gain was calculated throughout maneuvers and is observed against measures of the array distortion. These calculations are compared for a linear, planar and volumetric array. We will discuss the experimental results in addition to environmental modeling to determine the best achievable gains. [This work was supported by the Defense Advanced Research Project Agency (DARPA) under the Robust Passive Sonar (RPS) project.]

2:00

2pSP5. Boundary array size for teleconferencing. Dwight F. Macomber (Belar Electron. Lab., Inc., Devon, PA)

Boundary arrays for teleconferencing require large numbers of sensors to be effective. The paper describes efforts to estimate the number of microphones required for an effective teleconference boundary array. Measurements of many source-to-sensor room transfer functions (RTF) were made in a conference room of moderate size. The direct arrivals of these RTFs were time aligned in software to estimate the transfer function of a delay-and-sum boundary array. Using the image method of Allen and Berkley, another set of RTFs was generated for a virtual room acoustically similar to the measured room. These RTFs were time aligned to create a second virtual array. The S/Rs of these two virtual arrays compare favorably. The image method was used to estimate transfer functions of virtual boundary arrays of different sizes for a 60-cubic-meter room. Standard beamforming and matched-filter processing were used. TIMIT speech was convolved with the virtual array transfer functions to generate audio used for evaluation. A listening test ranked the auditory performance of four single-microphone reference systems and eight test arrays in the virtual room. For rooms of roughly 100 cubic meters, boundary arrays with approximately 100 microphones may provide good subjective performance.

2:15

2pSP6. Acoustic array data compression via Karhunen-Loeve transform. Frank A. Boyle and Thomas H. Phipps (Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029, boyle@arlut.utexas.edu)

Acoustic data compression has been explored in several contexts and several techniques exist. The applicability of each technique depends on the type of data processing that is intended. For example, audio data is often compressed via perceptual coding methods (e.g., mp3) in which quantization noise is distributed according to psychoacoustic principles. The Karhunen-Loeve (KL) transform presents an opportunity to compress surveillance array hydrophone-level data while preserving relevant features in beamformed displays. A KL codec was formulated and applied to acoustic test data from a horizontal line array. The results appear promising in that features of interest are preserved with significant data compression. The presentation will include a description of the algorithm as well as examples with actual data. [Work funded by ONR.]

2:30

2pSP7. Nonlinear weighting techniques to replace tradeoffs between mainlobe width and sidelobe suppression. Dale B. Dalrymple (Signal Processing Systems Div. of Information Systems Labs., 10070 Barnes Canyon Rd., San Diego, CA 92121, ddalrymple@islinc.com)

When the discrete Fourier transform is used to transform data from one domain to another as in narrowband beamforming or spectral analysis a weighting function can be applied as a vector multiplication before the transform or as a convolution after the transform to control mainlobe width in opposition to sidelobe suppression. Nonlinear operations applied to transform outputs or weighted versions of the transform outputs can combine mainlobe width control with sidelobe suppression. Applying weights before the transform requires the execution of multiple transforms for this process. The use of small kernals in the frequency domain on the output of a single transform provides a more efficient implementation. The set of small kernal summed cosine weights provides a variety of responses that can be combined nonlinearly retaining the best traits of each. Kernals can also be calculated adaptively.

2:45

2pSP8. Applications of beamforming and acoustic holography in an anechoic tank with surface reflections. Sea-Moon Kim, Youngchol Choi, and Yong-Kon Lim (Yuseong, P.O. Box 23, Daejeon 305-600, Korea)

Recently, our institute, KRISO/KORDI, constructed a small anechoic tank which has an anechoic lining at the four walls and bottom. Because surface reflections still occur, special care must be taken for acoustic measurements and array signal processing, for example, source localization and acoustic holography. In this paper the effects of the surface reflections to the array signal processing methods are investigated. First, estimated errors by both methods are analyzed for various measurement parameters. Beamforming experiments with vertical line arrays are performed. Holographic reconstruction of simple and complex acoustic sources is also performed to compare the true and estimated results.

Session 2pUW

Underwater Acoustics: Propagation and Modeling

Kevin B. Smith, Chair
Department of Physics, Naval Postgraduate School, Code PH/SK, Monterey, California 93943

Contributed Papers

1:00

2pUW1. Beam summation algorithm for wave radiation and propagation in stratified media. Tal Heilpern and Ehud Heyman (School of Elec. Eng., Tel Aviv Univ., Tel Aviv 69978, Israel, heyman@eng.tau.ac.il)

An efficient Gaussian beams summation (GBS) algorithm for tracking source excited wavefields in plane stratified media is introduced. It has two important features: (a) it involves an efficient calculation of the GB propagators, and (b) it involves rather sparse lattice of beams. For (a) we approximate the medium using layers with constant wavespeed gradient, and derive an efficient recursive algorithm for tracking the GB through such medium. This model not only reduces the number of layers, and thereby the algorithm complexity, in comparison with the conventional uniform layers model, but it also eliminates the reflection artifacts at layers interfaces. Property (b) is achieved by determining the beam expansion parameters for an efficient discretization of the source-excited beam spectra using a sparse lattice of beams. The algorithm has been validated and calibrated via thorough numerical comparisons with closed form ray solutions for source-excited fields in layered media. Perfect agreement between these independent solutions has been obtained in regions where the ray solution is valid, but the beam formulation also provided smooth and physically meaningful solutions in caustic regions where the ray solution fails.

1:15

2pUW2. Parabolic propagation in a weakly range-dependent duct: Approaching higher orders systematically. Robert F. Gragg (Naval Res. Lab., Code 7140, Washington, DC 20375, robert.gragg@nrl.navy.mil)

This work illustrates a technique that exploits energy-conserving transformations to split a CW field into a pair of components that propagate via uncoupled parabolic equations in opposite directions along the axis of a weakly inhomogeneous waveguide. A systematic series of these transformations produces this splitting at successively higher orders while avoiding backscatter. In the interest of clarity, the simplest possible nontrivial case is considered: waves of vertical displacement on a stretched membrane with a smooth density inhomogeneity along the y direction that forms a duct in the x direction. (This case is truly two-dimensional and its only environmental variable, density, is continuous.) This transformation technique provides an efficient means of incorporating the effects of weak environmental inhomogeneity in higher-order parabolic propagation. [Work supported by ONR.]

1:30

2pUW3. Modeling acoustic nonlinearities in marine sediments. B. Edward McDonald (Naval Res. Lab, Code 7145, Washington, DC 20375)

Investigations into the theoretical and numerical treatment of acoustic nonlinearities in marine sediments are motivated by mine countermeasures and other Naval interests related to explosions in shallow water. Agreement between theory and data for the nonlinearity parameter B/A has been demonstrated in early papers concerning saturated sands. In the more general case involving varying depth profiles of air and silt contents, however, it is observed that B/A values vary so greatly that deterministic nonlinear modeling approaches may have to be augmented to address environmental uncertainty. Comparisons are given between B/A values derived from static consolidation tests and available *in situ* data. Numerical profiles from the NPE shock propagation model are given to demonstrate the effects of environmental uncertainty. [Work supported by the ONR.]

1:45

2pUW4. A single-scattering solution that handles large contrasts across interfaces. Elizabeth T. Kusel, William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180), and Michael D. Collins (Naval Res. Lab., Washington, DC 20375)

Single-scattering and energy-conservation approximations have both proven to be effective for solving range-dependent ocean acoustics problems. The energy-conservation approach is usually used in acoustic models since it provides greater efficiency. The single-scattering approach is more general and more promising for problems involving elastic layers. A range-dependent medium is approximated in terms of a series of range-independent regions separated by vertical interfaces. The single-scattering solution is usually implemented in terms of an iteration formula, which may diverge when there is a large contrast in the material properties across an interface. Convergence can be improved by artificially splitting a vertical interface into a series of slices with smaller contrasts. For the idealized problem of scattering from a single stair step, a significant error may be introduced by neglecting multiple scattering between slices. However, the approach should be accurate for problems involving gradual range dependence, such as sloping interfaces. [Work supported by the Office of Naval Research.]

2:00

2pUW5. Coupled perturbed modes for a wedge waveguide. Chris J. Higham and Chris T. Tindle (Phys. Dept., Univ. of Auckland, Private Bag 92019, Auckland, New Zealand)

A transformation of the normal modes is described to accommodate the density jump found in range-dependent penetrable bottom problems. The transformation is necessary for the application of perturbation theory to a coupled mode model of a wedge waveguide. The combination of conventional coupled modes and perturbation theory yields an efficient modal theory for shallow water acoustics of range dependence. The method is applied to upslope propagation in the ASA benchmark 2 wedge [J. Acoust. Soc. Am. **87**, 1499–1510 (1990)].

2:15

2pUW6. Acoustical propagation modeling using the three-dimensional parabolic equation based code 3DWAPE within a multiprocessing environment. Kaelig Castor (CEA/DASE, BP 12, FR-91 680 Bruyeres-le-Chatel, France), Frédéric Sturm (Laboratoire de Mecanique des Fluides et d'Acoustique, 69134 Ecully Cedex, France), and Pierre Franck Piserchia (CEA/DASE, FR-91 680 Bruyeres-le-Chatel, France)

In some particular oceanic environments involving bathymetric slopes and horizontal sound speed gradients, the azimuthal coupling can be significant. Fully three-dimensional models are thus needed to account for horizontal refraction. These models need usually high computational re-

sources, especially for broadband calculations and/or for long-range paths where the 3-D effects are clearly accentuated. In this study, numerical simulations using the 3-D parabolic equation model 3DWAPE [F. Sturm, Acust. Acta Acust. **88**, 714–717 (2002)] are presented. The calculations are performed on a massively parallel computer providing a high computational efficiency. The message-passing interface (MPI) communication library is used. Two parallelization levels are considered. The first one allows a broadband-signal propagation by distributing independently on different processors the calculations for each frequency. The second one is based on a segmentation of the propagation matrices. Simple methods are used to perform an optimized multiprocessor implementation by equalizing processor workload, and split the number of processors between the two parallelization levels. An analysis of both speedup and efficiency of the algorithm is presented for several configurations. Computational time comparisons are shown for the 3-D ASA benchmark. The algorithm can also be applied to a realistic environment involving sound speed profiles and bathymetry data sets.

2:30

2pUW7. Sound propagation in shallow water with surface waves. Chris T. Tindle (Phys. Dept., Univ. of Auckland, Auckland, New Zealand) and Grant B. Deane (Scripps Inst. of Oceanogr., San Diego, CA)

Wavefront modeling can be used to calculate acoustic pulse propagation in shallow water with real surface waves. The method uses ray tracing but phase and amplitude are found from approximations to a phase integral. The lowest order approximation is the conventional geometric ray. Higher order approximations describe caustics, shadow zones and focusing. Reflection from sloping curved boundaries also leads to a range dependent horizontal wave number and requires modification of the phase integral and phase derivatives. Results are in good agreement with an experiment in which 10-kHz short pulses were transmitted in shallow water with 0.8-m peak-to-trough surface waves and a sloping bottom. [Work supported by ONR.]

2:45

2pUW8. A new primitive ray-tracing (PRT) model to compute forward and inverse acoustic propagation in a two-dimensional, range-dependent, sound-speed field. Sergey V. Vinogradov, Jerald W. Caruthers, Hans E. Ngodock (Dept. of Marine Sci., Univ. of Southern MS, 1020 Balch Blvd., Stennis Space Ctr., MS 39529), and Natalia A. Sidorovskaia (Univ. of Louisiana-Lafayette, Lafayette, LA 70504)

A primitive ray-tracing (PRT) model has been developed as part of a computational system to support potential tomographic observations. It is intended to estimate the variability in acoustic arrivals due to mesoscale oceanic circulation. The underlying algorithm is fully range dependent; it utilizes both vertical and downrange sound-speed gradients to compute ray trajectories. The ''primitiveness'' of the model means that it is simple enough to be inverted, while most existing forward acoustic models are too sophisticated computationally for that purpose. The inversion of this PRT takes in travel times and estimates possible oceanic structure, which further is being assimilated into an ocean circulation model. This paper provides details and results of acoustic computational experiments, along with a description of the PRT's application in a postulated tomographic observational system for the northeastern Gulf of Mexico.

3:00–3:15 Break

3:15

2pUW9. High-accuracy absorbing boundary conditions for high-order parabolic wave equations. Murthy N. Guddati and A. Homayoun Heidari (Dept. of Civil Eng., North Carolina State Univ., Campus Box 7908, Raleigh, NC 27695-7908)

Using a new systematic approach, a novel set of arbitrarily high-accuracy absorbing boundary conditions (HABC) is introduced for high-order parabolic equations. HABC is derived by finite-element discretization of the boundary, followed by applying a special integration scheme and imaginary stretching. The resulting boundary condition is then coupled with the interior and can be solved with the same numerical method as the interior, e.g., finite differences. The accuracy of the HABC is controlled by two parameters: (1) the number of absorbing layers (the order of the HABC); and (2) a reference phase velocity for each layer. The latter controls the wave number for which the HABC absorbs the wave field exactly. Hence, the absorption range, i.e., the distribution of the reflection coefficient, can be controlled by the coefficient matrices of the HABC. A specially designed explicit finite-difference scheme is used to solve the HABC. The added computational cost due to the HABC is negligible for practical purposes. The performance of the proposed boundary condition is shown through numerical examples for high-order parabolic equations, for different orders of the HABC and different sets of phase velocity parameters. Effective absorption of both propagating and evanescent waves is illustrated in the results. [Work supported by NSF.]

3:30

2pUW10. Fluctuations in sound transmission through velocity profiles that are periodic in range and in geo-time. Jacob George and Robert L. Field (NRL Code 7185, Stennis Space Ctr., MS 39529)

The lunar M2 tides cause variations in sound velocity profiles that are periodic in range and in geo-time. Acoustic transmissions through such an environment exhibit fluctuations at the primary tidal frequency as well as its overtones, due to nonlinear dependence on sound speeds. The number of tidal cycles within the source-receiver range is found to be a major factor that determines the magnitude and nature of the fluctuations. This has been investigated using parabolic equation (PE) models and by modal analysis. In the WKB model the important phase factor consists of a range integral of the horizontal wave number of each mode. A study of this integral has revealed a dramatic decrease in fluctuations when the number of tidal cycles within a constant source-receiver range increases. This result directly impacts transmissions that are at variable angles measured from the direction of the tides. Pulse transmission through the tidal environment has also been investigated, and the results are similar to those described above. [Work supported by ONR.]

3:45

2pUW11. Modeling the acoustical focusing properties of shoaling surf. Grant Deane (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238) and Chris Tindle (Univ. Auckland, New Zealand)

Shoaling surf focuses sound, creating high intensity caustics that impact the performance of underwater communications systems and sonars in the very near shore region. Simple expressions for the time-varying caustic locus and amplitude can be derived from the Kirchhoff–Helmholtz scattering integral. The analysis shows that a focus is formed at a range equal to one-half the crests radius of curvature when the wave is directly over a shallow water source, and moves toward the crest and decreases in intensity as the wave shoals shoreward. The analysis is compared with numerical scattering calculations based on measured surface gravity wave profiles, and also with experimental data in which 10-kHz, single cycle pulses were transmitted over a 40 m path in 6 m of water just north of Scripps Pier. The impact of the caustics on sonar performance will be discussed. [Work supported by ONR.]

4:00

2pUW12. Time-domain sound propagation through a ship wake. Xiao Di, Lee Culver, and David Bradley (Appl. Res. Lab, The Penn State Univ., University Park, PA 16802)

A time-series model has been developed to visualize high-frequency pulses propagating through the bubbly wake of a ship. The wakes can have very large void fractions, and thus strong inhomogeneities in sound speed, scattering cross section, and acoustic attenuation. The model uses the Green's function PE to calculate the pressure field at frequencies within the bandwidth of the acoustic pulse, followed by Fourier synthesis to calculate time-domain solutions over a range of times. Special care was required to preserve the phase of the pressure at each frequency. Our

concern is with quasihorizontal propagation and source–receiver ranges from about 100 to 1000 m, which contain effects of bubbles on propagation including attenuation, refraction, and scattering. We show predictions using a simple model for the distribution of bubbles in a ship wake. We find a very interesting frequency dependence for refraction and scattering. [Work supported by ONR Code 321.]

4:15

2pUW13. A geometrical model for surface ship wake. Boris Vasiliev (DRDC-Atlantic, 9 Grove St., Dartmouth, NS B2Y 3Z7, Canada, boris.vasiliev@drdc-rddc.gc.ca)

High-frequency geometrical measurements of a surface ship wake collected by other research groups will be presented. The general trends in the width data agree with the previously published measurements; however, the trends in the depth data differ from those published in the literature. On-going efforts in data analysis to develop a high-frequency surface wake model that accounts for wake width, depth and persistence will be discussed.

4:30

2pUW14. Identifying modes that produce late arrivals for low-frequency long-range propagation in shallow water. Harry DeFerrari, Irina Rypina, and Ilya Udovydchenkov (RSMAS, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

Model and data comparisons are presented for acoustic propagation over a frequency range from 100 to 3200 Hz ($D=145$ m, $R=10$ km). Above 0.8 kHz most all energy is reflected from the water–sediment interface. Arrival times of waterborne paths are in close agreement with normal mode and PE predictions when observed sound speed profiles are used as inputs. Generally, the received signal can be closely approximately with $f/50$ modes since steeper SRBR modes are attenuated. For frequencies below 0.5 kHz, arrivals from waterborne paths are still present, but an additional group of later arrivals is also observed. The timing of these arrivals suggests either late SRBR or more complicated modes from energy passing well into the bottom. Unfortunately, geoacoustic properties of the subbottom are not known for depth greater than a few meters. Bottom properties at depth are assumed and then propagation models are used to determine the modes that most likely produce the late arrivals. Some modes having group velocities that correspond to the arrival time of the late arrivals penetrate a few tens of meters into the bottom. Coherence in time and space is computed and compared for bottom penetrating and waterborne paths.

4:45

2pUW15. Probability density function methods for uncertainty analysis in underwater acoustics. Kevin R. James and David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, 2019 Lay Auto Lab., 1231 Beal Ave., Ann Arbor, MI 48109)

Forward modeling of underwater acoustic propagation is generally successful when environmental parameters and boundary conditions are known. Unfortunately, such information is seldom available at the requisite level of precision, and any imprecision introduces uncertainty into sound field predictions. Quantifying this sound-field uncertainty is important for applications of acoustic propagation models such as matched-field processing. This presentation describes a method for quantifying the underwater-sound-field uncertainty arising from imperfect knowledge of the environment and its boundaries. It is based on formulating and solving a transport equation for the joint probability density function (PDF) of the real and imaginary parts of a harmonic sound field. The appropriate equation is obtained by combining spatial derivatives of the PDF with physical laws drawn from guided wave mechanics. The inputs for solving the PDF-transport equation are known or assumed distributions of the uncertain parameters. Solutions can be readily reduced to expected values, uncertainties, and confidence intervals for the predicted sound field. Results for simple test cases involving range-independent isospeed underwater sound channels are considered, and compared to solutions obtained analytically or through Monte Carlo simulations. [Work sponsored by ONR.]

Session 3aAA

Architectural Acoustics: Coupled Volume Systems: Design Process and Implications

David T. Bradley, Cochair
Architectural Engineering, University of Nebraska-Lincoln, 1110 South 67th Street, Omaha, Nebraska 68182-0681

Ralph T. Muehleisen, Cochair
Civil and Architectural Engineering, Illinois Institute of Technology, 3201 South Dearborn, Chicago, Illinois 60616

Chair's Introduction—8:30

Invited Papers

8:35

3aAA1. Comparison of real world measurements and computer model results for a dedicated coupled volume system. David T. Bradley and Lily M. Wang (Architectural Eng., Univ. of Nebraska—Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681, dbradley@mail.unomaha.edu)

A dedicated coupled volume system consists of a larger main space connected through an opening, or coupling aperture, with a smaller auxiliary space specifically designed to produce a double slope effect (DSE). DSE can be achieved if the absorption characteristics in the secondary space are such that its decay time is longer than that of the main space, which in turn causes late-arriving energy to be fed back into the main space. This project focuses on the verification of computer modeling as an accurate method for analyzing dedicated coupled volume systems. Impulse response measurements were taken in a real world coupled volume hall. The amount of absorption in the main hall and the number of open apertures were systematically varied to create a series of configurations for which the impulse responses were measured. The same configurations of the hall were modeled in the computer aided acoustical modeling program, ODEON. Comparisons of the impulse responses from the real world measurements and the computer model are presented both qualitatively and quantitatively through examining pertinent objective acoustical measures, such as reverberation times (T30, T15) and clarity index (C80).

8:55

3aAA2. Scale-model studies of reverberant energy decay in a coupled-room system. Jason E. Summers[a], Rendell R. Torres, and Yasushi Shimizu[b] (Program in Architectural Acoust., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180)

Energy decay at high, middle, and low frequencies is measured in a reconfigurable 1:10 scale model of a two-room coupled system schematically representing a coupled-room concert hall. Measurements are used both to phenomenologically characterize decay behavior in each frequency range and to validate analytical and computational models. At high frequencies, decay behavior is characteristic of energy exchange between diffuse fields. Comparisons of computational geometrical-acoustics predictions based on randomized beam-axis tracing with scale-model measurements indicate errors resulting from tail-correction assuming constant quadratic growth of reflection density. Using ray tracing in the late part in place of tail correction is shown to correct this error. At midfrequencies, scattering at apertures results in dependence of decay behavior on aperture shape. High-frequency geometrical-acoustics and statistical-acoustics models are modified to account for wave effects at coupling apertures by including power transmission coefficients of apertures [J. E. Summers *et al.*, J. Acoust. Soc. Am. **112**, 2226 (A) (2002)]. Scale-model measurements evidence behavior predicted by these models. At low frequencies, coupling strength is related to unperturbed pressure distribution on coupling surfaces. Perturbation models [C. M. Harris and H. Feshbach, J. Acoust. Soc. Am. **22**, 572–578 (1950)] qualitatively explain these effects. [Work supported by the Bass Foundation.] [a] Current address: Acoustics Division, Naval Research Laboratory, Washington, DC 20375-5350.

9:15

3aAA3. Five sensitivities of the coupled volume/double sloped system. Michael Ermann (Virginia Tech School of Architecture + Design, 201 Cowgill Hall (0205), Blacksburg, VA 24061, mermann@vt.edu)

Stop chord measurements taken in an existing coupled volume concert hall, established statistically-based formulas, ray-tracing-software simulations, and headphone listening perception/preference studies imply a fickle relationship between the coupled volume concert hall, the double-sloped sound decay, and the listener. The two halls examined in this study suggest five sensitivities of the system: (1) To produce a double sloped sound decay, the coupled volume must be exceedingly more reverberant than the main hall to which it is attached. (2) To produce a double sloped sound decay, the sonic aperture linking the coupled volume to the main hall must be exceedingly small. (3) Maintaining low levels of background noise—always essential when designing spaces for music listening—is especially important if the double sloped impulse response is to be perceived above the noise floor. (4) Listeners who sit near the apertures probably hear something different than those who sit nearer to the center of the hall. (5) It is unclear whether subjects taking a paired comparison test prefer a double sloped auralization to a classic Sabine auralization.

9:35

3aAA4. Decay time uncertainty analysis in acoustically coupled spaces. Ning Xiang, Mendel Kleiner, and Tomislow Jasa (School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180)

In order to guide acoustic design and to verify the design goals, evaluations of multiple decay times in coupled spaces are of practical significance when designing coupled volumes in performing art spaces. Traditionally, however, identification of double- or multiple-sloped decay in room impulse response measurement has been considered very challenging. Previous works [Xiang and Goggans, J. Acoust. Soc. Am., **110**, 1415–1424 (2001); **113**, 2685–2697 (2003)] have applied Bayesian framework to cope with the demanding tasks in estimating multiple decay times from Schroeder decay functions, evaluated from measured room impulse responses. The Bayesian approaches have also been applied in systematic investigation of double-slope decay in coupled volumes with respect to volume ratio, absorption ratio, and aperture size [Bradley and Wang, J. Acoust. Soc. Am. **113**, 2188 (2003) (A)]. These recent applications of Bayesian approaches stimulate further study of uncertainties associated with decay time estimation. This paper discusses the estimation uncertainties and interrelationships between multiple decay times within Bayesian framework.

9:55

3aAA5. Measurement and subjective evaluation of the spatial modulation of late reverberation. Durand R. Begault (MS 262-2, NASA Ames Res. Ctr., Moffett Field, CA 94035, durand.r.begault@nasa.gov)

The acoustical characterization of a large rectangular space with uniformly distributed absorption is facilitated by a predictably linear reverberant decay of decibel sound pressure level. By contrast, large spaces with irregular shapes yield correspondingly irregular reverberant decays. A contributing factor is coupled space effects where the several reverberant decays occur simultaneously. When the mean free path for reflections is relatively large, as in a cathedral, temporal thresholds can be exceeded to the point where it is possible to hear spatially separated decays. The phenomenon of a slow amplitude modulation between separate locations can be described as moving late reverberation, which is audible, e.g., in a cathedral when an organ note is stopped. Measurements of this phenomenon were made in San Franciscos Grace Cathedral, a French Gothic-style Episcopal church, using multiple calibrated microphone impulse response recordings. Data on the temporal-spatial diffusion of reverberation from various spatial perspectives are presented, along with sound examples. Preliminary listening test data involving subjective evaluation of late reverberation movement is presented, where the temporal and spatial components of a synthetic impulse response is manipulated in a manner characteristic of similar large spaces.

10:15–10:25 Break

10:25

3aAA6. Coupled volumes in concert hall: Not just reverberation chambers. Damian Doria, Russell Johnson, Todd L. Brooks, and Ted Pyper (Artec Consultants Inc., 114 West 26 St., Fl. 12, New York, NY 10001)

Artec Consultants Inc. has incorporated coupled volumes into the design of purpose-built concert hall spaces for over three decades. Initial interest in coupled volumes for concert spaces focused on extending reverberation time and assisting loudness balance for brass and strings, but the geometric configuration and incorporation of other acoustics design innovations in Artec projects has varied not only reverberation time but also several other primary parameters. This paper reviews approaches to incorporating coupled volumes in concert halls and comments on the artistic value of their ability to instantly tailor the acoustics of a concert space. Artec's halls in Dallas, Lucerne, São Paulo, and Philadelphia are discussed in the context of their use of coupled volumes to enhance the performance environment for symphony and other performances.

Contributed Papers

10:45

3aAA7. Reverberated sound field modeling in coupled rooms using a diffusion equation. Alexis Billon, Vincent Valeau, Anas Sakout (LEPTAB, Univ. of La Rochelle, Av. M. Crpeau 17042, La Rochelle Cedex 01, France, abillon@univ-lr.fr), and Judical Picaut (LCPC Nantes, Bouguenais Cedex France)

Sound modeling in coupled rooms has attracted considerable attention in the past, but accurate and operational models are still needed. In recent papers, a diffusion equation based model has been applied with success to unusual room shapes. This approach allows nonuniform repartition of energy, and is especially relevant in room acoustics for long rooms or complex spaces such as networks of rooms. The present work aims at validating the behavior of the diffusion model in the case of two acoustically coupled rooms separated by an open area. In this purpose, the time-dependent diffusion equation is solved in three dimensions using a finite-element solver. It allows one to predict both sound attenuation and reverberation time at any point of the coupled rooms. Parameters influencing the coupling, i.e., room relative sizes, aperture size and room absorption areas, are investigated. Results are then compared with the classical statistical theory of coupled rooms. Finally, a comparison with experimental data in two coupled classrooms is also provided.

11:00

3aAA8. Searching for double-slope behavior in a coupled-volume concert hall. Christopher Storch (Cavanaugh Tocci Assoc., Inc., 327 F Boston Post Rd., Sudbury, MA 02478)

Impulse response measurements were recently conducted in a coupled-volume concert hall in an attempt to discover any double-slope behavior in the reverberant decay, a characteristic that had been subjectively heard in concerts. In the unoccupied hall, measurements were conducted in the primary volume with the secondary volume decoupled, 50% coupled, and 100% coupled. Results of these measurements will be presented. Speculation as to why the reverberant decays do not agree with subjective listening impressions will be discussed.

11:15

3aAA9. Functional-differential-equation models of reverberant energy decay in systems of coupled rooms. Jason E. Summers[a] (Program in Architectural Acoust., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180)

Statistical-acoustics (SA) models of reverberant energy decay in systems of coupled rooms conventionally represent energy densities of subrooms as solutions of systems of coupled ordinary differential equations. These models unrealistically assume that exchange of energy between

subrooms is instantaneous. Lyle has proposed a functional-differential-equation (FDE) SA model in which travel time of energy between subrooms is incorporated as constant delay times estimated from distances between centroids of adjacent subrooms [C. D. Lyle, Acoust. Lett. **4**, 248–252 (1981)]. Here, an FDE-SA model is proposed which more generally accounts for travel-time distributions and, further, incorporates previously introduced improved decay models within subrooms [J. E. Summers *et al.*, J. Acoust. Soc. Am. (in press)]. Monte Carlo methods are used to estimate travel-time distributions for simple geometries and allow for predictions with the FDE-SA model. Comparison of scale-model measurements and predictions of a computational geometrical-acoustics model with those of SA models shows that, contrary to Lyles findings, FDE-SA models must consider more than first moments of travel-time distributions in order to yield more accurate predictions. [This work was supported by the Bass Foundation.] [a] Current address: Acoustics Division, Naval Research Lab, Washington, DC 20375-5350.

11:30

3aAA10. Acoustical decoupling of a dome: The adaptive reuse of historic Angelus Temple. David Conant (McKay Conant Brook Inc., 5655 Lindero Canyon Rd., #325, Westlake Village, CA 91362, dconant@mcbinc.com)

With its 128-foot-wide, unsupported concrete dome springing 125 feet above the floor of its quarter-circle plan, the original 5300-seat mother church of the International Church of the Foursquare Gospel in Los Angeles (circa 1923) required adaptive reuse as a rock-n-praise Christian church. Challenges included squelching its 7-s RT60, decoupling the dome by eliminating its unusually severe focusing (largely responsible for repetitive echoes), especially strong at the intended audio mix position and stage, integrating acoustical and audio systems design to achieve 118 dBA uniformly throughout, while meeting full operation for opening services within 7 months. Because of its structurally weak dome and its National Register status, the principal acoustical solution prohibited covering, or supporting anything from, the focusing dome. Described is a method not known to have been applied elsewhere intervening about 95% of all the problematic, focusing sound rays with large suspended baffle/absorber panels specifically designed and positioned via laser model studies while achieving an attractive, dramatic array. The result is remarkably successful with respect to all intended acoustical, audio, and historical architecture goals.

11:45–12:00

Voluntary headphone paired comparison listening test to establish perception and preference levels of the double sloped decay.

WEDNESDAY MORNING, 17 NOVEMBER 2004 PACIFIC SALON 1, 8:25 TO 11:00 A.M.

Session 3aAB

Animal Bioacoustics: Marine Mammal Acoustics: Session in Honor of Ron Schusterman III

Patrick W. Moore, Chair
SPAWAR SSC-SD, 49620 Beluga Road, San Diego, California 92152-6506

Chair's Introduction—8:25

Invited Papers

8:30

3aAB1. Evoked-potential study of automatic gain control in the sonar of a false killer whale. Alexander Supin (Inst. of Ecology and Evolution, 33 Leninsky Prosp., 119071 Moscow, Russia), Paul Nachtigall, Whitlow Au, and Marlee Breese (Univ. of Hawaii, Kaneohe, HI 96744-1106)

Brain auditory evoked potentials (AE's) were recorded in a false killer whale Pseudorca crassidens trained to accept suction-cup EEG electrodes and to detect targets by echolocation. AEP collection was triggered by echolocation pulses transmitted by the animal. The target strength varied from 22 to 40 dB; the distance varied from 1.5 to 6 m. All the records contained two AEP sets: the first one of a constant latency (transmission-related AEP) and a second one with a delay proportional to the distance (echo-related AEP). The amplitude of echo-related AEPs was almost independent of both target strength and distance, though combined variation of these two parameters resulted in echo intensity variation within a range of 42 dB, thus indicating automatic gain control (AGC). The amplitude of transmission-related AEPs was independent of distance but dependent on target strength: the less the target strength, the higher the amplitude. Recording of transmitted pulses has not shown variation of their intensity depending on target strength. It is supposed that AGC mechanisms are a variation of hearing sensitivity depending on the target strength and release of echo-related responses from masking by transmitted pulses depending on the distance. [Work supported by ONR (USA) and RFBR Ministry of Science (Russian Federation).]

8:50

3aAB2. Acoustic monitoring of northern fur seals in the Bering Sea. Stephen J. Insley (Inst. of Marine Sci., Univ. of California Santa Cruz, Santa Cruz, CA), Bruce W. Robson (7305 9th Ave NW, Seattle, WA), and William C. Burgess (Greeneridge Sci., Santa Barbara, CA)

A substantial problem in determining whether anthropogenic noise affects marine animals is measuring how an animal responds to a noise source that is located in a remote pelagic environment and is ephemeral in nature. We are interested in measuring northern fur seal reactions to ship noise in the Bering Sea, Alaska, where the fur seals have started to decline. The timing of their decline coincides with an increased probability of interaction with the pollock fishing fleet. Whether fishing activity affects fur seal behavior, via direct competition for fish and/or indirect competition (e.g., acoustic disturbance or vessel avoidance), is unknown and difficult to study. We began by developing procedural tools to measure encounter rates and animal response. We deployed acoustic dataloggers on lactating females in order to monitor for ship signatures. This was referenced to movement data from satellite position fixes and pressure data from dive records. The combined dataset provides assays that measure an animal's movement (i.e., direction, speed, and relative effort via stroke rate), its overall activity budget (i.e., traveling, diving, and resting), and the pattern of diving behavior. The assays are now being field tested with a new set of instrument deployments during the 2004 fur seal breeding season.

9:10

3aAB3. Acoustic techniques provide insights into reproductive strategies of bearded seals, ♣*Erignathus barbatus*. Sofie M. Van Parijs and Chris W. Clark (Bioacoustics Res. Program, Lab. Ornithology, Cornell Univ., Ithaca, NY 14580, sofievp@nfh.uit.no)

Several pinniped species mate underwater. This makes it difficult to study their mating behavior and assess their reproductive strategies. Acoustic techniques improve our capacity to study these species in their natural habitat. Captive studies demonstrate that vocal and dive displays are produced only by males. Studies of free-ranging animals demonstrate that these displays are produced during the mating season. Males have individually distinct vocalizations. In the Norwegian Arctic, male bearded exhibit two strategies, with some males holding small discrete territories within which they display and others roaming over much larger areas. In Alaska, data from an acoustic array were used to explore male behavior over 18 years. Males exhibited long-term site fidelity. In the Norwegian Arctic, fjord systems leave large ice-free areas within which males establish territories. In Alaskan waters, male movements and strategies are influenced by fast ice cover.

9:30

3aAB4. Antimasking strategies of underwater vocalizations and hearing abilities of polar seals. Jack Terhune (Dept. of Biol., Univ. of New Brunswick, P.O. Box 5050, Saint John, NB E2L 4L5, Canada terhune@unbsj.ca)

Detection of underwater vocalizations by polar seals is limited by their auditory sensory abilities and external masking noises from abiotic (meteorological and ice noises) and biotic (conspecific calls) sources. Attributes that present the antithesis of masking noise characteristics are thought to enhance detectability of calls. Some proposed anti-masking strategies such as call repetition/rhythm patterns are supported by evidence of lower detection thresholds (1–5 dB), while others are not (e.g., abrupt onset and offset of calls). For frequency swept calls, downsweeps have lower detection thresholds (1–5 dB) than upsweeps. The majority of frequency swept calls (greater than 0.1 oct) of bearded (*Erignathus barbatus*), Weddell (*Leptonychotes weddellii*) and harp (*Pagophilus groenlandicus*) seals are downsweeps (89%, 86% and 63%, respectively). Temporal and frequency separation, call lengthening, and directional clues also reduce masking effects. Diverse call repertoires and calling behaviors of polar seals (e.g., courtesy rule) appear to have evolved characteristics that reduce the effects of abiotic and biotic masking. Characteristics of many anthropogenic underwater noises differ from sounds produced in nature. To estimate masking effects of anthropogenic noise on phocid communication, it is important to determine if the man-made noises are defeating existing antimasking strategies.

9:50–10:00 Break

Contributed Papers

10:00

3aAB5. Evoked potential measurements of the West Indian manatee modulation rate transfer function. David Mann, Brandon Casper, Mandy Cook (Univ. of South Florida, College of Marine Sci., 140 7th Ave. S., St. Petersburg, FL 33701, dmann@marine.usf.edu), Debborah Colbert, Joseph GaspardIII (Mote Marine Lab., Sarasota, FL 34236), Roger Reep (Univ. of Florida, Gainesville, FL 32610), and Gordon Bauer (New College of Florida, Sarasota, FL 34243)

Evoked potential measurements of two Florida manatees (*Trichechus manatus latirostris*) suggest that these herbivores have evolved an auditory system with high temporal resolution. The manatee modulation rate transfer function (MRTF) is maximally sensitive to a 600-Hz amplitude modulation (AM) rate. This modulation rate is midway between the AM sensitivities of terrestrial mammals (chinchillas, gerbils, and humans) (80–150 Hz) and dolphins (1000–1200 Hz). We also demonstrate evoked potential responses to carrier frequencies up to 40 kHz. These results suggest that manatees may have reasonable underwater localization abilities despite the high speed of sound underwater, which could be important in enabling them to localize oncoming boats.

10:15

3aAB6. Estimation of the acoustic reflectivity of a Florida manatee from physical measurements of animal tissue. Fernando Simonet, Jules J. Jaffe (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238, fsimonet@ucsd.edu), and Ann E. Bowles (Hubbs–Sea World Res. Inst., San Diego, CA 92109)

In order to better understand the cause of low acoustic reflectivity of the Florida manatee (*Trichechus manatus latirostris*), a series of measurements were taken from a stranded animal post-mortem. The samples came from an adult male that was struck and killed by a boat in April 2004. These measurements were used to calculate the speed of sound, density and attenuation within each sample. To achieve a good statistical representation, many subsamples (connective tissue, blubber and muscle) were extracted from ventral, dorsal and lateral samples collected from approximately the midpoint of the body (at the umbilicus). Results for the connective tissue samples indicate an average sound speed, density and attenuation of 1725 m/s, 1089 kg/m^3 and 5.21 dB/mm, respectively. These results establish a reflection coefficient of 0.12, or, in other words, only 12% of the incident acoustic wave is reflected in the water–skin interface.

10:30

3aAB7. Acoustic reflectivity measurments of sirenia (Florida manatees) at high frequencies. Jules S. Jaffe, Paul L. Roberts, Fernando Simonet (Marine Physical Lab, Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238), and Ann E. Bowles (Hubbs-Sea World Res. Inst., San Diego, CA 92109)

The Florida manatee (*Trichechus manatus latirostris*) is an endangered sirenian inhabiting shallow coastal waters of Florida (USA). Its population ($\sim$3300) is dwindling, with 25%–30% of mortalities occurring due to collisions with boats (95 losses in 2002). An active sonar system that detects the presence of the animals, and hence alerts the boater, could help reduce the collisions. Experiments were performed on six captive manatees with several calibrated transducers in a $33\times10\times3$ m^3 pool to learn how much sound is reflected by these animals. At a frequency of 171 kHz we transmitted a brief sinusoid. An underwater video camera was aimed along the axis of the range direction of the sound transmission, permitting the co-registration of animal and acoustics. The camera and sonar were calibrated together by translating a 38-mm tungsten carbide sphere (TS $=-39$ dB@ 171 kHz) in a separate test tank facility. Results indicate that the reflectivity of the animals (not strictly target strength) is somewhat low, in the -49 to -40 dB range.

10:45

3aAB8. Differential response by manatees to playbacks of sounds simulating approaching vessels. Jennifer L. Miksis-Olds, Percy L. Donaghay (Grad. School of Oceanogr., Univ. of Rhode Island, Narragansett, RI 02882, jmiksis@gso.uri.edu), James H. Miller (Univ. of Rhode Island, Narragansett, RI 02882), Peter L. Tyack (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), and John E. ReynoldsIII (Mote Marine Lab., Sarasota, FL 34236)

One of the most pressing concerns associated with the endangered Florida manatee is mortality due to collisions with watercraft. Watercraft collisions are the leading identified cause of adult mortality, resulting in greater than 30% of manatee deaths each year. Reducing adult mortalities is critical to the recovery of the manatee population, as population trends are more sensitive to adult deaths than to those of other age groups. Acoustic playback experiments were conducted to assess the behavioral responses of manatees to watercraft approaches. Playback stimuli were constructed to simulate a vessel approach to approximately 10 m in seagrass habitats. Stimulus categories were (1) silent control; (2) idle outboard approach; (3) planing outboard approach; and (4) fast personal watercraft approach. These results are the first to document responses of wild manatees to playback stimuli. Analyses of swim speed, changes in behavioral state, and respiration rate indicate that the animals respond differentially to the playback categories. The most pronounced responses, relative to the controls, were elicited by the personal watercraft. Quantitative documentation of response during playbacks will provide data that may be used as the basis for future models to predict the impact of specific human activities on manatee and other marine mammal populations.

WEDNESDAY MORNING, 17 NOVEMBER 2004 SHEFFIELD ROOM, 8:00 A.M. TO 12:15 P.M.

Session 3aAO

Acoustical Oceanography: Inverse Methods in Acoustical Oceanography

Peter Gerstoft, Chair

Scripps Institution of Oceanography, University of California San Diego, Marine Physical Laboratory, 9500 Gillman Drive, La Jolla, California 92093-0238

Contributed Papers

8:00

3aAO1. *In situ* measurement technique for compressional wave velocity dispersion. Charles W. Holland (Appl. Res. Lab., The Pennsylvania State Univ., State College, PA), Jan Dettmer, and Stan E. Dosso (Univ. of Victoria, Victoria, BC, Canada)

Marine sediments can be considered a two-phase medium composed of solid sediment particles and fluid-filled interstices. Many theories of propagation through such an assemblage predict a dispersive compressional wave velocity. Granular sediments (i.e., sands and gravels) have been predicted and reported to exhibit velocity dispersion and/or a nonlinear frequency-dependent attenuation in the frequency range 10^2-10^4 Hz (Kramers–Kronig relations require velocity dispersion if a nonlinear frequency-dependent attenuation is observed). One of the experimental challenges has been to measure velocity over several decades of frequency using a single technique so that observed velocity changes cannot be due to different biases in the measurement techniques. A simple *in situ* measurement technique is described that measures the critical angle as a function of frequency from $\sim10^2-10^4$ Hz and thus infers the velocity dispersion over that band. Measurements on the mid to outer continental shelf (Malta Plateau and the New Jersey shelf STRATAFORM area) show little or no velocity dispersion. This implies that the compressional wave attenuation for these sediments is linear or nearly linear over this band. [Work supported by the ONR and NATO Undersea Research Centre.]

8:15

3aAO2. Measuring sediment density and velocity gradients in the transition layer. Charles W. Holland (The Pennsylvania State Univ., Appl. Res. Lab., State College, PA), Jan Dettmer, and Stan E. Dosso (Univ. of Victoria, Victoria, BC, Canada)

The geoacoustic properties of marine sediments, e.g., bulk density and compressional velocity, commonly exhibit large variations in depth near the water–sediment interface. This layer, termed the transition layer, is typically of $O(10^{-1}-10^0)$ m in thickness. Depth variations within the transition layer may have important implications for understanding and modeling acoustic interaction with the seabed, including propagation and reverberation. In addition, the depth variations may contain significant clues about the underlying depositional or erosional processes. Characteristics of the transition layer can be measured directly (e.g., coring) or remotely. Remote measurements have the advantage of sampling without disturbing the sediment properties; they also have the potential to be orders of magnitude faster and less expensive than direct methods. It is shown that broadband seabed reflection data can be exploited to remotely obtain the depth-dependent density and velocity in the transition layer to high accuracy. Mid-shelf seabed reflection measurements in the Italian littoral are exploited to obtain transition layer properties. These properties agree with direct (i.e., core) measurements within the uncertainty esti-

mates. A slight but discernable bias between the two methods may be due to compaction during core sampling. [Work supported by ONR and NATO Undersea Research Centre.]

8:30

3aAO3. Range-dependent inversion for ocean and bottom properties from modal dispersion curves. Gopu Potty (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882), Subramaniam Rajan (Sci. Solutions, Inc., Nashua, NH 03063), and James Miller (Univ. of Rhode Island, Narragansett, RI 02882)

Range-dependent sediment acoustic properties estimated from modal dispersion data will be presented. Rajan *et al.* [J. Acoust. Soc. Am. **82**, 998–1017 (1987)] have developed a linear perturbation technique to determine the bottom acoustic properties using group velocity dispersion relation for range-independent environments. This method has been now extended to cover range-dependent inversions for the sediment acoustic properties. Results of inversions performed using synthetic data show that the range-dependent properties can be retrieved using data from multiple source-receiver combinations. If only one source-receiver combination is used, the inversion procedure determines the range averaged sediment properties. A range-dependent inversion technique will be applied to field data from ASIAEX East China Sea (ECS) Experiment. Broad band acoustic data from Wide Band Sources (WBSs) collected on a vertical line array in frequency range 10–400 Hz will be used for the inversion. Over 200 such WBSs were deployed during the experiment and many source–receiver combinations among these will be used to investigate the range variability of compressional wave speed profiles in the ECS experiment location. Results will be compared with core data, geophysical surveys and nonlinear inversions. [Work supported by ONR.]

8:45

3aAO4. Inversions for attenuation profiles using modal amplitude ratios. Gopu Potty and James Miller (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882)

Data from the ASIAEX East China Sea (ECS) experiment is used to estimate sediment attenuation coefficients as a function of frequency and depth. Modal amplitude ratios for modes in the frequency range 10 to 300 Hz are used to obtain the attenuation profiles. These inversions use broadband data from wide band sources (WBSs) deployed in the area. The modes are detected, identified and their amplitudes measured using a time-frequency wavelet analysis. A joint inversion for attenuation coefficient profile, water depth, source and receiver depths, range and empirical orthogonal functions in the water column is performed. This inversion differs from our earlier technique in that instead of modal attenuation coefficients we treat the attenuation coefficient profile as the primary unknown. This enables us to reduce the number of unknowns and hence allows us to use higher modes in the inversion. Data from many WBS charges will be used to estimate an average value for attenuation and its standard deviation. The results will be compared with core data and frequency and depth dependence will be examined. [Work supported by ONR.]

9:00

3aAO5. Validation of the rapid geoacoustic characterization (RGC) algorithm. Kevin D. Heaney and James Murray (Lockheed Martin ORINCON Defense, 4350 N. Fairfax Dr., Ste. 400, Arlington, VA 22003)

The rapid geoacoustic characterization (RGC) algorithm was developed to perform near-real time geoacoustic inversion in shallow water using the striations of passing surface ships of opportunity. It has been applied to inversion workshop data [Heaney, IEEE J. Ocean Eng. 2004 (1)] and measured surface ship data [Heaney, IEEE J. Ocean Eng. 2004 (1)]. It is currently under review for inclusion in the Navy's Ocean Atmospheric Media Library (OAML). As part of the OAML approval process a set of seven test cases was generated for the independent validation and verification of the code. These cases span the range of simulated to measured data, and range-independent to range-dependent environments. The method as well as results of the application of this algorithm to these test cases will be presented. The limitations of the algorithm will be addressed.

9:15

3aAO6. Geoacoustic inversions in shallow water using Doppler-shifted modes from a moving source. Eric M. Giddens and Michael J. Buckingham (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

Underwater acoustic measurements have been made of aircraft overflights of a shallow-water (15 m deep) environment off the coast of La Jolla, CA. The narrow-band harmonics generated by the engine and propeller penetrates the air–sea interface and excites the normal modes of the system with each mode having a characteristic Doppler shift. From a high-resolution FFT, the modal eigenvalues are determined from the modal frequencies. These eigenvalues are dependent on the acoustic properties of the sediment. For a Pekeris waveguide, the eigenvalues can be found from the transcendental dispersion relationship, providing a direct method for inverting for the density and sound speed in the sediment without the use of a complicated numerical model. The method is relatively straightforward and only requires a single sensor in the water column. The sensitivity of the inversion technique will be discussed in the context of simulated experiments as well as real data collected in the experiments performed north of Scripps pier, La Jolla. [Work supported by ONR and the ARCS Foundation.]

9:30

3aAO7. Synthetic data inversion of porous ocean sediment based on Biot model. Keunhwa Lee and Woojae Seong (Dept. of Ocean Eng., Seoul Natl. Univ., Seoul 151-742, Korea)

Geoacoustic inversion usually estimates parameters by modeling the ocean sediment as either acoustic or elastic media. Treating the sediment as porous medium and partial inversion for sediment moduli has been performed [N.P. Chotiros, J. Acoust. Soc. Am. **112**, 1853–1868 (2002)]. In this study, synthetic data inversion based on the full Biot model consisting of 13 unknown parameters is performed. At first, sensitivity tests are done to figure out how much a parameter affects the acoustic fields. Then, iterative inversions using synthetic data are performed based on the parameter's sensitivities. Sensitive parameters are searched first, while other insensitive ones are fixed at their representative values. Finally, the insensitive parameters are optimized with the previous found sensitive parameter values fixed. In addition to the full Biot model, the equivalent fluid model utilizing complex wave speed and density is used for the inversion to test if the Biot parameters can be retrieved from these inverted complex wave speed and density.

9:45

3aAO8. Subbottom profiling using a ship towed line array. T. C. Yang, Kwang Yoo, and Laurie Fialkowski (Naval Res. Lab., Washington, DC 20375)

This paper presents a technique for determining the effective depth of subbottom layers using a ship towed line array with a controlled source. Results are demonstrated using the MAPEX 2000 data. Local measurements of bottom layer thickness can be done using several methods: (1) Reflection profiling using a source and a receiver array at different ranges. (2) A high power downward looking boomer that can penetrate tens of meters of sediment layers. (3) A vertical line array using ambient noise. (4) Geo-acoustic inversion using acoustic data from a ship towed line array. The ship towed line array provides local measurements of the subbottom thickness without the requirement of high power sources and extended experimental setup, on one hand, and with minimal interference from shipping traffic on the other hand. We show here that a simple seismic profiling technique extended to the towed line array data provides the same

thickness profile as obtained by other techniques. It agrees with the geo-acoustic inversion method but can be done in real time. Conversely, it can be used to initialize the geo-acoustic inversion program. [Work supported by ONR.]

10:00–10:15 Break

10:15

3aAO9. Geoacoustic inversion and subbottom profiling with ambient noise. Chris H. Harrison (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy)

A total of six experiments using a drifting vertical array have been carried out between 2002 and 2004. The purpose was to investigate the bottom reflection properties by inversion of ambient noise (GAIN =GeoAcoustic Inversion of Noise) and to investigate subbottom layering (SUPRA-GAIN=SUb-bottom PRofiling using Ambient noise) at frequencies from a few hundred Hz to 5000 Hz. Areas covered include the fairly flat, layered sediments of the Malta Plateau and the rocky outcrops of the Ragusa Ridge in the Mediterranean south of Sicily. Interesting comparisons can be made with other surveying techniques such as the boomer and chirp sonar. Fourier theory states that the depth limit for such subbottom profiling is set by the frequency resolution of the deduced reflection loss. There are other factors that degrade the deeper returns and some pragmatic limitations have been established. Comparisons of results with models suggest that as far as acoustic reflection properties are concerned, in rocky localities the roughness of the rock is more important than the shear.

10:30

3aAO10. Characterizing depth dependent geoacoustic properties of the seabed using layering and reflection loss information inferred from ambient noise. Martin Siderius (Ctr. for Ocean Res., SAIC, San Diego, CA 92121) and Chris Harrison (NATO Undersea Res. Ctr., La Spezia, Italy)

Vertically beamformed ocean ambient noise data can be used to infer bottom reflection loss over a broad band of frequencies. Seabed layering introduces a fringe pattern into the reflection loss curve that is related, through a Fourier transform, to the two-way travel time to significant reflectors (caused by layering). The two-way travel times can be used to infer number of layers and depths. In this paper, a global search algorithm is applied to find the sound speed, density and attenuation in each seabed layer. The search is directed by the quality of fit between modeled and inferred reflection loss with the two-way travel times used as an additional constraint. This method avoids one of the usual problems associated with geoacoustic inversion: predetermining the number of significant layers to parametrize the seabed. If too few layers are chosen, the solution does not describe the seabed in enough detail and choosing too many leads to ambiguities in some of the inverted parameters. We will present the application of this technique to simulated and measured data from experiments in the Mediterranean.

10:45

3aAO11. The sensitivity of far-field measurements in shallow water to the geoacoustic properties of the seabed. Travis L. Poole (MIT/WHOI Joint Program in Oceanogr. and Oceanograph. Eng., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543) and George V. Frisk (Florida Atlantic Univ., Dania Beach, FL 33004)

Because sound propagating through shallow water interacts strongly with the seafloor, measurements of the pressure field can be used in geoacoustic inversion algorithms to determine the sound speed profile in the bottom. However, the sound most easily measured is that which is concentrated in the water rather than in the bottom. Furthermore, only sound that propagates at shallow angles reaches a receiver far from the source because only these shallow-angle paths undergo total reflection from the bottom as well as the surface. In light of this behavior, an investigation is made into the sensitivity of this shallow-angle portion of the pressure field to changes in bottom properties. Also addressed is the manner in which this sensitivity affects geoacoustic inversion schemes and their ability to accurately determine bottom properties from far-field pressure measurements. [Work supported by ONR.]

11:00

3aAO12. Estimation of particle sizes for a range of narrow size distributions of natural sands suspended in water using multifrequency acoustic backscatter. Christopher Smith (NCPA, 1 Coliseum Dr., Univ. of Mississippi, University, MS 38677), Daniel Wren (USDA-Natl. Sedimentation Lab., Oxford, MS 38655), and James Chambers (Univ. of Mississippi, University, MS 38677)

Particle size estimation of suspended sediment is an ongoing area of research in acoustics. A single transducer was used to measure multifrequency backscatter from a suspended sediment-carrying jet. A composite waveform composed of three frequencies was transmitted by the transducer. The sediment was composed of natural sand and covered a range of narrow size distributions. Amplitude ratios among the three frequencies were evaluated in the experimental data. The same ratios were predicted using scattering theory. The measured ratios were used with these theoretical equations to solve for an estimate of the particle sizes. The method works well at estimating large (300–850 micron) particle sizes, but overestimates the size of small (150–300 micron) particles.

11:15

3aAO13. Shallow water ocean sound speed estimation using travel times between multiple sources and receivers. Bruce D. Cornuelle, Philippe Roux, Tuncay Akal, William S. Hodgkiss, William A. Kuperman, H. C. Song (Scripps Inst. of Oceanogr., UC San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0230, bdc@ono.ucsd.edu), and Mark Stevenson (NATO Undersea Res. Ctr., La Spezia, Italy)

A 29-element source-receiver array has been developed and used for time reversal mirror (TRM) experiments at 3.5 kHz. The TRM hardware also has been used to collect waveguide impulse responses between the source array and a 32-element receive array located about 10 km away in 100 m of water. The data were analyzed using standard time-of-arrival methods from large-scale ocean acoustic tomography. Stable arrival peaks with sufficient signal-to-noise ratios are tracked over the duration of the experiment and identified with ray-like arrivals as predicted by both ray tracing and full-wave propagation codes. We then adjust the ocean sound speed field, bottom topography, and array element locations to produce the best match between observed and predicted travel times, using ray paths as the observation kernels in the inversion. Even with an imperfect knowledge of the bottom topography, it is possible to make sensitive, relative inversions, converting changes in travel times to changes in array element locations and ocean sound speed, assuming the bottom does not change. [Work supported by ONR.]

11:30

3aAO14. Frechet-derivative calculations of the environmental sensitivity of the arrival structure in shallow water. Jit Sarkar, Bruce Cornuelle, Philippe Roux, and W. A. Kuperman (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr. Mail Code 0238, La Jolla, CA 92093-0238, jit@mpl.ucsd.edu)

The expected sensitivity of a high-resolution tomographic inversion of a 10-km slice of shallow water has been studied. Parabolic equation calculations of the Frechet derivative of simulated data at 10-km range for measured sound-speed profiles were made. The computations are based on using an existing 29-element transmitter array together with a 32-element receive array separated by 10 km. The frequency bandwidth was 3.4 kHz. The Fourier syntheses of these results produce a sensitivity map for each temporal segment of the arrival structure.

11:45

3aAO15. Travel time difference and incoherent intensity tomography of the current and turbulence structures of shallow oceans. Tokuo J. Yamamoto (Flosonic Corp., 12200 SW 89th Ave., Miami, FL 33176, tyamamoto@bellsouth.net), Arata Kaneko, and Haruhiko Yamaoka (Hiroshima Univ., Hiroshima, Japan)

Acoustic waves propagating through irrotational current cause a travel time difference between the reciprocal transmission paths. When they propagate through turbulence, they are scattered and a part of the propagating acoustic wave energy becomes incoherent. The two acoustical principles lead to the two corresponding acoustic tomography methods for imaging current and turbulence. The acoustic cross-array tomography (ACT) experiment was conducted at the Kammon Strait on 17–20 March 2003. Eight ACT measurement units were distributed on both sides of the strait. Gold codes having a carrier frequency of 5.5 kHz were transmitted as source signals. The data obtained from the experiment were used in the travel time difference tomography for imaging the current field and in the incoherent acoustic intensity tomography for imaging turbulence in the strait water. The comparisons between the current fields and the turbulence fields reveal that the turbulence intensity is nearly proportional to the current strength. This indicates that the turbulence in the very shallow (roughly 11 m deep) strait water is generated within and may be transported outside the bottom boundary layer. [Work supported by ONR.]

12:00

3aAO16. Acoustic-oceanographic buoy—An easily deployable, reconfigurable, and multifunctional acoustic-oceanographic system. Sergio M. Jesus, Cristiano Soares, Antonio Silva (SiPLAB-FCT, Univ. of Algarve, PT-8005-139 Faro, Portugal), Jean-Pierre Hermand (Université Libre de Bruxelles, 1050 Brussels, Belgium), and Emanuel F. Coelho (NATO Undersea Res. Ctr., I-19138 La Spezia, Italy)

The concept of an easy to use and easy to deploy ocean acoustic tomographic (OAT) system is presented. The system is composed of a network of buoys and a data inversion online processor. This study concerns the individual node of that network—the acoustic-oceanographic buoy (AOB)—the data inversion technique and the testing of the system at sea. The AOB is a lightweight surface buoy with a vertical array of acoustic and temperature sensors to be hand deployed in a free-drifting configuration from a small boat. The data are locally stored and transmitted online to a remote station for processing and monitoring. Data inversion is based on a broadband matched-field tomography technique where known and unknown parameters are simultaneously searched for (focalization). *In situ* recorded temperature data serve for algorithm initialization and calibration. The AOB was successfully deployed in several consecutive days during two rapid environmental assessment sea trials in 2003 (Mediterranean) and 2004 (Atlantic). Data collected at sea also show that the AOB can be reconfigured as a receiving array for underwater coherent communications in the band up to 15 kHz. [Work supported by FCT and Royal Netherlands Navy—Projects NUACE, POSI/CPS/47824/2002 and REA, RNLNC/2003/04.]

WEDNESDAY MORNING, 17 NOVEMBER 2004 PACIFIC SALONS 6 & 7, 8:00 TO 10:30 A.M.

Session 3aBB

Biomedical Ultrasound/Bioresponse to Vibration: Acoustic Bioeffects

Yuri A. Pishchalnikov, Chair
Department of Anatomy and Cell Biology, Indiana University, 635 Barnhill Drive, Indianapolis Indiana 46202-5126

Contributed Papers

8:00

3aBB1. Ultrasonic tissue characterization of the beating hearts of mice: Evidence of anisotropy and of hypertrophy. James G. Miller, Attila Kovacs, Stephanie H. Posdamer, Kirk D. Wallace, Michael R. Courtois, Carla J. Weinheimer, and Mark R. Holland (Dept. of Phys. and Ctr. for Cardiovascular Res., Washington Univ.)

New insights into the disease-altered properties of the heart may be provided through the study of genetically manipulated mice. To extend methods developed for studies of the hearts of patients to the hearts of mice, we report the observation of anisotropy of backscattered ultrasound in the parasternal short-axis view of normal mouse hearts, the cyclic variation of backscatter in normal hearts, and alterations of that pattern of cyclic variation in hypertrophic hearts. Echocardiographic images were obtained using a clinical imaging system with a nominal 13-MHz linear array. In seven wild-type mice the anisotropy of mid-myocardial apparent backscatter arising from the angle of insonification relative to the local fiber orientation in the short-axis view was consistent with that observed previously in the hearts of healthy human volunteers. The magnitude of the cyclic variation of backscatter from the hypertrophic hearts of nine mice subjected to neonatal aortic banding was significantly less than that obtained from eight control, sham-operated mice. This study lays the groundwork for subsequent investigations in which genetically manipulated mice will be measured in order to clarify the mechanisms of altered ultrasonic scattering and attenuation of the disease-altered myocardium in the presence of myocardial anisotropy. [Work supported by NIH R37HL40302.]

8:15

3aBB2. Respiratory smooth muscle relaxation using vibrations. Youhua Du and Ahmed Al-Jumaily (Auckland Univ. of Technol., Private Bag 92006, Auckland, New Zealand)

Many respiratory disorders (asthma in particular) are associated with the shortening of the smooth muscles within the airway walls. This could be attributed to an allergic response that results in airway constrictions. Physically, the shortening of smooth muscles is caused by cross-bridges (the heads of myosin molecules attached to the actins). Muscle relaxation is normally achieved by special bio-chemically driven medications. This work has demonstrated that muscle relaxation can be achieved by mechanical excitation. The dynamic response of contracted tracheal smooth muscles to vibrations under isometric conditions is investigated. This is achieved using two-dimensional finite element modeling as well as experimental investigation. The main algorithm of the finite element modeling is to determine the variation in the muscles' stiffness caused by external excitation. However, the experimentation focused on using an electric field to stimulate the contraction and using mechanical vibration to produce the relaxation. The main outcome from this work is that smooth muscle relaxation can be achieved by contact mechanical vibration. The question that remains to be answered is can contracted smooth muscles be relaxed by noncontact oscillation? Answering this will definitely add to the body of knowledge for developing a drug-free methodology of relieving airway constrictions.

8:30

3aBB3. Effect of exposure duration on lung hemorrhage from low-frequency underwater sound. Diane Dalecki, Sally Z. Child, and Carol H. Raeman (Dept. of Biomed. Eng. and Rochester Ctr. for Biomed. Ultrasound, Univ. of Rochester, Rochester, NY 14627)

Previous research from our lab has shown that low-frequency (~100–1000 Hz) underwater sound exposure can produce lung hemorrhage in laboratory mammals. The threshold for damage is lowest for exposure at the resonance frequency of lung. In adult mice, the threshold for lung damage is ~2 kPa using a 3-min continuous wave (cw) exposure at the resonance frequency (~300 Hz). The objective of this study was to determine the effect of exposure duration on lung damage produced by low-frequency sound. Experiments were performed with anesthetized adult mice. The resonance frequency of each mouse lung was determined using an acoustic scattering technique. Mice were exposed cw at the resonance frequency for durations of 1, 5, 10, 30 s, 1 and 3 min. The threshold for lung hemorrhage was relatively independent of duration for exposures greater than or equal to 1 min. For short exposure durations, the threshold increased and the mean extent of lung damage decreased. In some instances, extensive damage to the lung was observed to occur even for the 5-s exposure duration. However, the likelihood of extensive lung hemorrhage increased with increasing exposure duration.

8:45

3aBB4. Dual frequency high intensity focused ultrasound to accelerate hemostasis. Michael R. Bailey, Adam D. Maxwell, Justin A. Reed, Wayne Kreider, Vera A. Khokhlova, Vesna Zderic, Shahram Vaezy, and Lawrence A. Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

In acoustic hemostasis, cavitation appears to help emulsify and heat tissue to form a paste which seals lacerations. Our objective was to mix two frequencies to enhance cavitation, increase heat deposition, and accelerate hemostasis. A focused transducer (curvature 5 cm) with a 4.3-MHz center element (diameter 2 cm) and a 125-kHz outer annular element (inner and outer diameters, 3 and 5 cm) was engineered. This dual frequency transducer was used to create lesions in a transparent gel phantom and to induce hemostasis in five rabbits. Hemostasis time was determined with both frequencies (4.3 MHz at 620 W/cm^2, 125 kHz at 4 W/cm^2) and with the high frequency only (4.3 MHz at 620 and 780 W/cm^2). In gels, dual frequency HIFU created larger, broader lesions than 4.3-MHz HIFU alone. Bubbles created by just low frequency could be seen and, with sufficient time, created small local lesions. In rabbits, hemostasis times (~1 min for 25-ml/min bleeds) were decreased 20% with dual frequency compared to the 4.3-MHz wave alone at the same intensity or same total electrical power. Dual frequency mixing can enhance heat deposition and accelerate hemostasis by HIFU. [Work supported by NSBRI SMS00203, NIH Fogarty, ONRIFO, and NSF BES0002932.]

9:00

3aBB5. Vector Doppler ultrasound for the detection of internal bleeding. Bryan Cunitz, Peter Kaczkowski, and Lawrence Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105-6698)

A vector Doppler (VD) ultrasound system has four advantages over color-flow Doppler (CFD) in identifying internal bleeding: (1) measures instantaneous velocity more accurately, (2) measures flows perpendicular to the transmitter, (3) distinguishes between flows at different depths, and (4) detects the 3-D direction of the flow. A real-time vector Doppler system has been built and tested *in vitro*. The system consists of one focused transmitting/receiving transducer (10 mm diameter, 50 mm curvature, 5 MHz frequency) and four receiver transducers (5 mm diameter, flat discs, 5 MHz frequency). The signals are passed through analog filters and preamplifiers and then digitized at 20 MHz to obtain a Doppler signal. A digital signal processing algorithm is implemented to calculate flow direction from a combination of the five received signals as well as remove error from systematic phase shifts inherent in the instrumentation. The software was also designed to decrease the time it takes a user to locate arterial flow vessels. The system was tested using two flow phantoms: a moving string to verify accuracy and precision of the velocity, and cornstarch flowing through a tube to simulate blood flow. For both phantoms, device accuracy was tested over a range of angles and velocities.

9:15

3aBB6. Echogenicity from disturbed blood flow by an eccentric stenosis under pulsatile flow. Dong-Guk Paeng, Min Joo Choi (Interdisciplinary Postgraduate Program of Biomed. Eng., Cheju Natl. Univ., 1 Ara 1, Jeju, Korea 690-756), Suk Wang Yoon (SungKyunKwan Univ., Suwon, Korea 440-746), and K. Kirk Shung (Univ. of California, Los Angeles, CA 90089-1451)

In order to investigate echogenicity from blood flow disturbance, a severe eccentric stenosis (~70%) was constructed in a polystyrene tube (9.5-mm inner diameter) of a mock flow loop. A GE LOGIQ 700 Expert system was used to collect longitudinal and cross-sectional B-mode images from porcine whole blood. Harmonic, power, and color Doppler, and B-flow modes were also collected as references. Flow speed and stroke rate were controlled under pulsatile flow. The ''black hole,'' a hypoechoic zone at the center stream, was observed upstream of the stenosis during systole, while an expanding ''bright ring'' appeared during diastole in cross-sectional images. These rings looked like eddies due to backflow during diastole in longitudinal images. A bright stream line appeared from the orifice of the stenosis along the higher flow-speed zone during systole. A hyperechoic parabolic profile appeared once during a pulsatile cycle in longitudinal images showing the periodic spatial variation further downstream under some flow conditions. The parabolic profiles seemed to be a bright ring in cross-sectional images. Both longitudinal and cross-sectional images were helpful to understand 3D structures of echogenicity from disturbed blood flow by an eccentric stenosis under pulsatile flow.

9:30

3aBB7. Coupled vibration of a fluid-filled and submerged vascular tube with internal transitional/ turbulent flow due to a constriction. Yigit Yazicioglu, Thomas J. Royston, Todd Spohnholtz, Bryn Martin (Univ. of Illinois at Chicago, 842 W. Taylor St., MC 251, Chicago, IL 60607), and Francis Loth (Univ. of Illinois at Chicago)

The vibration of a thin-walled cylindrical, compliant viscoelastic tube with internal flow and an axisymmetric constriction that results in transitional/turbulent fluid flow is studied theoretically and experimentally. Vibration of the tube is considered with internal fluid-coupling only and with coupling to internal flowing fluid and external stagnant fluid. The theoretical analysis includes the adaptation of a model for turbulence in the internal fluid and its vibratory excitation of and interaction with the tube wall. Theoretical predictions are compared with experimental measurements conducted on a flow model system using laser Doppler vibrometry to measure tube vibration and miniature hydrophones to measure fluid pressure. This study is relevant to and may provide further insight into vascular diagnostics via noninvasive acoustic measurements. [Research support: NIH EB002511 and HL55296, and Whitaker Foundation BME RG 01-0198.]

9:45

3aBB8. Forward modeling of propagation and scattering of acoustic waves for hyperthermia treatment. Koen W. A. van Dongen and William M. D. Wright (Dept. of Elec. and Electron. Eng., Univ. College Cork, Cork, Ireland)

Cancer is a major cause of death in the Western world, resulting in a high demand for effective therapies. Recently, there is increasing interest in treatment via high intensity focused ultrasound (HIFU) where ultrasonic waves are used to increase the tumor temperature to induce cell death. For successful treatment, information about the temperature is essential. Research is applied where temperature profiles are obtained via acoustic wavefields. In order to understand the measurements and to verify imaging methods the acoustic wavefields are modeled. The forward model is based on a nonlinear integral equation of the second kind, obtained via reciproc-

ity from the acoustic wavefield equations. The integral equation relates the unknown total wavefields with the known incident wavefields and inhomogeneities in density and compressibility. A conjugate gradient method is used to solve the integral equation to obtain a solution for this ill-posed problem. To reduce the computational costs, a technique called "marching on in anything" is applied in both the temporal and the spatial domain. This technique is especially successful where wavefields have to be computed for various transducer positions. Results will be shown where the propagation and scattering of acoustic waves are modeled by using the above method.

10:00

3aBB9. Computational methods for imaging small temperature changes using ultrasonic through-transmission tomography for thermal cancer treatment. Konstantin Bograchev and William M. D. Wright (Dept. of Elec. and Electron. Eng., Natl. Univ. of Ireland, Cork, Ireland, konstantinb@rennes.ucc.ie)

During thermal cancer treatment there are small temperature fluctuations (37–42.5°C) inside the human body. These fluctuations are very important to monitor for quality cancer treatment. For monitoring deep seated fluctuations it is effective to use ultrasonic through-transmission tomography. Computer modeling of temperature fluctuation reconstruction using the above technique has been carried out. The model considers a fan-bean tomography scheme (60 sensors, 2160 rays in total) and iterative techniques for solving the inverse problem. It also assumes the 2D temperature is being reconstructed in an area of tissue of 60×60 mm^2 size, and the temperature fluctuation size is 6.5 mm in diameter. Such heated regions can appear, for example, during thermal treatment using high intensity focused ultrasound (HIFU). Algorithms for solving the inverse problem using algebraic iterative methods (including EM and ART) have been developed. The model takes into account electronic measurement noise and angle positioning errors. The model uses bilinear interpolation and takes into account the interpolation error. It also takes into account *a priori* information about the heating process. It was shown that it is possible to reconstruct the above temperature fluctuation with a good quality.

10:15

3aBB10. Improved exposimetry on diagnostic ultrasound equipment using broadband complex-valued hydrophone frequency responses. Volker Wilkens and Hans-Peter Reimann (Physikalisch-Technische Bundesanstalt, Dept. 1.6, Sound, Bundesallee 100, 38116 Braunschweig, Germany)

Nonideal hydrophone frequency responses may impede correct acoustic output measurements on medical ultrasound equipment, in particular when high-frequency or nonlinearly distorted waveforms have to be detected. However, correct pressure waveforms and the associated standard pulse parameters can be obtained by impulse deconvolution if the nonideal frequency response of the hydrophone $M(f)$ is provided with high frequency resolution both in amplitude and phase in a broad frequency range. The complex-valued calibration data required can be obtained by a novel secondary hydrophone calibration technique using broadband, nonlinearly distorted, focused ultrasound pulses and a very broadband optical multilayer hydrophone as reference receiver. The calibration data obtained for a bilaminar membrane hydrophone are applied to improve exposure measurements on a commercial diagnostic ultrasound machine. In comparison to the results obtained by the proposed pulse deconvolution method using the complete broadband complex-valued transfer function $M(f)$, an overestimation of the positive peak pressure by up to 50%, an underestimation of the rarefactional peak pressure by up to 11%, and an overestimation of the pulse intensity integral by up to 28% are found if the common evaluation method using the voltage-to-pressure transfer factor at the acoustic working frequency $M(f_{awf})$ is applied.

WEDNESDAY MORNING, 17 NOVEMBER 2004 PACIFIC SALON 3, 8:00 A.M. TO 12:00 NOON

Session 3aMU

Musical Acoustics: Computer Music I

Julius O. Smith, Chair

CCRMA, Music Department, Stanford University, Stanford, California 94305

Invited Papers

8:00

3aMU1. Stiff piano string modeling: Computational comparison between finite differences and digital waveguide. Julien Bensa (Laboratoire d'Acoustique Musicale, Universite Pierre et Marie Curie, Paris, France, bensa@lam.jussieu.fr), Stefan Bilbao (Queen's Univ., Belfast), Richard Kronland-Martinet, Thierry Voinier (Laboratoire de Mecanique et d'Acoustique, Marseille, France), and Julius O. SmithIII (Stanford Univ.)

As is well known, digital waveguides offer a computationally efficient, and physically motivated, means of simulating wave propagation in strings. The method is based on sampling the traveling wave solution to the ideal wave equation and linearly filtering this solution to simulate dispersive effects due to stiffness and frequency-dependent loss; such digital filters may terminate the waveguide or be embedded along its length. For strings of high stiffness, however, dispersion filters can be difficult to design and expensive to implement. It is shown how high-quality time-domain terminating filters may be derived from given frequency-domain specifications which depend on the model parameters. Particular attention is paid to the problem of phase approximation, which, in the case of high stiffness, is strongly nonlinear. Finally, in the interest of determining the limits of applicability of digital waveguide techniques, we make a comparison with more conventional finite difference schemes, in terms of computational cost and numerical dispersion, for a set of string stiffness parameters.

8:25

3aMU2. Interconnecting digital waveguides and finite difference structures: Recent advances. Matti Karjalainen and Cumhur Erkut (Helsinki Univ. of Technol., Lab. of Acoust. and Audio Signal Process., P.O. Box 3000, FI-02015 HUT, Finland)

Digital waveguides and finite difference schemes have been generally considered as alternative techniques in physical modeling of spatially distributed systems. The main difference between these techniques is in the variables used in the computation; in digital waveguide simulations traveling-wave components are used, whereas physical variables are typical in finite difference schemes. Although both techniques are known to provide exact modeling of ideal one-dimensional band-limited wave propagation and both of them can be composed to approximate higher-dimensional mesh structures, they usually become computationally different in the presence of losses, dispersion, and nonlinear interactions. Despite these differences, successful examples of hybrid waveguide/finite-difference models have been proposed in the literature. In these references, the finite-difference sections typically account for local nonlinear interactions. More recently, a general, unifying, and DSP-oriented framework is introduced, where the functional equivalence of digital waveguides and finite difference schemes is systematically elaborated and the conditions of building mixed models are studied. The present lecture will summarize the new framework and provide demonstrations of hybrid models. Real-time sound examples will be presented. [Work supported by EU ALMA project (IST-2001-33059).]

8:50

3aMU3. The digital waveguide mesh: Applications in acoustic modeling and current research directions. Damian T. Murphy (Media Eng. Group, Dept. of Electron., Univ. of York, Heslington, York, UK)

The digital waveguide mesh (DWM) was first introduced as a modeling technique for musical acoustics applications in 1993 as a natural extension to the 1D digital waveguide that is widely used in physical modeling sound synthesis. DWM related work at York has focused on a number of areas with notable results including reverberation simulation using the 2D triangular mesh, modeling enclosed spaces with the 3D tetrahedral mesh, improved anechoic absorbing boundary conditions (ABCs) and the analysis and validation of spatial properties of DWM-based room impulse response measurements. This paper presents recent results from three current areas of research. *RoomWeaver* is a development environment that facilitates research into the application of DWM-based models for virtual acoustic spaces, incorporating new hybrid mesh-types in 2D and 3D through the use of KW-pipes. The second parallel research area involves the development of new boundary formulations that extend the anechoic ABC to the more general frequency dependent case with variable diffuse reflection control. Finally, *Digitract* is a speech synthesis research tool based on a 2D DWM model of the human vocal tract that offers improved control over formant bandwidths and greater naturalness in the production of speech sounds.

9:15

3aMU4. Nonlinear transverse string vibration: Model comparisons and numerical methods. Stefan Bilbao (School of Music, Queen's Univ., Belfast, BT7 1NN, UK, s.bilbao@qub.ac.uk.)

There has recently been an increase in activity in developing digital models of nonlinear string vibration, for sound synthesis purposes; in a musical setting, such behavior is of interest under large-amplitude plucked string conditions. The usual starting point for such a model is a one-dimensional time-dependent second-order partial differential equation, which is discretized to yield a recursion; finite difference schemes are perhaps the most straightforward approach. There are, however, many suitable models; a general nonlinear transverse wave equation is taken as a starting point, and several approximations, namely the Kirchhoff–Carrier equation, digital waveguide type (i.e., tension-modulated) formulations, and the linear wave equation, are examined. The various solutions are compared and contrasted, both in terms of travelling wave propagation and the evolution of frequency content. The possibilities of and problems inherent in developing well-behaved numerical methods are discussed on a case-by-case basis, calling particular attention to the use of energetic quantities in discrete time as a means of controlling stability.

9:40

3aMU5. 3D mesh design to evaluate sense of place in closed spaces. Stefania Serafin and Federico Fontana (Aalborg Univ. Cophenhagen, The Eng. College of Cophenhagen, Lautrupvang 15, DK-2750, Ballerup)

The waveguide mesh defines ideal numerical schemes for the simulation of acoustic linear pressure wave fields, characterized by having constant parameters along the propagation domain. Using the waveguide mesh it is possible to accurately simulate closed spaces of arbitrary size and dimensions. We propose a real-time implementation of the 3D waveguide mesh which runs under the Pure data platform. The parameters the user can control, in real-time, are the mesh dimensions, the damping characteristics of the surfaces and the listening point and sound source positions. In previous experiments, we tested the listeners' ability to evaluate distance, by placing different sound sources inside an acoustic tube simulated by using the 3D mesh. In this paper we are interested in evaluating humans' ability to perceive the size of an acoustic enclosure. We calculated the impulse response of 3D meshes of different sizes and shapes, and asked the listener to rate such responses according to the perceived size of the enclosure. Results of this experiment will be shown during the presentation.

10:05–10:15 Break

10:15

3aMU6. Musical effects of the digital pressure-controlled valve. Tamara Smyth, Jonathan S. Abel, and Julius O. SmithIII (CCRMA, Dept. of Music, Stanford Univ., Stanford, CA 94305, tamara@ccrma.stanford.edu)

In computer music, a digital simulation of the pressure-controlled valve is required to model many wind instruments and vocal systems. The method used for simulating the valve usually depends on how closely it corresponds to any of three simple configurations for valves in acoustic tubes. The classification of the valve is determined by whether an additional upstream or downstream pressure causes the valve to open or close further on the corresponding side. In this work, a configurable model of a pressure controlled valve is implemented, allowing the user to design a valve simply by changing the model parameters. The user may design three simple models: one that is blown closed (like woodwinds or reed-pipes), one that is blown open (as in simple lip-reeds, the human larynx, harmonicas and harmoniums) and the transverse model (as in the bird's vocal organ, the syrinx). Using the computer model, the musical applications of different valve configurations and topologies (where valves are arranged in series and in parallel) will be examined. For several cases, observations will be made on how the valve couples differently with the tube-resonator and how this affects the produced sound.

10:40

3aMU7. Real-time timbre analysis as a technique for live synthesis control. Miller S. Puckette (Dept. of Music, UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0326, msp@ucsd.edu)

The possibility is explored of controlling synthesis using an instrumental or other sound source, not by trying to recreate the time-varying timbre of the musician's sound exactly, but instead by making the output sound reflect changes in the input, in a way a musician can control in order to produce an interesting stream of synthetic sound. First the range of spectra available from the sound source, and also that available from the synthesis technique, are estimated. A very simple measure of timbre is used that maps any sound into a point in a ten-dimensional space. (As a byproduct, the ten-dimensional volume spanned by a sound source gives a rough measure of timbral variability.) Then synthesis parameters are chosen for the synthesis algorithm by searching a database of known output timbres.

11:05

3aMU8. Piano sound synthesis: Relating physical phenomena to sound quality. Julien Bensa and Daniele Dubois (Laboratoire d'Acoustique Musicale, 11 rue de Lourmel, 75015 Paris, France, bensa@lam.jussieu.fr)

The piano is a complex instrument with a large number of mechanical elements, the majority of which contribute to the sound production. The physical characteristics of these elements together with their interaction influence the timbre of the piano sound. Thus, in order to give a precise description of the behavior of this instrument and effectuate a satisfactory sound synthesis, the totality of the physical phenomena that are part of the sound production ideally should be taken into account. However, due to the complexity of the sound production system, this is not possible. Still, works on piano sound synthesis have made it possible to model (usually non-real-time) high quality piano sounds. These models represent a great tool when studying the perception of piano timbre. We therefore propose, thanks to sound modeling and subjective studies, to determine the perceptual effect of phenomena involved in a sound production system. This approach hopefully will give a better understanding of the relation between the physical behavior of the instrument and the perception of its sound quality, and hereby give clues to how the piano model can be simplified without loss of quality. This is crucial for the conception of real-time high-quality synthesis models.

Contributed Papers

11:30

3aMU9. Equivalence of the digital waveguide and finite difference time domain schemes. Julius O. SmithIII (CCRMA, Music Dept., Stanford Univ., Stanford, CA 94305)

It has been previously understood that the finite difference time domain (FDTD) method for numerically solving partial differential equations is equivalent to the digital waveguide (DW) method over the DW solution space. However, prior references have assumed that the DW solution space is a strict subset of the FDTD solution space, so that the FDTD method exhibits solutions not available in the DW method. In this talk, based on arXiv.org e-print physics/0407032, we show that there is in fact a one-to-one correspondence between the two solution spaces, and we give a simple, explicit transformation for the 1D case (e.g., for modeling vibrating strings). The ability to transform FDTD models to DW models and back provides a means of guaranteeing passivity and/or stability of specific FDTD models with specific boundary conditions. Furthermore, wide classes of passive boundary conditions can be specified for the FDTD method based on the linear transformation of known passive terminations in the digital waveguide case. Similarly, physically precise FDTD excitations (input signals) can be defined based on their DW counterparts.

11:45

3aMU10. Spectral accuracy of numerical integration schemes for acoustic simulation. John Amuedo (Signal Inference Corp., 3267 Butler Ave., Los Angeles, CA 90066)

Physical simulation of time-dependent acoustical systems often requires advancing one or more state derivatives of such a system forward in time. Design of a discretization for such a system, and choice of particular numerical integration methods within an acoustical model, determine its physical integrity, numerical accuracy, stability, and realizability. Selection of an appropriate discretization and integration approach requires balancing indigenous numerical properties of a physical model, desired physical accuracy of results, stability constraints on allowable time-step size, computational complexity, and memory requirements. Physical simulation of musical instrument and cochlear mechanics may impose additional demands on physical accuracy, as choice of a particular computational

method colors perceptual fidelity of audio output. This talk analyzes magnitude and phase errors, and stability characteristics of zero- through fourth-order members of the families of numerical integrators most commonly employed for solving time-dependent acoustics problems. The methods compared include explicit Adams–Bashforth predictors, implicit Adams–Moulton correctors, backward differentiation integrators, and conditionally stable schemes including the Simpson, Tick and ''leap frog'' methods.

WEDNESDAY MORNING, 17 NOVEMBER 2004 SUNSET ROOM, 9:00 TO 11:20 A.M.

Session 3aNS

Noise, Architectural Acoustics, Animal Bioacoustics, Engineering Acoustics, Musical Acoustics, Physical Acoustics, Committee on Standards and Underwater Acoustics: Special Session to Honor the Contributions of Robert W. Young to Acoustics

Alan H. Marsh, Chair
16072 Santa Barbara Lane, Huntington Beach, California 92649-2155

Chair's Introduction—9:00

Invited Papers

9:05

3aNS1. Robert W. Young, acoustical consultant, 1940 to 1990. Douglas K. Eilar (Eilar Assoc., 539 Encinitas Blvd., Ste. 206, Encinitas, CA 92024, deilar@eilarassoc.com)

Robert Young was hired by Vern O. Knudsen in the early 1940s to assist in the acoustical research work being conducted in support of the U.S. Navy's efforts to develop improved methods to use sound signals to detect and monitor ships and submarines. In this capacity, Bob served as an in-house consultant on all acoustical aspects of that work. After the war, Bob Young formed a partnership with Bob Gales to provide acoustical consulting services for a large variety of clients. Consulting activities included projects in architectural acoustical design, measurement and evaluation of environmental or community noise, neighborhood noise complaints, and the development of acoustical criteria for use in city and county noise ordinances. Bob Young also helped draft the noise regulations of the State of California. This paper describes highlights of some of the acoustical consulting work undertaken by Robert Young over five decades from the 1940s to the 1990s.

9:25

3aNS2. Robert Young's contributions to musical acoustics. Uwe J. Hansen (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47809)

Of Bob's 32 publications in musical acoustics 17 dealt with tuning and intonation. These publications span a time period of 43 years. Among them are two reviews in JASA, one in 1954: ''Twenty-five Years of Musical Acoustics,'' and one in 1979: ''Fifty years of musical acoustics in the Acoustical Society of America.'' Some of Bob's important contributions to tuning and intonation will be discussed, one of which was the invention with Allan Loomis of the Chromatic Stroboscope while he was employed at the C.G. Conn Music Company in the 1930s. Dan Martin, in a review of the device, called the Chromatic Stroboscope the most-precise frequency-measuring instrument available for acoustical research in the audio-frequency range. Current activities in musical acoustics in view of Bob's 25- and 50-year review comments will be considered.

9:45

3aNS3. The impact of Robert Young on national and international standards. Paul Schomer (Schomer and Assoc., Inc., 2117 Robert Dr., Champaign, IL 61821)

Bob's enthusiasm and active participation in standards from modest beginnings for the Z24 Committee in the 1950s to the important needs for national and international standards in the 1960s to the 1990s was truly remarkable. Bob was a most-conscientious and detailed reader of drafts, redrafts, and new editions of documents for IEC/TC 29, ISO/TC 43 and SC 1, ISO/TC 108, and our four ANSI S-Committees. He collected and used his own experiences and even conducted his own experiments to obtain data to support his proposals. Among other areas, Bob worked for 50 years on national and international Standards related to instruments, measurements, terminology, and scales and ratios for plotting. Bob was noted for his perseverance; if he believed in something, he had to make it happen. Some things that Bob had to make happen were the use of A-weighting and time-average for measurement and evaluation of environmental noise; here he was quite successful. Bobs achievements over a 50-year Standards history helped to push the USA and the world towards better acoustical instruments and measurements, clearer and more-consistent acoustical terminology, and clearer graphs and figures, but with a continuing effort to avoid dBA as the unit of A-weighted sound level.

10:05–10:20 Break

10:20

3aNS4. Contributions of Robert W. Young to the understanding of Space Shuttle sonic booms. Evelyn C. Young (Escondido, CA) and T. L. Young (Sea Space Corp., 12120 Kear Pl., Poway, CA 92064)

From 1981 to 1993, Robert W. Young measured 30 sonic booms generated during Space Shuttle landings at Edwards Air Force Base. The chronology of these measurements is based on the recollections of Evelyn Young who assisted Robert Young on all of these excursions. From the early days of the first Space Shuttle landings, with their unknowns of flight paths and landing times, to the routine landings 12 years later, the successes, failures, challenges, and results are described. Anecdotes complete a picture of the effort needed to accumulate the acoustical records that resulted in the presentation of six papers in the period from 1984 to 2002 relating to Space Shuttle landings. Excerpts from the paper ''Sonic Booms of Space Shuttles approaching Edwards Air Force Base, 1988-1993'' are presented to illustrate the ground tracks, flight paths, and sonic boom waveforms of landings of the Shuttles *Columbia* and *Atlantis*.

10:40

3aNS5. Bob Young and his contributions to animal bioacoustics and underwater acoustics. William C. Cummings (5948 Eton Ct., San Diego, CA 92122-3203, oshundbs@san.rr.com)

Quintessential component and Past President of ASA, Robert W. Young was truly a scientist *par excellence* for all seasons. His professional accomplishments excelled in large measure with outstanding assistance of his talented wife, Evelyn. Whether noise, psychoacoustics, architecture, propagation, music, patents, underwater sound, animal bioacoustics, or other discipline(s), his trademark was always distinguished by elegant design and very proper presentation. Whenever new or experienced co-workers needed a bit of correction or retuning, Bob Young (known to many of us as ''Uncle Fussy Ears'') was the one to seek for sage guidance. All contemporary acoustics manuscripts from the Naval Electronics Laboratory (NEL), later Naval Ocean Systems Center (NOSC), had to pass his scrutiny. The wise chose to seek early comment from Bob Young or chance the possibility of endless editorial notations or internal memos, oftimes pages long. For example, one quickly learned never to describe a whale's sound as ''loud'' in lieu of ''high sound source level.'' With a few examples of his pioneering studies on animal and underwater acoustics we'll revisit some of Bob Young's early landmark contributions and hear about a rare privilege of being so close to such authority, propriety and exactitude.

11:00

3aNS6. Robert W. Young and the A-weighting for acoustical measurements. George S. K. Wong (Inst. for Natl. Measurement Standards, Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

In 1985, the ASA Technical Committee S1: Acoustics, drafted the amendment ANSI S1.4A to the current sound level meter standard ANSI S1.4-1983. The aim was to provide A-weighting tolerances to frequencies above 16 kHz for measurement of short-duration transient signals. Dr. Young gave a lot of support during the preparation of that amendment. The current international sound level meter standard IEC 61672-1 (2002): Specifications, has A-weighting tolerances that allow less capable microphones to be used in Class 1 instruments. Under the above IEC standard, microphones that are not capable of measuring sound beyond 16 kHz can be used in IEC Class 1 sound level meters. Some pros and cons will be discussed.

WEDNESDAY MORNING, 17 NOVEMBER 2004 PACIFIC SALONS 4 & 5, 8:00 A.M. TO 12:00 NOON

Session 3aPA

Physical Acoustics: Materials Characterization, Scattering, and Time Reversal

Albert Migliori, Chair

National High Magnetic Field Laboratory, Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, New Mexico 87545

Contributed Papers

8:00

3aPA1. Unexpected elastic softening in delta-plutonium. A. Migliori, H. Ledbetter, and J. B. Betts (Natl. High Magnetic Field Lab., Los Alamos, NM 87545)

Elastic-constant measurements on a Pu 2.4-at.-% Ga quasiisotropic polycrystal reveal remarkable elastic softening as temperature increases from ambient to 550 K. Unexpected softening appears in both the bulk modulus B and the shear modulus G, thus in all quasi-isotropic elastic stiffnesses Cij such as the Young modulus and the constants. The dB/dT slope gives a (lattice) Gruneissen parameter $=6.0$, enormously higher than a typical f.c.c.-metallic-element value of 2.4. Especially, this high value from dB/dT measurements disagrees strongly with the value from volume measurements of -0.26. The dB/dT slope exceeds that measured previously at lower temperatures. Also, it exceeds that expected from high-temperature Debye–Waller-factor measurements. A two-level model used successfully previously to interpret this alloy's unusually low thermal expansion fails to explain the large $dCij/dT$. We comment on other possible explanations for plutonium's odd anharmonic behavior. These concepts include magnetism, $5f$-electron localization-delocalization, and vibrational entropy. Our measurements on a Pu–Ga polycrystal agree remarkably well with a Kröner-theory average of previous measurements on a same-composition monocrystal. [Work supported by the State of Florida, the National Science Foundation, and the NNSA.]

8:15

3aPA2. Wave interaction on acoustic resonators with and without hysteresis. Lev A. Ostrovsky (Zel Technologies and Univ. of Colorado, 325 Broadway, R/ET0, Boulder, CO 80305)

Nonlinear interaction of counterpropagating waves in solids is considered by using a general approach taking into account only the cumulative effects for which the corresponding terms in the equations give nonzero contribution at a wave period. Different stress–strain relations are accessed, including those with hysteresis which serve as basic models for the recent acoustic experiments with rock and metals. It is shown that the model of a ring resonator interaction considered earlier is applicable for calculating the response of a resonance bar used in the majority of recent experiments. In particular, the ratio of the hysteretic dissipation coefficient and the relative sound speed variation due to nonlinearity remains independent on the amplitude. However, the nonlinear wave distortions and, hence, harmonic generation, occurs differently from that in a traveling wave.

8:30

3aPA3. Novel approach for overcoming effects of diffraction in measurements of the nonlinear B/A parameter in liquidlike media. Kirk D. Wallace, Mark R. Holland, and James G. Miller (Dept. of Phys., 1 Brookings Dr., Campus Box 1105, Washington Univ. in St. Louis, St. Louis, MO 63130, kirk.wallace@wustl.edu)

Finite amplitude measurements of the nonlinear B/A parameter in a liquidlike medium such as soft tissue can be complicated by the diffraction pattern associated with a bounded ultrasonic source. The goal of this work is to design and implement an approach based on the use of a stainless steel delay-line to position the specimen of interest either at the last axial maximum or in the far-field portion of the diffraction pattern from a piston source. A Burgers equation enhanced, nonlinear angular spectrum simulation was used to design the experimental apparatus by modeling the ultrasonic fields. A 0.6-mm-diam hydrophone receiver was mechanically scanned in a water-bath to measure the finite amplitude ultrasonic fields emerging from the delay-lines. A thin-plate tungsten resonator was used to suppress the fundamental relative to the second harmonic. As evidence of the effectiveness of this approach for overcoming complications of data analysis arising from diffraction, the nonlinear parameter B/A, the small-signal speed of sound, and the attenuation coefficient were successfully measured over a wide range of isopropanol-water mixtures. Measurements of B/A had an average uncertainty of less than 5% over the range of values associated with biological tissues (4.6 to 11.9). [Work supported by NIH R01HL72761.]

8:45

3aPA4. The impact of the Lennard-Jones parameter choice on theoretical acoustic absorption and dispersion curves. Andi G. Petculescu and Richard M. Lueptow (Northwestern Univ., Dept. of Mech. Eng., 2145 Sheridan Rd., Evanston, IL 60208, a-petculescu@northwestern.edu)

The acoustic vibrational relaxation in polyatomic gas mixtures is modeled based on binary collisions treated via pairwise transition probabilities. The sensitivity of the model to the values of the Lennard-Jones parameters (potential well depth and collision diameter) is investigated for nitrogen–water–methane mixtures. The choice of the collision diameter has a far greater effect on the transition probabilities and subsequently on the acoustic absorption than that of the Lennard-Jones potential depth. In addition, 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 mixture composition. Thus, a binary mixture of moist nitrogen with 338 ppm water is considerably more sensitive to changes of the water collision diameter than is a mixture of nitrogen, 338 ppm water and 1% methane. [Work supported by NASA.]

9:00

3aPA5. Average grain-size (g_s) estimation in polycrystalline channels. Dov Hazony (EECS, Case, Cleveland, OH 44106, dxh2@cwru.edu), J. Lawrence Katz (Univ. of Missouri, Kansas City, MO 64108-2784), and Gerhard Welsch (Cleveland, OH 44106)

Ample experimental evidence suggests that the impulse response of a polycrystalline channel is close to the Gaussian wave-form when the number of grains is sufficiently high. Moreover, a corresponding periodic cellular simulation of the channel, where each cell contains two reactive and one dissipative discrete elements, makes it possible to associate an impulse response-$f(t)$-with the model. It will be shown that also $f(t)$ converges to the Gaussian waveform when the number of cells in the model is relatively large. Furthermore, it will be seen that, for the model to fit the channel, the cell-size must equal the average grain-size. It then follows that g_s may be calculated from the variances of the incoming and outgoing signals and the respective delay between the first two waveforms. Data will be provided and a formula will be given. [Work supported by NIH.]

9:15

3aPA6. Acoustic diffraction by deformed edges of finite length: Theory and experiment. Timothy K. Stanton and Dezhang Chu (Woods Hole Oceanograph. Inst., Dept. of Appl. Ocean Phys. and Eng., Woods Hole, MA 02543, tstanton@whoi.edu)

The acoustic diffraction by deformed edges of finite length is described analytically and in the frequency domain through use of an approximate line-integral formulation originally applied to deformed finite-length cylinders [T. K. Stanton, J. Acoust. Soc. Am. **86**, 691–705 (1989)]. In this case of edges, the factor to the integral specific to the scattering per unit length by the cylinder is replaced by one specific to edge diffraction. The formulation is applied to laboratory data involving elastic disks of finite thickness by incorporating an approximate double-edge diffraction formula for three-sided semi-infinite barriers based on modified solutions of Pierce [A. D. Pierce, J. Acoust. Soc. Am. **55**, 941–955 (1974)]. The diffraction in the backscatter direction by individual elastic disks was measured using a variety of disks spanning a range of diameters, thicknesses, and two material properties. A broadband acoustic source was used and the disks rotated over 360° in 1° increments. Given the overlap between some of the different classes of echoes, the analysis was restricted to the leading double-edge echo which was resolved over much of the range of orientation. There was generally good-to-excellent agreement between this approximate formulation and the data over much of the range of frequency and orientation angles. However, the formulation greatly overpredicted the diffraction from the thinnest disks.

9:30

3aPA7. Higher order acoustic diffraction by edges of finite thickness. Dezhang Chu and Timothy K. Stanton (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Generalized temporal and spectral solutions of acoustic diffraction by edges of finite thickness involving higher orders diffraction are presented. The solution in the time domain is an extension of Medwins double-edge second-order diffraction solution [H. Medwin *et al.*, J. Acoust. Soc. Am. **72**, 1005–1013 (1982)], which is based on the Huygens principle, while the solution in the frequency domain is an extension of the asymptotic formulas for the double-edge second-order diffraction via amplitude and phase matching given by Pierce [A. D. Pierce, J. Acoust. Soc. Am. **55**, 943–955 (1974)]. Since both solutions are approximate they have different advantages and disadvantages. The contribution from higher order diffraction depends strongly on the thickness and geometry of the edges. Comparison between the theory and experimental diffraction data demonstrates that for edges of finite thickness, the first order diffraction alone underestimates the total diffraction field significantly, especially when the flat end of the two edges directly faces the transducers (i.e., edge-on incidence). [Work supported by ONR.]

9:45

3aPA8. Analysis of sound propagation in a fluid through a screen of scatterers. Yves C. Angel (INSERM 556, 151 cours A Thomas, 69424 Lyon Cedex 03, France) and Christophe Aristegui (Laboratoire de Mecanique Physique, 33405 Talence Cedex, France)

A multiple scattering analysis in a nonviscous fluid is developed in detail in order to predict the coherent sound notion in the presence of disordered heterogeneities, such as particles, fibers, bubbles or contrast agents. Scatterers can be homogeneous, layered, shell-like with encapsulated liquids or gas, nonabsorbing or absorbing, and can take a wide variety of shapes. *A priori* imposed limitations or physical assumptions are absent in the derivation, whether they concern the expected response of the fluid-scatterer mixture, the scatterer size relative to wavelength, the scatterer concentration or the screen thickness. As in any multiple scattering formulation, a closure assumption is invoked. Closed-form results for the backscattered and forward-scattered wave motions on either side of the screen of scatterers are obtained. The scatterer–fluid mixture is shown to behave as an effective dissipative medium from the standpoint of the coherent motion. It is found that the effective medium is fully described once two parameters are determined: the effective wavenumber and the reflection coefficient for the associated half-space screen. Remarkably, both parameters depend only on the far-field scattering properties of a single scatterer. Figures illustrate the results for selected scatterers.

10:00–10:15 Break

10:15

3aPA9. Acoustical characterization of scatterer motion in a reverberant media. Stephan G. Conti (Southwest Fisheries Sci. Ctr., 8604 La Jolla Shores Dr., La Jolla, CA 92037), Philippe Roux (Marine Physical Lab., La Jolla, CA 92093-0205), Julien de Rosny (Laboratoire Ondes et Acoustique ESPCI, 75005 Paris, France), and David A. Demer (Southwest Fisheries Sci. Ctr., La Jolla, CA 92037)

It has been demonstrated that the total scattering cross section of a moving scatterer in a reverberant media can be estimated from ensembles of reverberation time series. In order to obtain an accurate estimate of the total scattering cross section, the positions of the scatterer must be uncorrelated between consecutive time series. Here we show that as the displacement of the scatterer increases between time series, the measured total scattering cross section converges asymptotically to the absolute cross section. If the positions of the scatterer are correlated between consecutive time series, the measured cross section is lower than the absolute value. Theoretical formulas for the measured cross section versus average displacement are demonstrated using simulations and experiments for a ballistic motion or a random walk. Inversion of the scattering cross section measurements are also used to estimate the average displacement of the scatterers between consecutive acoustic pings. This inversion technique is applied to experimental measurements in tanks with a standard copper sphere following a ballistic motion and live fish in random motion.

10:30

3aPA10. Ultrasonic wave propagation in particulate mixtures. Melissa L. Mather, Richard E. Challis, and Andrew K. Holmes (Ultrason. and NDE Lab., School of Electron. and Elec. Eng., Univ. of Nottingham, University Park, NG7 2RD Nottingham, UK)

Measurements of ultrasonic compression wave attenuation as a function of frequency can provide a means to monitor suspensions of microscopic particles in liquids on-line in a process setting. Such mixtures are important in many industrial sectors including chemical, mineral, pharmaceutical, agrochemical and food. The ultrasonic data can be used to estimate the dispersed phase particle size distribution and to quantify dynamic effects such as flocculation and crystallization. The technique requires a detailed mathematical model to relate the ultrasonic wavenumber to mixture properties, and there are two distinct approaches to this—*scattering* and *hydrodynamic*. Although both have been shown to adequately describe sound propagation in dilute mixtures (up to ~5% v/v), current models fail for concentrated suspensions of high density particles. The applicability of existing and future models of wave propagation in these suspensions will be greatly improved with more rigorous formulations for the *local viscosity* in the vicinity of a suspended particle, viscosity being the principal loss mechanism in such mixtures. In this paper five different theoretical formulations for local viscosity are investigated; their effects on wavenumber when included in both scattering and hydrodynamic propagation theories are compared, with reference to experimental data.

10:45

3aPA11. Distinction between adhesive and cohesive defects in a structural bond. Vlasie Valentina, Rousseau Martine (Laboratoire de Modelisation en Mecanique, Universite Paris 6, Case 162, 4 Pl. Jussieu, 75252 Paris Cedex 05, France), Champaney Laurent, and De Barros Silvio (Universite de Versailles Saint Quentin en Yvelines, 78035 Versailles Cedex)

The ultrasonic guided modes in an aluminum/epoxy/aluminum trilayer are used to characterize the adhesion and the cohesion of this structure. Different models are compared. The exact model (I) takes into account the adhesive layer; the two aluminum/adhesive interfaces are assumed to be perfect. The rheological model of the interfaces (II) consists to describe the two contact zones by a distribution of springs; the boundary conditions at the two aluminum/adhesive interfaces are given by Jones. The rheological model of the adhesive (III) consists to replace the whole adhesive layer by a distribution of springs with inertia; the Jones conditions are then written at the aluminum/aluminum interface. The theoretical comparison between the models (I) and (II) [respectively, (III)] gives the stiffnesses of the interface (respectively of the adhesive layer). For the model (II), the theory/experiment comparison allows us to define interface damage, i.e., adhesive defects. For the model (III), we define global damage in the adhesive layer, i.e., cohesive defects. Nondestructive tests are realized on samples with different levels of adhesion and any cohesive defect. The rheological models confirm an interface damage and any global damage. The same conclusions are obtained by destructive tests.

11:00

3aPA12. Damage detection in composites using the time-reversal acoustics method. Alexander Sutin, Eric Roides, and Armen Sarvazyan (Artann Labs., 1753 Linvale-Harbourton Rd., Lambertville, NJ 08530)

Time-reversal acoustics (TRA) can provide focusing of broadband frequency pulses in complex structures where different types of elastic waves can propagate. In this work, TRA focusing efficacy in the presence of mechanical damage in a material to test the integrity of various composites has been explored. The experiments were conducted on the following materials: (1) Static and cyclic loaded composite samples made of Grade 11 Garolite. (2) A Garolite plate with local damages created by an air gun bullet. (3) Epoxy/carbon fiber composites with mechanical damage. Experiment results demonstrated that a TRA focused signal passing through a damaged area has a greater level of distortion. This effect is manifested in the shape of a waveform of the TRA focused signal. In the presence of the defect, the density of the acoustic energy in the TRA focused waveform redistributes, decreasing in the center and increasing at the sides. The distortion of the TRA focused signal was significant for the composite samples affected by cyclic loading even when there were no visible defects. [Work supported by NASA.]

11:15

3aPA13. Time-reversal of waves near critical angle on a liquid–solid interface. Francois Vignon, Didier Cassereau, Mathias Fink, Jean-Francois Aubry (Laboratoire Ondes et Acoustique, ESPCI, 10 rue Vauquelin, 75005 Paris, France, francois.vignon@loa.espci.fr), Fabrice Marquet, and Pierre Gouedard (ESPCI, 75005 Paris, France)

When sending a plane ultrasonic wave towards a liquid–solid interface, the reflected wave is affected, depending on the attack angle. In particular, when ultrapassing the critical tranverse angle, the reflection

coefficient becomes complex, which in practice has the effect of dephasing the reflected wavefront. If this reflected wave is time-reversed and sent back towards the interface, the reflected wave of this time-reversed wave should not present this dephasing, as the time-reversal process restores the original phases. A theoretical and experimental study of these phenomena has been done. The dephasing of the reflected waves when passing the critical angle is observed, and as expected the signals after time-reversal are in phase. Nevertheless, a significant attenuation of the time-reversed signal is observed for attack angles around the critical transverse angle, as energy goes with the transmitted wave, which is not time-reversed. In a second part, it is shown by simulations and experiments how the fact of working at critical angle influences focusing of waves, by time-reversal or simple time-delays methods.

11:30

3aPA14. Optimization of time-reversal acoustic focusing. Alexander Sutin, Eric Roides, and Armen Sarvazyan (Artann Labs., 1753 Linvale-Harbourton Rd., Lambertville, NJ 08530)

Time-reversal acoustic (TRA) focusing systems are typically based on an acoustical resonator (a cavity) in which the multiple reflections of a bouncing acoustic pulse act as a virtual phased array. The efficiency of the TRA focusing system depends on the spectral and temporal parameters of the focused pulse and parameters of the resonator, such as its shape, size, and the relative position of its transducers. The acoustical properties of the resonator material, parameters of the sonicated medium, and the coupling of the TRA resonator and sonicated medium may also affect efficiency. Optimization of such a multiparametric system is a complex problem which is nearly impossible to achieve empirically without an adequate theoretical model. A theoretical model for predicting the efficacy of TRA focusing has been developed and experiments testing theoretical predictions were conducted. The experiments were conducted using a 10-channel TRA system developed in Artann with several broadband resonators in the frequency band of 0.1–2 MHz. The measured dependences of the focused signal on the acoustic pulse spectrum and duration, coupling of the resonator with the sonication medium, and the modes of TRA signal formation (analog or binary) confirmed theoretical predictions. [Work supported by NIH.]

11:45

3aPA15. Decomposition of the time reversal operator applied to subwavelength localization of scatterers. Jean-Gabriel Minonzio, Claire Prada, Mathias Fink, and Alexandre Aubry (Laboratoire Ondes et Acoustique, ESPCI, 10 rue Vauquelin, 75231 Paris, Cedex 05, France, jean-gabriel.minonzio@loa.espci.fr)

The decomposition of the time reversal operator (DORT method in French) is an ultrasonic detection method based the analysis of the whole pulse echo responses of an array. The eigenvalues and the eigenvectors of the time reversal operator (i.e., singular values and singular vectors of the array response matrix) provides information on localization of scatterers in the insonified medium. Recent experiments shows how DORT can also be used as a characterization method of subwavelength scatterers as cylinders [Minonzio *et al.*, J. Acoust. Soc. Am. accepted May (2004)]. It has been shown that multiple singular values are associated with a single scatterer and that the singular vectors are combinations of the normal modes (monopole, dipole) of vibrations. In this study, interactions between two subwavelength cylinders less than a wavelength apart are described. Multiple scattering affects the singular value distribution. Experiments and calculations have been carried out on nylon and metal cylinders with diameters ranging from 0.07 to 0.2 mm. The central frequency is 1.5 MHz, and the scatterers are 5 cm away from the array. The estimator multiple signal characterization (MUSIC) is used for subwavelength localization of scatterers [Prada *et al.*, J. Acoust. Soc. Am. **114**, 235–243 (2003)].

WEDNESDAY MORNING, 17 NOVEMBER 2004 ROYAL PALM SALON 5, 8:30 A.M. TO 12:00 NOON

Session 3aSA

Structural Acoustics and Vibration: Surface Ship Vibration and Noise

Raymond W. Fischer, Chair

Noise Control Engineering Incorporated, 799 Middlesex Turnpike, Billerica, Massachusetts 01821

Invited Papers

8:30

3aSA1. Surface ship underwater radiated flow noise. Christ A. F. de Jong (TNO TPD, 2600 AD Delft, The Netherlands), Leonard J. van Lier (TNO TPD), and Johan Bosschers (MARIN, 6700 AA Wageningen, The Netherlands)

An increasing focus on naval operations in littoral waters and the advances in weapon and sensor capabilities are driving an increased interest in the control of the underwater signatures of naval platforms. The technology for the control of machinery and propeller noise is well advanced, but there is a short fall in knowledge of the mechanisms that govern the underwater radiated noise due to the flow along the hull and appendages, especially for surface ships. The United States and Netherlands Navies have started a joint technology research and development program on *Mechanisms and Prediction of Surface Ship Radiated Flow Noise*, running from 2004 until 2007. This paper gives an overview of the research program and a summary of the results of the initial studies that have been carried out to date. Several possible noise source mechanisms have been identified on the basis of a literature survey and the results of a series of underwater noise measurements on full-scale frigates. The objective of the program is to develop and validate scaling laws and theoretical models for predicting flow noise levels of surface ships.

8:55

3aSA2. Bow-thruster noise remediation in ROGER REVELLE, AGOR-24. Neal A. Brown (NAB & Assoc., Inc., 5745 Friars Rd., Unit 113, San Diego, CA 92110, nab@world.std.com)

A program was initiated for Scripps Institution of Oceanography, sponsored by ONR, to silence the bow thruster in the research vessel ROGER REVELLE, AGOR-24. The ship suffered habitability problems during long-term dynamic positioning operations. A one-half-scale mockup of the ship's bow thruster, sea chest, guard grids, and local hull surfaces was constructed for air tests. By measuring velocity distributions at the (absent) impeller location while separately drawing air through the mockup system at high

velocity, it was demonstrated by calculation that the offending noise originated in unsteady cavitation of the impeller of the vertical-axis thruster. Several modifications to improve the flow were designed and evaluated in the mockup. Further, a modified impeller was designed on the basis of cavitation avoidance. The new impeller was designed-in-detail, manufactured, and installed along with several structural modifications of the thruster and sea chest, approved by the thruster manufacturer, during ship overhaul. No additional acoustic treatments were applied. Quantitative measurements of compartment levels of noise and vibration, and structural and water-borne noise were carried out to compare with premodification measurements. Results now indicate an absence of bow-thruster cavitation noise in adjacent compartments during DP operations. Thrust performance was preserved.

9:20

3aSA3. Experimental and analytical investigation of surface ship hull structural acoustics using physical models. Matthew A. Craun, Daniel P. Diperna, Gerald P. Carroll, William J. Martin (NSWCCD, West Bethesda, MD 20817, craunma@nswccd.navy.mil), Charles N. Corrado, and Bruce M. Abraham (Appl. Physical Sci. Corp., New London, CT 06355)

An experimental and analytical investigation was undertaken to characterize surface ship structural acoustics. LDV scans were made of scale models representing midhull sections of conventional and double-hull ship designs. The models were placed in a reverberant tank and excited at multiple points with tankage empty and full of water. In addition to LDV scans, hydrophones were utilized to measure radiated power. Local and global hull response for the models were examined and compared up to a frequency of 5 kHz. Hull response was further processed to determine the supersonic velocity components. Wave number analysis was conducted to characterize wave-type propagation along the hull structure. Periodic highlights were observed for wave aliasing due to bulkheads and hull stiffeners. Theoretical models were developed and exercised to develop an understanding of the experimental data. The experimental effort provides a data set to verify numerical tools. The combined experimental and theoretical modeling efforts provide a foundation to develop concepts for acoustic signature control.

9:45

3aSA4. Structural acoustic modeling issues for surface ship hull structures. Nathan Martin, K.L. Chandiramani, Richard Brown (BBN Technologies, 50 Moulton St., Cambridge, MA 02138, nmartin@bbn.com), Robert Haberman, and Robert Dees (BBN Technologies, Mystic, CT 06355)

The complex 3D character of surface ship hull structures provides a challenging configuration for detailed analytic predictions of hull response and radiation. This paper reviews key structural-acoustic response mechanisms for doubly-framed surface ship hulls and assesses the implications of these mechanisms for the prediction of hull response and radiation.

10:10–10:25 Break

Contributed Papers

10:25

3aSA5. Vibratory response of surface ships predicted by energy finite element approach. Kuangcheng Wu (Dept. of Signatures and Hydrodynamics, Northrop Grumman Newport News, 4101 Washington Ave., Newport News, VA 23607) and Nickolas Vlahopoulos (Univ. of Michigan, Ann Arbor, MI 48109)

Energy finite element analysis (EFEA) and energy boundary element analysis (EBEA) have been successfully developed and validated for a submerged plate and a cylindrical shell (Zhang *et al.*, J. Sound Vib. **263**, 21–46) and have been applied to calculate flow noise of a submerged cylindrical shell under turbulent boundary layer excitation (K. Wu and N. Vlahopoulos, ASA 143rd Meeting). Recently, new capabilities are under development to extend EFEA/EBEA formulations for surface ship applications. The presentation will discuss the developmental work and present some numerical examples in validating the development. [Work Support for this research is provided by ONR Code 334.]

10:40

3aSA6. Development of ship radiated noise transfer functions and near-field to far-field measured signature comparisons. Raymond Fischer and Leo Boroditsky (Noise Control Eng., Inc., 799 Middlesex Tnpk., Billerica, MA 01821)

Various methods exist to determine the vibration level on the wetted hull plate of a vessel—whether they be measured or predicted. Transfer functions are discussed that relate the hull vibration levels and ship's physical parameters to the far-field radiated noise. These methods are useful to predict the noise from fisheries research vessels and other vessels such as work or cruise boats. The former need to "sneak-up" on fish to perform stock assessment and the latter may need to operate in environmentally sensitive areas. Predicted radiated noise is compared to measured noise to assess the accuracy. Data is also presented on measured noise data for several vessels at various separation distances approximating near- and far-field conditions. These data were taken with the vessel at zero speed.

10:55

3aSA7. Minimization of the vibration transmission through the propeller-shafting system in a submarine. Paul Dylejko and Nicole Kessissoglou (School of Mech. and Manufacturing Eng., The Univ. of New South Wales, Sydney NSW 2052, Australia, paul.dylejko@student.unsw.edu.au)

Ships and submarines are efficient sources of underwater radiated noise. In the low-frequency range, the main sources of ship vibration and noise are the working machinery such as the diesel engines, gearboxes and generators, and the propeller/propulsion system. Rotation of the propeller results in discrete tonals at the blade passing frequency and its harmonics. In addition, in the case of a submarine, hydrodynamic forces acting on the propeller due to the fluctuating pressure field are transmitted through the propeller-shaft and thrust bearings, resulting in axial excitation of the submarine hull. A hydraulic dynamic absorber, also known as a resonant changer, is used to minimize the vibration transmission through the propeller-shafting system, in order to prevent excitation of the hull axial resonances. The present work is concerned with minimizing the vibration transmission to the hull by optimizing the virtual spring, mass, and damper parameters associated with the resonant changer in a submarine. The dynamic response of the propeller-shafting system, including the propeller shaft, thrust bearing, thrust block foundation, and resonant changer, are characterized using four pole parameters.

11:10

3aSA8. Hybrid deterministic-statistical modeling of stiffened plate structures. Rolf H. Lande, Robin S. Langley, Jim Woodhouse (Cambridge Univ. Eng. Dept., Trumpington St., Cambridge, CB2 1PZ, UK., rhl23@cam.ac.uk), and Philip J. Shorter (ESI US R&D, San Diego, CA 92130)

Statistical energy analysis (SEA) is based on equations of power balance between resonant subsystems, and, given the excitation power input to the system, the resulting average vibration level in each subsystem may be calculated. However, due to the low modal density of beams, SEA cannot fully account for energy flow along stiffening members of typical stringer-skin constructions used in the aerospace and marine industries. Thus, in a conventional SEA model of a built-up structure to which excitation is applied through a stiffening member, it is not clear how to compute the power input to each resonant subsystem and the response of the nonresonant stiffening framework. The problem is addressed by using a coupled deterministic-statistical approach: the stiffeners are modeled deterministically, while the reverberant two-dimensional structural components are modeled as power absorbing systems, in which the power is carried by statistical cylindrical waves propagating towards an energy sink. This power is then reinjected into an SEA model of the system. The numerical features of the method and some new results from a beam-plate system (compared with Monte Carlo simulations) are presented. The results indicate that the SEA power input as well as the response of the stiffener is modeled correctly.

11:25

3aSA9. Effects of fluctuating shear stresses on the low wave number wall pressure spectrum. Y.-Fan Hwang (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804–0030)

The well-known Kraichnan–Phillips theorem states that the low wave number spectrum of the wall pressure on a smooth wall beneath an incompressible and inviscid turbulent boundary layer flow is proportional to the square of wave number, and the limiting value of the spectrum is zero as the wave number approaches zero. In practice, however, measurements do not support this theoretical conclusion from ideal fluids. A theory on the possible contributions by the fluctuating viscous shear stresses to the nonvanishing pressure spectrum at zero wave number has been discussed in the literature [D. M. Chase, ''Fluctuations in wall-shear stress and pressure at low streamwise wave numbers in turbulent-boundary-layer flow,'' J. Fluid Mech. **225**, 545–555 (1991)]. Preliminary evidence to support the theory, based on the result of comparing measured near-zero wave number wall pressure and shear stress spectra, is discussed in this paper. [Work supported by ONR, Code 333.]

11:40–12:00
Panel Discussion

WEDNESDAY MORNING, 17 NOVEMBER 2004 CALIFORNIA ROOM, 9:00 A.M. TO 12:00 NOON

Session 3aSC

Speech Communication: Perception in First and Later Languages (Poster Session)

Tessa C. Bent, Chair

Linguistics Department, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon.

3aSC1. Selective influences of nonphonemic variability on early word recognition. Leher Singh (Commun. Disord., Boston Univ., 635 Commonwealth Ave., Boston, MA 02215) and Katherine White (Brown Univ., Providence, RI 02912)

Experience with words is necessarily episodic: Each encounter with a word is distinct from previous encounters of that word due to contextual or indexical factors. Nevertheless, we must recognize phonetically distinct instances of a word as lexically equivalent. The normalization problem is particularly challenging for prelexical infants who may not know which acoustic dimensions are phonemic, evidenced by findings that young infants only recognize words when words are perceptually similar across encounters (e.g., spoken by talkers of the same gender). By contrast, older infants are able to recognize dissimilar tokens of words, raising the question of why young infants consider nonphonemic acoustic detail in word recognition. In a series of studies investigating conditions that disrupt early word recognition, it was found that young infants treated as deterministic acoustic detail that may be phonemic in any given language (e.g., vowel length, pitch). Conversely, they effortlessly disregarded details that would never prove phonemic in any language (e.g., amplitude, extralinguistic noise). This suggests infants approach the normalization problem conservatively, initially ascribing relevance to all potentially phonemic details when constructing a lexicon. Later in development, infants perceive as relevant only those details that are phonemic in their language, revealing striking parallels to native phonological attunement in infancy.

3aSC2. Adapting automatic speech recognition methods to speech perception: A hidden semi-Markov model of listener's categorization of a VC(C)V continuum. Terrance M. Nearey (Dept. of Linguist., Univ. of AB, Edmonton, AB, T6G 2E7 Canada)

This study reanalyzes the categorization of 144 synthetic VC(C)V stimuli by 13 listeners. Medial silent gap duration and frequencies of pre- and postgap $F2-F3$ transitions were independently varied to cover the response set / aba, abda, ab#ba, ada, adba, ad#da/. A simple logistic model was shown to work very well for the complex response patterns [T. Nearey and R. Smits, J. Acoust. Soc. Am. **24**, 111, 2434 (2002)]. The former analysis required a static ''spoon-fed'' description of the variable duration

stimuli in terms of synthesis parameters. A new analysis will be presented using a hidden semi-Markov model (HSMM), which is a fully automatic dynamic pattern recognizer. The HSMM includes explicit state durations and is fitted to perceptual data by optimization methods allied to the maximum mutual information (MMI) approach from the automatic speech recognition literature. Given only the waveforms of the stimuli as input, the trained HSMM provides an extremely good fit to the perceptual data. The fully automatic HSMM performs better than the spoon-fed static logistic model, with rms error values of 4.8 versus 5.9 percent, respectively. The flexibility and generality of the HSMM/MMI framework for perception models will be sketched. [Work supported by SSHRC.]

3aSC3. Contrastive backward effects of nonspeech tones on speech perception. Travis Wade and Lori L. Holt (Dept. Psych., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213)

Nonspeech stimuli influence phonetic categorization, but effects observed so far have been limited to precursors' influence on perception of following sounds. However, both preceding and following speech affect phonetic categorization. This asymmetry in nonspeech and speech effects raises questions about whether general auditory processes play a role in context-dependent speech perception. Here, experiments test whether the asymmetry stems from methodological issues rather than genuine mechanistic limitations. To determine whether backward effects of nonspeech on speech may be achieved when listeners are sufficiently encouraged to incorporate later-occurring acoustic events, a series of experiments examined perception of CVC words with [da]–[ga] series onsets followed by embedded tones and one of two possible final consonants. When the final consonant was required for word identification, subjects showed clear contrastive effects; more [d]-initial words were heard with higher-frequency tones approximating a [g] third formant location, and vice versa. More limited effects were observed when subjects identified only the initial consonant and when no final consonant was present. Results are discussed with respect to a contrastive mechanism not speech specific but operating at a relatively high level, taking into account spectral patterns occurring over extended periods before and after a target event. [Work supported by NIH.]

3aSC4. Stability in perceptual assimilation: Talker and vowel context effects. Sang-hee Yeon, Ratree Wayland, James Harnsberger, and Jenna Silver (Univ. of Florida-Gainesville, Gainesville, FL 32611)

A study was conducted to examine methodological and phonetic context variables in the discrimination and identification of non-native speech sounds and the stability of these patterns across experimental conditions. Native speakers of American English were presented with two voicing contrasts produced by three Hindi speakers, [k]–[g] and [ʧ]–[ʤ]. They appeared word-initially and in three vowel contexts [a, i, u] in a forced choice identification test and in a categorical AXB discrimination test. In the discrimination test, listeners were assigned to two conditions: (1) trials in which stimuli were produced by a single talker and (2) trials in which each stimulus was produced by a different talker. Single vs multiple talkers within trials were used to determine if a stronger correlation would be observed between identification and discrimination if listeners were forced to rely on more abstract representations in long-term memory when coping with talker variability. Multiple vowel contexts were employed to examine the stability of perceptual assimilation patterns. The results of pilot studies have demonstrated a significant vowel context effect on both identification and discrimination effect as well as a significant talker effect. The implications of the results for models of cross-language speech perception will be discussed.

3aSC5. Optimizing measures of the perceptual assimilation of stop consonants. James Harnsberger, Sang-hee Yeon, and Jenna Silver (Univ. of Florida - Gainesville, Gainesville, FL 32611)

A study was conducted to examine methodological variables in the perceptual assimilation of non-native speech sounds. Native speakers of American English were presented with two voicing contrasts produced by three Hindi speakers, [k]–[g] and [ʧ]–[ʤ]. They appeared word-initially and in three vowel contexts [a, i, u]. Listeners were assigned to two groups. The first group was administered a categorical AXB discrimination test and a traditional forced-choice identification test in which stimuli were presented in isolation. The second group was administered the same discrimination test, but the identification test involved judging stimuli in the same context in which they appeared in the discrimination test (AXB identification). Pilot studies have shown a significant correlation with predicted discrimination scores based on the identification test, but the correlation was stronger between the AXB discrimination and AXB identification tests. A similar study was also conducted with Hindi listeners identifying and discriminating six Korean stop contrasts ([t–th], [t–t'], [t'–th], [k–kh], [k–k'] and [k'–kh] from one talker. These results highlight the importance of selecting optimized tasks in order to test hypotheses concerning the role of linguistic experience in speech perception.

3aSC6. Examining the role of nonadjacent phonotactic dependencies in the perception of spoken language. Conor T. McLennan, Paul A. Luce, and Robert LaVigne (Dept. of Psych., Univ. at Buffalo, Buffalo, NY 14260, mclennan@buffalo.edu)

At the previous ASA meeting we presented results on experiments investigating the learning of nonadjacent phonotactic dependencies in the perception of spoken language. Previous research had provided evidence for the learning of nonadjacent dependencies at a grammatical level (e.g., Gomez, 2002). Our research examined whether such learning was possible at a phonological level. Each learner was exposed to CVC nonword stimuli produced by one of two artificial languages. Both languages contained the same adjacent dependencies so learners could only distinguish the languages by acquiring dependencies between the initial and final Cs (i.e., the nonadjacent dependencies). Our results demonstrated that subsequent perception was facilitated for stimuli with the same nonadjacent phonotactic dependencies. In the current study, we extended this line of research by evaluating the role of nonadjacent phonotactic dependencies in nonwords that had been lexicalized. In particular, we asked whether the previous effect of facilitation would now manifest itself as inhibition, presumably due to the competition between lexicalized items sharing nonadjacent phonotactic dependencies.

3aSC7. Effects of context and frequency shifts in vowel identification. Peter F. Assmann, Catherine M. Glidden (School of Behavioral and Brain Sci., Univ of Texas at Dallas, Box 830688, Richardson TX 75083, assmann@utdallas.edu), and Terrance M. Nearey (Univ. of AB, Edmonton, AB T6E 2G2, Canada)

A high-quality vocoder (Kawahara's STRAIGHT) was used to study the effects of frequency shifts on the identification of vowels in syllables along a continuum from "bit" to "bet." Natural speech versions of "bit" and "bet" were recorded by an adult female talker in the carrier sentence, "Please say the word _again." Synthesized versions were constructed with upward and downward shifts in formant frequencies (FF; scale factors 0.9, 1.0, or 1.1) in combination with shifts in fundamental frequency ($F0$; scale factors 0.8, 1.0, or 1.2). When the carrier sentence and target syllable were both shifted, listeners reported more "bit" responses as the $F0$ shift increased and the FF shift decreased. Systematic shifts in the identification functions were also found when only the carrier was shifted. Overall, the results are consistent with a model that simulates the categorization responses of human listeners based on statistical distributions of acoustic measurements in natural speech.

3aSC8. The role of multimodal sensory feedback in the perceptual learning of speech under a novel acoustic transformation. Kevin T. Webster and Lorin Lachs (California State Univ., Fresno, 5310 N. Campus Dr. M/S PH11, Fresno, CA 93740)

This experiment examined spoken word recognition during and after perceptual learning of an unusual acoustic transformation of speech. Previous research has found that adults and children receive more benefit from cochlear implants when auditory information is effectively integrated with visual lipreading information. Because cochlear implant users have little, if any, preoperative experience with auditory information, these findings suggest that perceivers are able to capitalize on already-obtained lipreading skills in order to effectively learn to perceive new acoustic information about speech. This may be due to the fact that auditory and visual speech are lawfully related to the same physical event, a spoken utterance. Alternatively, it is possible that the addition of any relevant stimulus, regardless of its lawful relations with auditory speech, will aid in the process of learning how to deal with new perceptual information. To test these hypotheses, normal-hearing participants were trained with three different conditions of perceptual input while learning to perceive frequency-inverted speech: auditory-alone, auditory with orthographic visual stimulation, and auditory with visual information about the lips and face. The results of this study showed interesting patterns of perceptual learning and are discussed in terms of their implication for current theories of perceptual integration.

3aSC9. A phonological asymmetry and its perceptual consequence: The case of the velar nasal onset. Chandan R. Narayan (Dept. of Linguist., Univ. of Michigan, 4080 Frieze Bldg., Ann Arbor, MI 48109)

The perception of nasal consonants in onset and coda position by native English and Cantonese speakers was investigated. English exhibits a three-way nasal place contrast in coda position (/m, n, ŋ/) but only a two-way contrast in onset (/m, n/). Cantonese shows the three-way contrast in both positions. Previous perceptual investigation into the English asymmetry found that the [n]-[ŋ] contrast in onset position was difficult to discriminate for native English speakers. [Larkey *et al.*, Percept. Psychophys. **23**(4), 299–312 (1978)]. The present study investigated this asymmetry using real-speech stimuli in a cross-language experiment. In an AXB task (where *same* pairs were different tokens from the same place category and *different* pairs were tokens from different categories), both English ($n=12$) and Cantonese ($n=6$) listeners accurately discriminated all three nasal place contrasts in coda position. Crucially, Cantonese listeners accurately discriminated the [n]-[ŋ] contrast in onsets (95% correct) while English listeners were less successful (69% correct). The results are interpreted as a perceptual consequence of the asymmetry of nasal place in English, where the lack of a phonological contrast diminishes its discrimination. Future work will extend these findings to an investigation of the development of nasal place discrimination in infancy.

3aSC10. The influence of language experience on perception: The case of Spanish and English. Kathleen Currie Hall, Elizabeth Hume, Keith Johnson (Dept. of Linguist., Ohio State Univ., 1712 Neil Ave., Columbus, OH 43210-1298), and Amanda Boomershine (Ohio State Univ., Columbus, OH 43210)

Native speakers can distinguish sounds that serve a contrastive function in their language better than those that do not, but the role of allophony in speech perception is understudied. The present study tests the influence of the allophonic/phonemic distinction on ratings of perceptual similarity given by Spanish and English listeners to pairs of stimuli contrasting [d], [r], and [ð]. These sounds have different phonological relations in the two languages: [d] and [r] are allophonic in English and contrastive in Spanish, while [d] vs [ð] are allophonic in Spanish but contrastive in English. The study also investigated to what extent, if any, degree of experience/proficiency with a non-native language influences perception. The results suggest that native language but not level of language proficiency in L2 affects rated similarity. In particular, whether or not sounds are in an allophonic or contrastive relationship in the native language shapes how these sounds are perceived. [Work supported by NIDCD Grant R01 DC04421-4.]

3aSC11. Vowel discrimination in early bilinguals: Evidence of phonetic reorganization. Anders Hojen and James E. Flege (Univ. of Alabama at Birmingham, CH19 301, 1530 3rd Ave. S., Birmingham, AL 35294-2041)

The issue of whether new categories may be established for vowels found in a second language (L2) but not the native language (L1) remains controversial. The present study compared the discrimination of English vowels by native English speakers, early Spanish-English bilinguals, and Spanish monolinguals ($n=20$ each). Vowels in the control contrast (/i/-/u/) were likely to be heard as two distinct Spanish vowels by Spanish speakers, whereas the vowels in three test contrasts (/e$^{\text{I}}$/-/I/, /ɑ/-/ʌ/, /o$^{\text{ʊ}}$/-/ʊ/ were likely to be heard as a single Spanish vowel. To avoid the ceiling effects often seen in cross-language vowel discrimination research, within-trial $F0$ variation was introduced into the categorial AXB test used here. The 64 trials testing each contrast were presented in two randomized blocks differing in interstimulus interval (0 vs. 1000 ms). All three groups obtained high scores for the control contrast (means=88% to 98% correct), but only the Spanish monolinguals obtained low ($<65\%$) scores for the three test contrasts. All early bilinguals showed significantly above-chance performance ($p<0.01$ by binomial probability), but only three Spanish monolinguals did so. The results are interpreted to mean that the early bilinguals established new categories for English vowels. [Work supported by NIH.]

3aSC12. Training effects on native and novel linguistic categories. Grant L. McGuire (The Ohio State Univ., 222 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210, mcguire@ling.osu.edu)

Categorical perception (e.g., Liberman *et al.*, 1957) and the perceptual magnet effect (Kuhl, 1991) are frequently invoked to help explain how the perceptual system deals with highly variable speech input by warping the perceptual system in favor of native contrasts. Research into $L2$ acquisition demonstrates that this warping can be overcome and non-native categories can be introduced to listeners with extensive training (e.g., Logan *et al.*, 1991); further studies into novel category acquisition have shown a perceptual magnet effect for nonspeech stimuli (Guenther *et al.*, 1999). This paper addresses native English speakers' acquisition of a novel category, the velar fricative /x/, by testing the effects of category labeling and discrimination training on subjects sensory trace and context coding auditory modes. The results demonstrate training affects the novel category differently from the native categories, with labeling training creating a perceptual magnetlike effect only in the novel category. Both types of training increased subjects' abilities to discriminate categories, but mixed testing and training types (i.e., categorical testing with noncategorical training and vice versa) resulted in the greatest improvements.

3aSC13. Perception of the English tense/lax vowel contrast by native speakers of Russian. Maria V. Kondaurova (Program in Linguist., Purdue Univ., West Lafayette, IN 47907, mkondaur@purdue.edu) and Alexander L. Francis (Purdue Univ., West Lafayette, IN 47907)

The present study examined the effect of native language background on the perception of American English tense and lax front unrounded vowels ([i] and [I]). These vowels are distinguishable according to both spectral (vowel quality) and temporal (vowel duration) properties. Nineteen native Russian and 12 native American English listeners identified stimuli from a two-dimensional beat-bit continuum varying in 10 spectral and nine duration steps. English listeners used predominantly spectral quality when identifying these vowels, but also showed some reliance on duration cues. In contrast, Russian listeners failed to employ vowel quality to differentiate English tense and lax vowels which can be explained by

their assimilation to the same Russian /i/ phoneme in a ''single category assimilation'' process (Best, 1995). Instead, they relied almost entirely on the duration cue, despite the absence of distinctive vowel length in Russian. These results could be interpreted as supporting the desensitization hypothesis (Bohn, 1995): When spectral differences are insufficient to distinguish a non-native vowel contrast, duration will be employed by default. Alternatively, the strength of Russian listeners' reliance on duration may instead reflect the influence of prior experience with subphonemic duration contrasts within the lexical stress system of the Russian language.

3aSC14. An acoustic cue reliance pattern in second language speech perception. Yasuko Ito (Dept. of Second Lang. Studies, Univ. of Hawaii at Manoa, Moore Hall 570, 1890 East-West Rd., Honolulu, HI 96822, yasukoi@hawaii.edu)

The present study employed acoustic analysis to examine which acoustic cues, spectral (formant) or temporal (duration), native Japanese speakers rely on to identify four mid- and low back vowels in English as a second language. It is generally agreed that Japanese possesses five vowels, whereas English has 15 including diphthongs. This discrepancy is considered to cause confusion to the Japanese learners when they identify English vowels. Furthermore, Japanese possesses a long–short vowel contrast, which may help or cause confusion to the Japanese learners in the English vowel identification. In the present study, a perception test, which was a forced-choice identification task, was administered to Japanese learners of English, and the target back vowels used in the test were later examined acoustically. The results indicated that the learners confused the target back vowels that are temporally similar while they are spectrally different. The acoustic results, together with the patterns of confusion in identifying English back vowels, suggest that these Japanese learners of English are relying primarily on temporal cues rather than spectral cues.

3aSC15. Auditory training with phonetic variability and acoustic enhancement: A comparison of /r/–/l/ training techniques for Japanese adults. Paul Iverson, Kerry Bannister, and Valerie Hazan (Dept. of Phonet. and Linguist., Univ. College London, 4 Stephenson Way, London NW1 2HE, UK)

High-variability phonetic training (HVPT) techniques have been successful for helping Japanese adults learn the English /r/–/l/ distinction. This success has been attributed to the fact that HVPT exposes listeners to naturalistic phonetic variability, from many talkers and phonetic contexts [e.g., Logan *et al.*, J. Acoust. Soc. Am. **89**, 874–886 (1991)]. The present investigation examined whether naturalistic phonetic variability is essential for effective training. HVPT (natural recordings from ten talkers) was compared with three conditions, in which signal processing was used to alter the phonetic variability of these natural recordings: all enhancement ($F3$ set to extreme values); perceptual fading (amount of $F3$ enhancement progressively decreased during training); and secondary cue variability (variability in $F2$ and duration cues progressively increased during training). The results thus far suggest that all of these training techniques were similarly effective, improving /r/–/l/ identification by an average of 15–20 percentage points for Japanese adults. It appears that the presence of naturalistic phonetic variability may not be critical to the effectiveness of HVPT. [Work supported by ESRC.]

3aSC16. Mandarin and English listeners' perception of Mandarin tones in isolation and in context. Tessa Bent (Northwestern Univ., Dept. of Linguist., 2016 Sheridan Rd., Evanston, IL 60208, t-bent@northwestern.edu)

A comprehensive model of cross-language speech perception must account for all aspects of phonological structure. However, current models focus primarily on phonemic categories and exclude cross-language perception of prosodic structure. The cross-language interaction of prosodic categories in non-native listeners is difficult to model in current frameworks because multiple prosodic units must be considered as possible sources of transfer and interference including lexical tone, word stress, pitch accent, and phrasal intonation patterns. This study tested the hypothesis that perception of non-native prosodic contours is affected by both the relationship between the native and non-native languages' prosodic contour categories and the unit length over which the contours apply. Native English and Mandarin listeners' discrimination of the four Mandarin lexical tones in isolated monosyllables and three-syllable sequences was tested. Reactions times to correctly discriminated ''different'' trials were entered into a multidimensional scaling analysis. In the English listeners' similarity spaces, the configuration of the four lexical tones differed in the three-syllable context compared to their perception in isolation. In the Mandarin listeners' similarity spaces, the isolated tones were arranged in triangular formation with the rising and dipping tones in close proximity, but for the three-syllable stimuli, all tones were well separated. [Work supported by NIH/NIDCD.]

3aSC17. Production and perception of American English vowels in different consonantal contexts by native Japanese speakers. Takeshi Nozawa (Faculty of Economics, Ritsumeikan Univ. 1-1-1 Noj.-Higashi, Kusatsu, Shiga 525-8577 Japan), Ratree Wayland (Univ. of Florida at Gainesville), and Elaina M. Frieda (Auburn Univ.)

This is a follow-up study to our previous work on the discriminability and perception of American English vowels by native Japanese speakers. Native Japanese speakers' production of English vowels was compared to their perception. Participants were recorded saying English words and nonwords with varying CVC contexts with a delayed repetition task in order to assess vowel intelligibility. The accuracy of their productions was assessed by native English listeners to determine whether their utterances were perceived as intended. In general, vowels that were difficult for Japanese speakers to differentiate in perception were also difficult to differentiate in production. In keeping with our previous perceptual results, the effect of consonantal contexts also played a significant role in production. For instance, Japanese speakers discriminated /æ/ better when it was preceded by velar stops and their production of these same CVs received higher identification rates. Further, /≥/-/ʌ/ was a difficult perceptual contrast regardless of the consonantal context and Japanese speakers' productions of these vowels were also unintelligible. [Work supported by Grant-in-Aid for Scientific Research (C)(1)(1410635).]

3aSC18. *L*2 speakers' recognition memory for English sentences with prosodic cues. Eun-Suk Choi (Univ. of Hawaii, 1811 East-West Rd., MSC960 Honolulu, HI, eunchoi@hawaii.edu)

This study investigated $L2$ learners' recognition memory for English sentences with prosodic cues and examined the effect of English proficiency on recognition memory for English prosody. Twenty Korean $L1$ speakers and ten English $L1$ speakers participated in two tasks. The Korean $L1$ speakers were grouped according to their proficiency level: 10 each for high and low proficiency level. In a discrimination task, participants' attention was explicitly focused on the contrasts in prosody and the associated differences in meaning, followed by a recognition task that tested whether prosodic cues were used to recognize the sentences. The results demonstrated that high-proficiency $L2$ speakers of English were better at using prosodic cues to interpret spoken sentences and remembering their meaning accordingly than low-proficiency $L2$ speakers. High-proficiency $L2$ speakers showed better recognition memory of the prosodic content of English sentences than for their lexis, while low-proficiency $L2$ speakers were better with the lexical content than prosodic cues. Native speakers and high-proficiency $L2$ speakers of English did not differ in recognizing meaning differences cued by prosody. However, high-proficiency $L2$ speakers of English were significantly worse at rejecting sentences containing lexis that they had not heard than $L1$ speakers.

Session 3aSP

Signal Processing in Acoustics: Time Reversal Signal Processing

Alireza A. Dibazar, Cochair
University of Southern California, Biomedical Engineering, 2650 McClintock Street, Los Angeles, California 90089-1451

Hassan H. Namarvar, Cochair
University of Southern California, Biomedical Engineering, 2650 McClintock Street, Los Angeles, California 90089-1451

Contributed Papers

8:45

3aSP1. Performance of a multichannel time-reversal receiver design in a highly reverberative environment. James V. Candy, Brian Guidry, Christopher Robbins, Andrew Poggio (Univ. of California, LLNL, P.O. Box 808, Livermore, CA 94551), David Chambers, and Alan Meyer (Univ. of California, Livermore, CA 94551)

The development of multichannel time-reversal (T/R) communications in a highly reverberative environment was discussed previously [Candy *et al.*, J. Acoust. Soc. Am. Suppl. **115**, 2467 (2004)]. This paper focuses on the development and performance of a 1-bit receiver in the multichannel case. Here, we discuss the performance of a suite of multichannel TR 1-bit receiver realizations, similar to the multichannel case, using an acoustic 8-element T/R array and a set of client stations. The performance of these processors on both simulated and experimental data is discussed in detail. The experiment is provided by a stairwell between two floors of a noisy building. The stairwell is populated with obstructions (pipes, rails, wall, etc.) and a 90-deg landing—clearly a highly reverberative environment. It is shown that the 1-bit multichannel receivers perform quite well when compared to their full-dynamic-range 16-bit counterparts, and are able to reliably extract the transmitted code from the noisy measurements.

9:00

3aSP2. Time-reversal communication through a highly scattering medium. David Chambers, Christopher Robbins, Brian Guidry, and Ahmad Elayat (Lawrence Livermore Natl. Lab., Box 808, L-154, Livermore, CA 94551, chambers2@llnl.gov)

An ultrasonic time-reversal array system is used to transmit communication signals across an aluminum slab with 195 holes drilled through it. Multiple scattering and mode conversion stretches a 0.4-microsecond transmitted pulse into a received signal that rings for over 75 microseconds with little attenuation. Communication through such a medium is subject to severe multipath distortion. Four variations of a time-reversal receiver are used to mitigate multipath distortion and allow error-free transmission of a variety of bit sequences [see Candy *et al.* J. Acoust. Soc. Am. Suppl. **115**, 2367 (2004)]. Performance of point-to-point receivers constructed from different segments of the Green's function (channel response) is compared. [This work was performed under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.]

9:15

3aSP3. Detection enhancement using multiple time-reversed guide sources in shallow water. David C. Calvo, Charles F. Gaumond, David M. Fromm, Joseph F. Lingevitch, Richard Menis, Ellis Kim, and Geoffrey F. Edelmann (Acoust. Div. Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320)

Detection in a monostatic, broadband, active sonar system in shallow water is degraded by propagation-induced spreading. Improved detection has previously been achieved in the 3- to 3.5-kHz range by using the information in time-reversed one-way transmissions from multiple spatially separated guide sources. Detection results in the wider 500- to 3500-Hz band from an experiment south of the Hudson Canyon off the coast of New Jersey are presented in this talk. The target as well as the guide source was an echo repeater positioned at various water depths and ranges. Various subapertures of the 64-element NRL source receiver array were used to transmit 0.25-s LFMs and detect their echoes. The guide source data is analyzed to show the amount of information content depending on aperture and bandwidth. The improvement in probability of detection using multiple guide-source signals is compared with a baseline probability using matched filtering. Improvement in detection as a function of the number of guide sources and their range is also discussed. [Work supported by ONR.]

9:30

3aSP4. Experimental testing of broadband decomposition of the time reversal operator (DORT) in shallow water. David M. Fromm, Charles F. Gaumond, Richard Menis, David C. Calvo (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320), Joseph F. Lingevitch, Geoff F. Edelmann, and Elisabeth Kim (Naval Res. Lab., Washington, DC 20375-5320)

Results are shown from the application of DORT to at-sea data taken in an area south of Hudson Canyon off the New Jersey coast during Spring 2004. A vertical source/receiver array with 64 hydrophones spanning the water column of 100-m depth was operated in 500-Hz bands between 500 and 3500 Hz. Sets of four, five, or six beams were transmitted at small angles from horizontal and the response from an echo repeater was recorded on 64 channels. The resulting data were analyzed using broadband DORT. Specific transmissions were used to quantify the degradation of coherence due to temporal variability of the source/receiver, environment, and echo repeater. The isolation in the singular vectors of the echo repeater using sets of independent source beams is comparable to that from the coherence tests. The singular spectrum of the noise field is compared with the expected results from that of a random matrix. Backpropagation images based on at-sea data and propagation using *in situ* XBT measurements are shown. [Work supported by ONR.]

9:45

3aSP5. Source level measurements of a cavitation noise source in a water tunnel using a time reversal mirror. Christopher Barber (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804, dcb150@psu.edu) and Gerald C. Lauchle (Penn State Univ., University Park, PA 16802)

An experiment conducted in 2001 at the Penn State Garfield Thomas Water Tunnel demonstrated the use of a time reversal mirror to acquire acoustic measurements of a cavitation noise source. Preliminary results presented at the 142nd Meeting of the ASA (Fort Lauderdale, FL) indicated an improved capability to measure low signal-to-noise ratio (SNR) sources compared to measurements made using omnidirectional or sum-beam sensors. Final experimental results, including calibrated source level estimates, are presented and compared to results from conventional measurements. Key experimental findings regarding the capabilities and limitations of the technique, as well as applicability to acoustic measurements in other types of flow-facilities, are discussed. [Work supported by ONR Code 334, Naval Surface Warfare Center Carderock Division, and the Penn State Applied Research Laboratory.]

Session 3aUW

Underwater Acoustics: Very High Frequency [0 (100 kHz)] Boundary Interactions

Gary J. Heald, Chair
DSTL, Naval Systems, Winfrith Technology Centre, Dorset, Great Britain

Chair's Introduction—7:55

Invited Papers

8:00

3aUW1. Mechanisms of seafloor scattering: Roughness vs discrete inclusions. Anatoily N. Ivakin (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, ivakin@apl.washington.edu)

A model of high-frequency scattering from the seafloor is developed taking into account discrete inclusions in the sediments. The model allows prediction of bottom reverberation, given material parameters of the sediment and inclusions and the size–depth distribution of inclusions. The frequency-angular dependencies of the seabed backscattering strength for various types of the sediment and inclusions are calculated and discussed. An environmental data set obtained at SAX99 site (near Walton Beach, Florida), including the sediment particle size–depth distribution and the water–sediment interface roughness spectra, was used to compare contributions of volume and roughness components of the seafloor scattering. It is shown, in particular, that contribution of gravel and shell inclusions and coarse sand fraction in total scattering can be dominating (over roughness) at very high frequencies (about 100 kHz and higher) and grazing angles above critical (about 30 deg), while roughness at SAX99 site is likely a dominating mechanism of bottom scattering at lower frequencies and grazing angles below critical. A combined model, taking into account both roughness and volume discrete scattering, is shown to be a good descriptor of bottom reverberation in a wide frequency-angular range. Possibilities for inversion of various sediment parameters from backscattering data are discussed. [Work supported by ONR, Ocean Acoustics.]

8:25

3aUW2. Acoustic backscatter at very high frequencies from rough seabeds. Richard J. Brothers (QinetiQ, Bincleaves Technol. Park, Weymouth, Dorset DT4 8UR, UK), Gary J. Heald (DSTL, Winfrith, Dorset, DT2 8XJ, UK), Gary Robb, Timothy G. Leighton (Univ. of Southampton, Southampton SO17 1BJ, UK), and Justin Dix (Southampton Oceanogr. Ctr., Southampton, SO14 3ZH, UK)

Acoustic backscatter data were gathered from a variety of seabed types, both in the laboratory tank and in coastal waters. Data were gathered from sandy sediments in the tank, with a variety of characterized rough surfaces, using narrow-band pulses at frequencies between 100 and 950 kHz. Wideband data gathered at sea were obtained at frequencies between 80 and 650 kHz. Data gathered at sea from the Acoustic Range at QinetiQ Bincleaves included backscatter from the natural sandy seabed, but also scatter from several artificial sediments (sand, gravel, and pebbles) placed in a rotator. The latter equipment allowed acoustic interrogation of the same patch of seabed from multiple angles. Experimental data are compared with fluid, poroelastic, and discrete scatterer models, with a view to recommendations for the modeling of seabed backscatter in the frequency band 100 kHz to 1 MHz.

8:50

3aUW3. The small slope approximation + (SSA+) for rough surface scattering. Shira L. Broschat and Yanqiu Wang (School of Elec. Eng. and Computer Sci., Washington State Univ., P.O. Box 642752, Pullman, WA 99164-2752)

To be practical, a rough surface scattering model must be easy to implement and easily incorporated into existing models. At the same time, it must be accurate enough to give useful results. Here ''practical'' is defined to mean (1) no more than N-D integration away from low grazing angles, where N represents the dimension of the surface, (2) no more than $2N$-D integration at low grazing angles, and (3) accuracy to within a few dB. In previous work it was shown that at low forward grazing angles an approximation to the nonlocal small slope approximation (NLSSA) satisfies the first and third criteria but not the second. It was also shown that the NLSSA cross section can be written as the sum of the lowest-order SSA cross section and an additional term, the latter of which can be considered a correction term for the lowest-order SSA. In this paper the SSA+, which is based on the NLSSA but includes three approximations to the correction term, is presented. The SSA+ satisfies all three criteria for a ''practical'' scattering model at low forward grazing angles. Numerical results are presented. [Work supported by ONR.]

9:05

3aUW4. Scintillation index of ocean surface forward scattered HF acoustic signals. Benjamin Cotté, David Bradley, and R. Lee Culver (Appl. Res. Lab., Penn State Univ., State College, PA 16804)

The intensity fluctuations of surface bounce path measured during a field test near San Clemente Island in August 2002 are presented. Signals used are CW pulses with 0.14- and 1.0-ms durations and FM sweeps with 1-, 7-, 13- and 22-kHz bandwidth centered at 20 and 40 kHz. Source and receiver were separated by about 700 m with source depth ranging from 10 to 67 m and receiver depth ranging from 44 to 217 m. Data were collected over a 5-day period for low sea states (0 to 1). The dependence of the surface bounce path scintillation index on frequency, grazing angle, pulse length and bandwidth has been estimated. The results are then compared with an existing model based upon the correlation function of the ocean wave spectrum. Scatter of the data about the model is being investigated from the standpoint of other parameters. [Work supported by ONR Code 321US.]

9:20

3aUW5. Broadband hf backscattering from a smooth sand surface. Charles F. Greenlaw and D. Van Holliday (BAE SYSTEMS, 4669 Murphy Canyon Rd., San Diego, CA 92123)

Measurements of backscattering from smoothed sand surfaces at frequencies from 200 kHz to 4 MHz and grazing angles from 10° to 80° were made in a laboratory tank. The sand was a well-sorted fine sand with all organics and shell fragments removed. The surface was smoothed prior to making measurements at multiple locations using three different broadband transducers (0.5, 1.0, and 2.25 MHz) driven by an impulse. Waveforms from 32-ping coherent averages were recorded, the spectra averaged and normalized by the transducer beamwidths (as an analog for ensonified area), and joined to produce a relative backscattering spectrum for the sand surface at each grazing angle. The results suggest that scattering is mediated by sandgrain-scale roughness at high frequencies and is reasonably consistent with Lambert's rule up to near-normal incidence. At lower frequencies, the scattering strength jumps abruptly higher at grazing angles above about 50°, which is consistent with some models for volume scattering. [Work supported by ONR.]

9:35

3aUW6. Calibrated seafloor acoustic backscatter imagery with a 160-kHz multibeam sonar. Christian de Moustier and Alexandre Schimel (Ctr. for Coastal and Ocean Mapping, Univ. of New Hampshire, 24 Colovos Rd., Durham, NH 03824)

A fully calibrated multibeam swath sonar operating at 160 kHz was used to collect seafloor acoustic backscatter data in Portsmouth Harbor, NH. Acoustic backscatter levels integrated over each beam's footprint are compared with levels integrated over the sonar pulse's footprint commonly used in sidelooking sonar imagery of the seafloor. In both cases, concurrent swath bathymetry is used to convert arrival angles measured at the sonar into angles of incidence relative to the seafloor. These results are contrasted against an extensive sedimentary and acoustic database obtained in 2001 with high frequency seafloor mapping sonars O(100–500 kHz) while assembling a common data set for the 2nd International Conference on High-resolution Surveys in Shallow Water 2001 (http://www.ccom.unh.edu/shallow/shallowsurvey.htm).

9:50

3aUW7. A comparison of laboratory reflection coefficient data with predictions from a time domain model of sea floor scatter using the small slope approximation (BoRIS-SSA). Marcia Isakson (Appl. Res. Lab., Univ. of Texas, 10000 Burnett Rd., Austin, TX 78758), Gaetano Canepa, and Eric Pouliquen (NATO Underea Res. Ctr., La Spezia, Italy)

Reflection coefficient data has the potential for sediment model verification and remote sensing of sediment parameters. However, a significant component of the bistatic acoustic response is scattering from the rough interface. For example, in typical ocean conditions, surface roughness can alter the apparent normal reflection coefficient by 15 dB or more over a frequency range from 10 to 100 kHz. Therefore, it is important to quantify the contribution from scattering to determine the smooth surface reflection coefficient from bistatic data. The Bottom Response from Inhomogeneities and Surface (BoRIS-SSA) code uses the small slope approximation to calculate bistatic response from rough surfaces. BoRIS-SSA can incorporate both stochastic realizations of the ocean bottom or data from stereo-photographic equipment. In this study, bistatic laboratory data taken at 50–150 kHz from a smoothed and roughened interfaces will be compared to BoRIS-SSA calculations. The model will be considered for its ability to predict both the average value of the bistatic response and the distribution of the data.

10:05–10:20 Break

10:20

3aUW8. Backscattering by a partially submerged circular cylinder at grazing incidence. Kyungmin Baik and Philip L. Marston (Phys. Dept., Washington State Univ., Pullman, WA 99164-2814)

It is not trivial to use ray theory to describe backscattering by partially submerged or partially buried targets because in many cases of interest, Fresnel zones associated with specular points are not fully exposed and there are transitions in the number of rays when more of the target is illuminated. The Kirchhoff approximation for a partially illuminated cylinder has been used for normal incidence [P. L. Marston, J. Acoust. Soc. Am. **114**, 2302 (2003)]. The present work concerns the application and testing of the Kirchhoff approximation for a partially submerged cylinder at grazing incidence. The Kirchhoff approximation removes the singularities of elementary ray theory. For a cylinder illuminated broadside near 30° grazing angle the backscattering amplitude evolves in a nonmonotonic way as the cylinder is lowered through the air–water interface. When backscattering measurements are taken in such a way that only specular contributions are included, the evolution of many (but not all) of the observed features is accounted for. The measurements typically have ka near 13. [Work supported by ONR.]

10:35

3aUW9. Parametric synthetic interferometric sonar for bottom and subbottom imaging and mapping. Manell E. Zakharia, Sidonie Hibral, Claire Pollet, and Emmanuel Rigaud (French Naval Acad., EN/IRENAV, EA3634 BP 600, F-29240 Brest armees, France, zakharia@ecole-navale.fr)

Parametric sonar and synthetic aperture processing are commonly viewed as competing solutions for the detection of buried objects. In the framework of the SITAR project (EC funded project for the detection and characterization of buried waste in the Baltic Sea), both techniques are used concurrently in order to get the best resolution for both bottom and subbottom images. The primary frequency of the prototype is around 100 kHz, leading to a secondary one ranging from 10 to 20 kHz. The difference in penetration will be used to discriminate proud from buried targets. Two lines of receiving arrays, were used in order to provide SAS interferometric mapping. Six SAS images were thus obtained from the same spot, at the same time and with comparable resolution (two high frequency images and a low frequency one for both top and bottom arrays). Three maps were then obtained by applying interferometric processing to data issued from each frequency range. The paper will compare the images and maps obtained in the three frequency ranges and correlate a target's elevation (or burial) to its reflectivity.

10:50

3aUW10. Bayesian inference for the propagation of measurement-based model parametrizations of bottom scattering strength to reverberation prediction uncertainty. Kevin D. LePage (Naval Res. Lab., Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375) and Charles W. Holland (Penn State Univ., State College, PA)

Measurements conducted on the Malta Plateau south of Sicily and at a site north of Elba in the Tyrannian Sea are used as the basis for the construction of bottom-scattering parameter probability density functions using Bayesian inference. The volume scattering and rough surface scattering prediction capability of the OASES model [Schmidt and Kuperman, J. Acoust. Soc. Am. (1995), and LePage and Schmidt, J. Acoust. Soc. (2003)] are used to model the bottom-scattering measurments in order to facilitate the extraction of these paremeter density functions. Issues regarding the resolution and resolvability of the bottom-scattering parameters are discussed, and comparisons between Bayesian probability density functions for reverberation predictions based on the scattering measurements and measured reverberation data are made. [Work supported by the ONR Capturing Uncertainty DRI.]

11:05

3aUW11. The computation of rough surface scattering from a finite interval of roughness. John A. Fawcett (DRDC Atlantic, Dartmouth, N.S. B2Y 3Z7, Canada)

The computation of scattering from a rough, pressure-release surface is a problem of much interest in many applications. In this paper, the exact numerical solution of scattering from a finite interval of roughness is considered. This interval is a portion of an otherwise flat surface that is infinite in extent. One numerical approach is to consider a truncated infinite integral equation. However, unless one is willing to use large computational intervals, perhaps far exceeding the extent of the roughness, the solution to the truncated problem may not approximate the infinite interval solution very well for some situations. In this paper, a system of equations is derived for a closed curve including the interval of roughness and a portion of the curve in the surrounding fluid. The system is solved using the method of wavefield superposition. The method is applicable to the case of both positive and negative deformations of the flat surface. As well, the method is easily extendable to the case that the roughness is part of an upper surface of a waveguide.

11:20

3aUW12. Time reversal of interface reverberation. Joseph F. Lingevitch (Naval Res. Lab., Code 7145, 4555 Overlook Ave. SW, Washington, DC 20375) and Philippe Roux (Univ. of California—San Diego, La Jolla, CA 92093-0701)

Bottom reverberation can be used as a probe-source of opportunity to construct a time-reversal focus that scans along a rough interface. We present results from a tank experiment at ultrasonic frequencies in which reverberation from a sandy bottom was collected on a 64-element source/receiver array. The monostatic reverberation measurements are windowed, time-reversed, and then transmitted. When measureable backscatter is present in the time window, energy is observed to focus at a range corresponding to that window. Two factors contribute to the back-scattered energy level: first, the propagation paths for a particular incident ping will nonuniformally ensonify the interface and, second, the local interface scattering strength will vary with range. A method to capture reverberation at all ranges and an application to realistic shallow water ocean waveguides are also presented.

11:35

3aUW13. Influence of small-scale turbulent fluctuations on direct-path coherence from 4 to 20 kHz. Kevin B. Smith and Ozer Eroglu (Dept. of Phys., Naval Postgrad. School, Monterey, CA 93943)

A numerical study is presented with results from a sensitivity study of wave front spatial coherence. Specifically, the influence of small-scale (order of tens of meters to tens of centimeters) water-column turbulence on vertical coherence of a direct path, LFM chirp at a source–receiver separation of 500 m is examined. Parameters chosen as variable included turbulence strength, spectral cut-off scale (and combined rms value), source depth, receiver sub-array depths, and background sound speed structure. Preliminary analysis suggests that the background profile may play a more significant role than originally anticipated. The generality of this is considered in the context of measures of profile ''smoothness'' and ''curvature,'' as previously quantified by others. Potential implications include the use of background profile knowledge to arrange source–receiver geometry for optimal signal reception when coherent processing is advantageous (e.g., comms, active detection, etc.). The model results also show that the Fresnel scale of the turbulent pertubations may dominate the coherence structure, as predicted by theory. The impact of turbulence strength and frequency appear to behave as expected. [Work supported by ONR.]

Session 3pED

Education in Acoustics: Acoustics Education Prize Lecture

Uwe J. Hansen, Chair
Department of Physics, Indiana State University, Terre Haute, Indiana 47809

Chair's Introduction—1:10

Invited Papers

1:15

3pED1. Grappling with pithy problems: The education in acoustics of John William Strutt (aka Lord Rayleigh) and of the rest of us. Allan D. Pierce (Boston Univ., Boston, MA 02215, adp@bu.edu)

There is a little of Rayleigh in each of us, so we might benefit from some selective emulation, and educators might adopt some of the stimulations that contributed to Rayleigh's success. As a child, Rayleigh loved dabbling in scientific experimental projects, but his formal education was greatly influenced by the contemporary view that mathematics was a respectable alternative to the classics. When he entered Cambridge, he was ''decidedly less advanced in mathematical skills than the best of his contemporaries,'' but this situation changed, largely due to the stimulus of the Cambridge environment, to an intrinsically competitive nature, and to the influence of one of the greatest educators of all time—Edward John Routh. Rayleigh was coached to solve problems, and he excelled at this. After graduation, Rayleigh embarked on a program of self-education and developed a strategy for combining his love of experimentation with his more recently acquired problem-solving skills. Details of this self-education are related. Extensive illustrations are given of problems such as might have been presented to Rayleigh as a student, such as he might have presented to students himself, and such as might to good purpose be presented to acoustics students of today.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

Ann E. Bowles, Chair
Hubbs Sea World Research Institute, 2595 Ingraham Street, San Diego, California 92109

Chair's Introduction—2:05

Invited Papers

2:10

3pID1. Hot topics in structural acoustics: New directions and challenges. Courtney B. Burroughs, Dean E. Capone (Appl. Res. Lab., Penn State Univ., State College, PA 16804), and Karl Grosh (Mech. Eng. Dept., Univ. of Michigan, Ann Arbor, MI 48109)

Although structural acoustics is an integral part of many areas of research of interest to ASA members, it is the control of noise and vibration in marine vehicles that has underpinned much of the activities in structural acoustics. With shifts in emphasis, and therefore research funding, we face new challenges to maintain a healthy and vibrant level of activity in structural acoustics. Because of the complexities of most structural systems of interest, characterization and visualization of the acoustic responses present challenges. Multi-channel measurement systems are becoming more widely used in structural acoustics research. Numerical methods in structural acoustics continue to improve, moving up in frequency to close the mid-frequency gap. With new numerical and measurement methods, tailoring of physical and geometric properties for the control of acoustic performance are becoming more viable. The control and harvesting of energy in vibrating structural systems is also a subject of continuing research. Minimization of electronic components, e.g., microelectromechanical systems (MEMS), is presenting new challenges in structural acoustics. Biomechanics provides another domain of interest for the structural acoustician. In this paper, we will present an overview of the many challenges we feel are hot topics facing the structural acoustics community in the Society.

2:30

3pID2. Hot topics in animal bioacoustics. Andrea M. Simmons (Dept. Psych., Box 1853, Brown Univ., Providence, RI 02912)

Recent technical achievements and conceptual advances in the study of communication and orientation by sound will be the focus of this presentation. In several different species (parrots, bullfrogs, big brown bats, dolphins), the use of multiple acoustic sensors to track the movements of animals and to map complex acoustic scenes under realistic field conditions has emphasized the plasticity of vocal behaviors. Behaviors such as the feeding buzzes of echolocating bats, once categorized as stereotyped by species, have now been shown to be variable and modifiable on the individual level. Similarly, interactions between senders and receivers in natural choruses are highly dynamic, yet seem to follow specific rules. Field studies have identified the production and use of new types of acoustic signals, such as the ultrasonic sounds produced by a species of Asian frog, and the use of seismic signals by mole rats and elephants. The vestibular system also plays an important role in acoustic orientation, such as flight and foraging behaviors of echolocating bats. Understanding the interactions between auditory and vestibular systems on both behavioral and physiological levels will lead to a richer understanding of animal bioacoustics.

2:50

3pID3. Hot topics in architectural acoustics: Practice and research. K. Anthony Hoover (Cavanaugh Tocci Assoc., Inc., 327 F Boston Post Rd., Sudbury, MA 01776, thoover@cavtocci.com) and Mendel Kleiner (Rensselaer Polytechnic Inst., Troy, NY 12180)

Architectural acoustics is concerned with all aspects of sound in and around buildings. The first part of the talk will address practical and applied topics that include footfall and impact sound isolation (especially floor underlayments), health care facilities (especially MRI noise and vibration control), learning-spaces acoustics (including the recent standard, ANSI S12.60.2002), and health and safety issues of ''acoustical'' products (including duct lining and flame/toxicity issues for sound-absorptive materials). An overview will also be presented of the National Council of Acoustical Consultants (NCAC) ''Member Forum,'' which is a members-only on-line forum for discussion and new information, specifically for acoustical consulting. The second part of the talk will focus on current research issues in architectural acoustics worldwide as mirrored by recent conventions, conferences, and meetings, such as the German Acoustical Society (DEGA), the Institute of Acoustics (IoA), and the International Congress on Acoustics (ICA) with special emphasis on auralization and scattering.

WEDNESDAY AFTERNOON, 17 NOVEMBER 2004 PACIFIC SALON 3, 1:00 TO 2:30 P.M.

Session 3pMUa

Musical Acoustics: Computer Music II--Demonstrations

Julius O. Smith, Chair
CCRMA, Music Department, Stanford University, Stanford California 94305

Informal demonstration and Workshop session supplementing lecture session 3aMUa—Computer Music

3pMUa1. *RoomWeaver* and *Digitract*: Two digital waveguide mesh acoustic modeling reseach tools. Damian T. Murphy, Mark J. Beeson, Jack Mullen, and David M. Howard (Media Eng. Group, Dept. of Electron., Univ. of York, Heslington, York, UK)

RoomWeaver is a digital waveguide mesh (DWM) development environment for virtual acoustic spaces. Scripting is used to define room geometry, object, surface and material properties, together with source and receiver parameters. Different mesh types and topologies are accommodated through a user-extensible plug-in architecture with the selected DWM grown to fit the defined room at simulation run-time. Visualization options allow mesh behavior and wave propagation to be observed in 2D or 3D. Excitation signals can be selected accordingly and resultant room impulse responses can be obtained in mono, stereo, binaural or surround formats. *Digitract* is a speech synthesis research tool that uses 1D and 2D digital waveguide models to simulate the human vocal tract. 1D simulation is already well established and *Digitract* allows direct comparison with 2D DWM models that offer improved control over formant bandwidths and greater naturalness in synthesized speech. Mesh types and topology can be defined together with reflection values at lip, glottis, and, in the 2D case, inner wall boundaries. Vibrato and pitch-slide can be introduced to the input and dipthongs and tripthongs can also be synthesized. 3D graphical feedback illustrates the behavior of the acoustic pressure waves present within the tract model.

Session 3pMUb

Musical Acoustics: Voice, Pitch and Perception

Diana Deutsch, Chair
Department of Psychology, University of California, San Diego, La Jolla, California 92093

Contributed Papers

1:30

3pMUb1. Perceptual grouping of musical sequences: Pitch and timing as competing cues. Kamil Hamaoui and Diana Deutsch (Dept. of Psych., Univ. of California, San Diego, La Jolla, CA 92093)

A series of experiments was conducted that explored the roles of pitch and timing cues on perceptual grouping of musical patterns, in which tones were presented at rates slower than those giving rise to stream segregation. Sequences of 12 tones were constructed in which a pitch cue (size of pitch distance between adjacent tones) and a temporal cue (size of temporal gap between adjacent tones) were set in opposition to each other. Either the pitch cue suggested groupings of three and the temporal cue suggested groupings of four, or vice versa. Listeners judged for each sequence whether they heard groupings of three or four tones. When the sequences traversed a semitone scale, grouping judgments were made on the basis of timing in a remarkably fine-grained fashion, and the effects of timing were related in an orderly fashion to the size of pitch distance cue employed. However, when the sequences possessed a hierarchical pitch structure, grouping on the basis of pitch dominated judgments instead. The effect of hierarchical pitch structure was substantial, even when compared with control sequences in which parameters such as interval size and contour were held constant. Implications of these findings for music perception and performance are discussed.

1:45

3pMUb2. The acquisition of absolute pitch for the mainstreamed, special educational needs and academically talented under Lau Chiu Kay Music Educatherapy. C. Kay Lau (Hong Kong Music Home for Handicapped Normal Talented Children Ltd., 1/F Hong Lok House, 475-475A Nathan Rd., Yaumatei, Kowloon, Hong Kong)

Absolute pitch (AP) is an ability to identify or produce tones without references. It is rare among musicians. This study examined the effectiveness of Lau Chiu Kay Music Educatherapy in training students to develop AP and identified factors that exerted significant influences. Three academically talented, 103 mainstreamed and 38 special educational needs students at Music Home, with an age range of 2 to 25, constituted the sample. It was a case study incorporating a one-group-pretest-post-test quasi-experiment. An AP assessment and a questionnaire were utilized. Results showed that (a) Lau Chiu Kay Music Educatherapy helped all subjects to develop AP. (b) Singing absolute solfge, accurate tuning, interest and attention were influencing factors, whereas the gender, aptitude, age of onset, language background, acute memory, auditory sensitivity, pitch identification practice, inheritance, tinnitus and chromesthesia were not. (c) AP could be developed within one to two months and during the starting grade of the piano learning. (d) The verbal encoding strategy was the major processing strategy, etc. Following theories can be suggested: (a) AP is innate. (b) AP is a successful outcome of the tonal-absolute solfge with the octave designation labelling association. (c) AP may have a sensitive period. (d) AP is developmental etc.

2:00

3pMUb3. Tone language and absolute pitch: Prevalence among American and Chinese conservatory students. Diana Deutsch (Dept. of Psych., Univ. of California, San Diego, La Jolla, CA 92093, ddeutsch@ucsd.edu), Trevor Henthorn (Univ. of California, San Diego, La Jolla, CA 92093), Elizabeth Marvin (Eastman School of Music, Rochester, NY 14604), and HongShuai Xu (Capital Normal Univ., Beijing, 10037, China)

Absolute pitch, defined as the ability to name or produce a musical note of particular pitch without benefit of a reference note, is extremely rare in the U.S. and Europe; this rarity has so far been unexplained. This paper reports a substantial difference in the prevalence of absolute pitch in two normal populations, in a large-scale study using direct, on-site testing, without self-selection from within the target populations. The subjects were students in two major music conservatories in the U.S. and in China. The Chinese subjects ($n=88$) all spoke the tone language Mandarin, in which pitch is used to convey the meaning of words. The American subjects ($n=115$) were all nontone language speakers. For both groups, the earlier the age of onset of musical training, the greater the prevalence of absolute pitch; however, its prevalence was far greater among the Chinese than the U.S. group, for each level of age of onset of musical training. The findings indicate that the potential for acquiring absolute pitch is universal at birth, and can be realized by enabling infants to associate pitches with verbal labels during the period in which they acquire the features of their native language.

2:15

3pMUb4. Fluctuations as a measure of frequency ratios in equally tempered scales. Alpar Sevgen (Dept. of Phys., Boğaziçi Univ., Bebek 34342, İstanbul, Turkey)

Complexity analysis of equally tempered scales (ETS) and the dynamic models that mimic this analysis indicate that tones tend to stay as far apart from each other as possible in scales employed in music. A quantitative measure of *distances* (frequency ratios) between the tones is therefore needed. For ETS with N semitones, M tones and interval structures $\mathbf{n}=\{n_1,n_2,\ldots,n_M\}$, where n_i is the number of semitones between adjacent tones, the average distance between tones is $N/2$, independent of $\mathbf{n}$. The variance $\sigma(\mathbf{n})$ indicating the magnitude of fluctuations around the mean distance provides such a measure. Scales related by cyclical permutations and mirror imaging have the same variance. Complementary scales (e.g., major $\leftrightarrow$ pentatonic) have their variances related, $\sigma^2(\mathbf{n})=|a|\sigma^2(\mathbf{n}_C)+b$, where a and b are constants independent of the scale structure, and variances of complementary scales behave in tandem. Thus the smaller the variance the greater is the distance among the tones. Increased fluctuations measure the amount of bunching. Numerical studies of variances agree closely with the results of the complexity analysis, namely that the major scale, the pentatonic scale and the modes have their tones placed as far apart from each other as possible.

2:30

3pMUb5. Measurement and analysis of Byzantine chant frequencies and frequency intervals. Kyriakos Tsiappoutas, George E. Ioup, and Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148)

A Byzantine music chant piece performed by a well-recognized chanter is used in order to measure experimentally the mean frequencies of the first five tones (D–A) of the Byzantine music diatonic scale. The experimental frequencies are compared with two traditional theoretical scales. It is found that a scale performed by a traditional chanter is very close to the frequencies of the proposed scales, except for tone F. The performer's standard deviation from the mean frequencies for each tone is determined. The amount of deviation is not provided by theory. By comparison of these results to the notion of pitch discrimination from psychophysics, it is further established that the frequency differences are small not only physically (frequency), but also in terms of perception (pitch). The attraction effect (secondary tones are higher in frequency when found between a lower and a higher main tone and lower in frequency when between a higher and a lower main tone) is tested for all analyzed tones (D–A); it is found that the sole tone subject to the attraction effect is the only secondary tone (E). The frequencies characterizing the attraction effect have not been explained in theory and are quantified experimentally for the first time.

2:45

3pMUb6. Constant interval distance between $R1$ and $R2$ of Spanish vowels sung at high pitch by sopranos. Eduardo Castro-Sierra, Miguel Zenker (HIMFG/ENM-UNAM, Xicotencatl No. 126, 04100 Mexico, D.F., Mexico), Adrian Poblano (InCH, CNR, Mexico, D.F., Mexico), and Alberto Ramirez-Trevino (CINVESTAV, Mexico, D.F., Mexico)

Recently, Joliveau *et al.* [Nature **427**, 116 (2004)] have studied the frequency distance between $R1$ resonance frequency and $f0$ of sopranos singing in English. Whereas at low-pitch singing this resonance frequency was roughly constant, at high-pitch singing $R1$ followed $f0$. This tendency held for vowels without lip rounding but not for those with lip rounding. Using CSRE45, at National School of Music of UNAM in Mexico City, $R1$ and $R2$ of the five tonic vowels (/a/, /e/, /i/, /o/, /u/) of Spanish words extracted from a traditional Mexican children's song spoken or sung by 11 native-speaking sopranos with an average of 5 years of classical operatic training were analyzed. While speaking or singing at low or mid pitch, either resonance frequency fell within the distribution characteristic of each vowel, forming clusters. However, when singing at high pitch, there was a constant interval difference, irrespective of singer, between $R1$ and $R2$ of every vowel. Since it has a simple vowel system, and the frequency separations between formants of different vowels are relatively large, sopranos singing at high pitch in Spanish appear to easily maintain the interval difference between $R1$ and $R2$, and thereby the clarity of the vowels.

WEDNESDAY AFTERNOON, 17 NOVEMBER 2004 ROYAL PALM SALONS 3 & 4, 2:00 TO 3:00 P.M.

Session 3pSP

Signal Processing in Acoustics: Signal and Image Processing: Applications

Hua Lee, Chair
Department of Electrical Engineering, University of California, Santa Barbara, California 93106-2991

3p WED. PM

Contributed Papers

2:00

3pSP1. Application of dynamic synapse neural networks on identification and localization of nonspeech sounds. Alireza A. Dibazar and Theodore W. Berger (Univ. of Southern California, 3650S McClintock Ave., OHE500, CA 90089-1451)

This paper focuses on the dynamic synapses neural network (DSNN) for identification of nonspeech sounds, including the chambering of a gun, as well as localization of identified sounds. The algorithm employed consisted of extracting DSNN features from sounds and classification of features based on Gaussian mixture models (GMMs). To extract DSNN features, a single neuron with a presynapse including a 14th order differential equation, first order post-synapse, and first order inhibitory feedback was used. After training, network parameters were used as features. The classification task was then formulated as an estimation of conditional joint probability. Classification results were collected from both chambering identification and localization. Successful localization was defined as correct identification of speaker of origin. The number of training and testing samples were 120 and 160, respectively. The system performed 96.88% and 90.00% correct identification and localization of test samples.

2:15

3pSP2. Classifying sidescan sonar images using self organizing maps. Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, jioup@uno.edu), Marlin L. Gendron, Maura C. Lohrenz, Geary J. Layne (Naval Res. Lab., Stennis Space Ctr., MS 39529), and George E. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

Self organizing maps (SOMs) can be used for computer-aided classification of objects found in two-dimensional snippets of sidescan sonar images. SOMs are briefly discussed, including the choice of features or attributes as well as various types of input data. The inputs can be, for example, the data values themselves, either raw or processed images; the amplitudes of the Fourier transform coefficients of the data; the wavelet transform coefficients of the data; the energies of the horizontal, vertical, or diagonal wavelet coefficients; the autocorrelation of the data; the Hartley transform coefficients of the data; the cepstrum; the dimensions of the object; or the sonar bright spot and shadow character. Tabular results and two-dimensional maps showing the groupings of measured and processed sidescan data are presented. Comparisons are made with human classifications of the same images. [Research supported in part by NRL-ASEE Summer Faculty Research Program.]

2:30

3pSP3. High-resolution three-dimensional imaging with synthetic-aperture forward-looking sonar systems. Hua Lee (Dept. of Elec. and Comput. Eng., Univ. of California, Santa Barbara, CA 93106, hualee@ece.ucsb.edu)

The concept of synthetic-aperture imaging is to utilize the scanning of data-acquisition transducer(s) to create a larger aperture span for resolution improvement. In the area of underwater sonar imaging, the synthetic-aperture imaging modality has been successfully applied to side-looking sonar systems to improve the cross-range resolution. Forward-looking sonar systems are designed to operate in simple range-profiling mode mainly for obstacle avoidance. Typically, the forward-looking systems are equipped with a physical transducer array, linear or circular. The image formation procedure produces an instantaneous two-dimensional range map converted from the time-delay profile of the echoes. As a result, the three-dimensional landscape is compressed into a two-dimensional range profile. Consequently, the depth information is lost, which seriously compromises the effectiveness of the system's capability. In this paper, the synthetic-aperture technique is applied to the forward-looking systems. The backward propagation image algorithm is applied for three-dimensional image reconstruction. With motion compensation procedure, similar to that in the side-looking synthetic-aperture imaging, high-resolution three-dimensional dynamic imaging capability can be achieved. In addition to the retention of the depth information, the resolution of the image profiles is greatly improved with dynamic updating. [Research supported by the UC MICRO program and Sonatech.]

2:45

3pSP4. Digital algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications. Kelly Fitz (School of Elec. Eng. and Computer Sci., Washington State Univ., Pullman, WA 99164-2752) and Sean A. Fulop (Cal. State Univ., Fresno, CA 93740-8001)

The *channelized instantaneous frequency* of a signal $x(t)$ is $\mathrm{CIF}_x(\omega,T)=(\partial/\partial T)\arg(F_h(\omega,T))$, where F_h is the short-time Fourier transform of $x(t)$ using window function h. The *local group delay* of a signal is $\mathrm{LGD}_x(\omega,T)=-(\partial/\partial\omega)\arg(F_h(\omega,T))$. For each point $F_h(\omega_0,T_0)$ in the STFT, the $f-t$ coordinates $[\mathrm{CIF}(\omega_0,T_0),t-\mathrm{LGD}(\omega_0,T_0)]$ pinpoint the local mean of the Rihacek distribution of complex signal energy, and this reassignment of the STFT magnitude yields a spectrogram from which a lot of blurriness is removed. Two algorithms which compute such a reassigned spectrogram are exemplified and evaluated. One is based on the theory of Nelson ["Cross-spectral methods for processing speech," J. Acoust. Soc. Am. **110**(5), 2575–2592 (2001)]; the second implements the equations of Auger and Flandrin ["Improving the readability of time-frequency and time-scale representations by the reassignment method," IEEE Trans. Signal Process **43**(5), 1068–1089 (1995)]. The empirical performance of each technique is qualitatively evaluated; both methods dramatically improve upon the standard spectrogram, and also surpass the naive benchmark provided by first-difference approximation of the phase derivatives. High-resolution spectrograms are provided for both music and speech signals; e.g., the time-frequency nature of a single glottal pulsation can be observed. Applications to sound morphing are also demonstrated, e.g., morphing voices and musical instruments.

Plenary Session, Business Meeting and Awards Ceremony

William A. Kuperman, President
Acoustical Society of America

Business Meeting

Presentation of Certificates to New Fellows

Michael A. Ainslie
Klaus Genuit
Robert F. Gragg
K. Anthony Hoover
Patricia A. Keating
Volker Mellert
Robert E. Remez
Philippe Roux
Shihab Shamma
Yoiti Suzuki
Stephen C. Thompson
Manel E. Zakharia

Presentation of Awards

Science Writing Award in Acoustics for Journalists to Ian Sample

Science Writing Award for Professionals in Acoustics to Stephen C. Thompson

Rossing Prize in Acoustics Education to Allan D. Pierce

Student Council Mentoring Award to David T. Blackstock

Silver Medal in Acoustical Oceanography to D. Vance Holliday

Silver Medal in Biomedical Ultrasound/Bioresponse to Vibration to James G. Miller

Silver Medal in Engineering Acoustics to John V. Bouyoucos

Honorary Fellowship to Walter V. Munk

Session 4aAA

Architectural Acoustics, Noise, Psychological and Physiological Acoustics, Speech Communication and Committee on Standards: Implementation of Classroom Acoustics I

Michael R. Stinson, Chair
National Research Council, Institute for Microstructural Science Montreal Road, Ottawa, Ontario K1A OR6, Canada

Chair's Introduction—7:40

Invited Papers

7:45

4aAA1. Forging improvements in classroom acoustical quality. Angelo Bellomo (OEHS, Los Angeles Unified School District, Los Angeles, CA 90017)

Major investments are currently being made in school infrastructure. In Los Angeles alone, 90 new schools are to be constructed in the next 10 to 15 years, and hundreds of existing schools will undergo major renovation. Similar efforts are underway in large urban school districts throughout the country. This presents a unique opportunity to improve acoustical quality in classrooms for future generations of teachers and students. While some school districts are embracing this opportunity, there are substantial challenges to be faced before improvements in classroom acoustical quality are widely realized. Despite recent adoption of the ANSI standard, there are inconsistent views among school administrators and designers on the extent of the problem and the urgency to act. Additionally, there is insufficient data on the incremental costs of achieving improved acoustical quality in new construction or modernization projects. Meanwhile critical design and purchasing decisions are underway. There are promising examples of how school districts can successfully meet these challenges and forge improved acoustical quality in the learning environment. The focus of this discussion is on the use of policy reforms to drive better design and purchasing decisions, and how these decisions can be further influenced by community involvement and interagency collaboration.

8:05

4aAA2. Experiences implementing ANSI S12.60 in classrooms. R. Kring Herbert and John Erdreich (Ostergaard Acoust. Assoc. 200 Executive Dr., W. Orange, NJ 07052)

Issues in meeting the ANSI 12.60 classroom acoustics standard fall into two categories: overcoming design team unwillingness to implement recommended criteria and finding practical methods of meeting issues of reverberation control and HVAC noise control. Solutions to HVAC noise control are emerging; an example is the use of fan coil units in new construction. Education of owners and designers is an ongoing process. It is hoped this paper contributes to a dialogue between the parties involved in school construction.

8:25

4aAA3. Source attenuating HVAC equipment—Just the facts. Arthur Hallstrom (Trane, American Standard, 1515 Mercer Rd., Lexington, KY 40511)

Current classroom designs range from NC 63 to NC 25. The ASA/ANSI 12.60 Classroom Standard sets a relatively low background sound requirement for classrooms. Achieving 12.60 levels economically requires the maximum in source sound attenuation. This paper will provide a manufacturer's perspective on current and low noise HVAC products and the resulting classroom sound levels. Predictive acoustics software will be used to predict the space sound power level and some indication of equipment cost will be provided in order to assist designers in what is the best approach to low noise classroom design.

8:45

4aAA4. Case study of a prototype elementary school designed to meet the ansi classroom acoustics standard. Gary Siebein, Robert Lilkendey, Hyun Paek, and Stephen Skorski (Siebein Assoc., Inc., 625 NW 60th St., Ste. C, Gainesville, FL)

The school board standards group that prepares design criteria for a large district, an architectural firm specializing in educational environments, a contractor and an acoustical consultant formed a team to develop a series of bid alternates for a prototype elementary school to upgrade the acoustical performance of the school to meet the ANSI classroom acoustics standard. Three primary areas of work included room finishes to reduce reverberation times to 0.60 s or less; selecting wall assemblies and details to reach the STC 50 required in the standard to reduce sound transmission between rooms; and HVAC system design to reduce noise produced by the air-conditioning system to the 35 dBA required in the standard. Only minor adjustments had to be made to the finish materials and wall assemblies in the prototype school design to meet the requirements of the standard. Major changes in design concept and implementation had to be made to the HVAC system design to meet the 35 dBA requirement. The paper describes the prototype design, the design processes used to develop the modifications and the cost impacts of the proposed changes.

9:05

4aAA5. Ten years of classroom acoustical design. Angelo J. Campanella (Campanella Assoc., 3201 Ridgewood Dr., Columbus (Hilliard), OH 43026)

Bradley found [J. Acoust. Soc. Am. **80**, 837–845] with word score tests that S/N and room reverberation time (RT) affect speech intelligibility, and that computed U50 [J. Acoust. Soc. Am. **80**, 846–854] was maximum for rooms having RT$<$0.6 sec and quieter than 35 dBA. This was applied to teleconferencing rooms [J. Acoust. Soc. Am. **111**, 2411(A)], and classrooms for distance learning [J. Acoust. Soc. Am. **107**, 2861(A)]. Economic strategies to reduce classroom RT and HVAC noise were developed. RT from 0.5 to 0.7 s and NC less than 30 are economically achieved: Absorption is provided by acoustical ceiling tiles of NRC$>$0.6. Carpets help. Chalk and cork boards provide some wall absorption, while in-use posters and book storage add more later. Central air handling systems promote NC-25 HVAC design. Noisy supply (VAV) air boxes are placed in the corridor ceiling. The transmitting manifold is lined. A flexible duct link to diffusers provides more noise attenuation at no added cost. Larger supply air diffusers satisfy NC-25. Common ceiling return air plenums compromise noise isolation between classrooms. High CAC ceiling tiles and return air openings near the room center provide NIC$>$40 versus recommended NIC50. Sketches and some cost estimates are provided.

9:25

4aAA6. Using computer modeling and auralization to provide classroom listening experiences. Robert C. Coffeen (School of Architecture and Urban Design, Marvin Hall, The Univ. of Kansas, Lawrence, KS 66045)

Computer modeling and computer auralization will produce comparative listening experiences for architects, building owners, and building users based on selected interior surface shapes, interior finish materials, and ambient noise levels. Will these techniques provide easily understood audible information that will assist classroom designers, school administrators, and teachers in making proper assessments regarding classroom acoustics? Auralizations for both good and bad classrooms, based on ANSI S12.60 2002, will be presented for comparative listening.

9:45–10:00 Break

10:00

4aAA7. Acoustical renovation of portable classrooms for cochlear implanted pupils. David Lubman (14301 Middletown Ln., Westminster, CA 92683, dlubman@ix.netcom.com) and Tim Lopez (Riverside County Office of Education, Riverside, CA 92502)

Anticipating increased enrollment of children with cochlear implants, the Riverside County Office of Education undertook the acoustical renovation of two portable classrooms in California's Riverside County: Wildomar Elementary School in Wildomar and Katherine Finchy Elementary School in Palm Springs. The aim was to improve the acoustical environment to make the classrooms suitable for pupils with severe hearing disabilities (reduce noise and reverberation). Lacking an acoustical standard specific for cochlear implanted pupils, the acoustical goals chosen were those of the ANSI acoustical standard S12.60-2002. Key changes were intended to reduce interior noise from HVAC and improve the facade insulation to reduce exterior noise intrusion. The resulting acoustical improvements are documented. Costs and lessons learned are discussed.

10:20

4aAA8. Classroom acoustics in a Leadership in Energy and Environmental Design (LEED) certified school. Felicia M. Doggett (Metropolitan Acoust., LLC 40 W. Evergreen Ave., Ste. 108, Philadelphia, PA 19118, felicia@metropolitanacoustics.com)

Clearview Elementary School in Hanover, Pennsylvania, is the first school in that state to achieve LEED (Leadership in Energy and Environmental Design) certification. Additionally, it is only one of three K–12 schools nationwide to achieve certification at the Gold Level. A study of the classroom acoustics has been conducted in this school to evaluate how well a Gold Level LEED-certified building compares to criteria set forth in ANSI S12.60-2002. Also addressed is the success of implementing this standard in schools in the greater Philadelphia area and New Jersey, and the receptive attitude that architects in this region have for the standard. Both Philadelphia and the State of New Jersey are in the midst of a large effort to update old school facilities and build new ones.

10:40

4aAA9. Classroom acoustics and architects: Sightlines and sound. Terence J. Williams (Terence Williams Architect Inc., 102.2957 Jutland Rd., Victoria, BC V8T 5J9, Canada)

All too often, acoustics in classroom design is a forgotten or ignored subject. Design teams consider structural, mechanical and electrical implications but overlook the fundamental issue of hearing and being heard. As an architect with an interest in architectural acoustics, particularly performance arts venues, we have taken a pragmatic approach, some might say unscientific, to classroom design. A series of case studies is presented that illustrates the challenge of classroom acoustics from an architect's point of view: not ''what does it look like,'' but ''how does it sound''! Because classroom acoustics are so closely aligned to the visual presentation of information, seeing and being seen becomes critical also. Sightlines and sound continue to vie for attention. The challenge is making it work.

11:00

4aAA10. Classroom acoustics and ANSI S12.60: A British perspective. Peter A. Mapp (Peter Mapp Assoc., Copford, Colchester, Essex, UK)

At the same time that ANSI S12.60 was being developed, UK recommendations for classroom acoustics were also being drawn up (Building Bulletin BB 93). The resulting standards turned out to require similar goals and set very similar criteria (e.g., 0.6 s RT). The British document however, also set a criterion for intelligibility of 0.6 STI for open-plan classrooms and teaching areas. Apart from making some general comparisons between the standards, the paper looks at the practicality of measuring STI in classrooms. The practical problems of making measurements under occupied conditions as well as the influence that the sound source itself can have on the measurements are discussed. An empirical relationship between STI and the reverberation time of a classroom is established which may be of use to designers. The effects of source directivity on the resultant STI are discussed and measurement data are presented that compare different loudspeaker types including omnidirectional (dodecahedral array), single-cone units, and head/mouth simulators.

11:20

4aAA11. The acoustic design of schools in England and Wales. Bridget M. Shield (Dept. of Eng. Syst., Faculty of Eng., Sci. and Built Environ., London South Bank Univ., London SE1 0AA, UK, shieldbm@lsbu.ac.uk)

Legislation governing the acoustic design of schools was introduced in England and Wales in July 2003. The acoustic performance standards which must be achieved are published by the Department for Education and Skills (DfES) in ''Building Bulletin 93: Acoustic Design of Schools.'' The UK government has recently announced plans for a major school buildings program; over 200 new secondary schools are to be built and many existing schools are to be refurbished. The DfES has published designs for ''exemplar schools'' by many of Britain's leading architects. However, current trends in architectural design such as the use of highly reflective materials and the introduction of large open plan spaces make it difficult to comply with the specified acoustic performance standards. There are also conflicts between the acoustic and other requirements, such as the desire for natural ventilation in classrooms. This paper outlines the regulations, presents data to show that many existing schools in England do not comply with previous or current standards, and gives examples to illustrate the effects of various acoustic treatments of classrooms. Issues concerning implementation of the regulations are discussed and examples of ''good'' and ''bad'' acoustic designs of schools are given.

11:40

4aAA12. A challenge to the Acoustical Society of America and its members to help accelerate adoption of the American National Standards Institute (ANSI) standard for classroom acoustics. Louis Sutherland (27803 Longhill Dr., Rancho Palos Verdes, CA 90275-3908), David Lubman (Westminster, CA 92683), and John Erdreich (Ostergaard Acoust. Assoc., West Orange, NJ 07052)

The ANSI Standard on classroom acoustics, S12.60-2002, has the potential to greatly improve the quality and accessibility of public education. To realize its potential the standard must be put to use. The standard is now voluntary unless it is adopted by a state or school district or is incorporated into individual school construction bidding documents. In some parts of the US, the ANSI standard now gets consideration for construction of new or substantially upgraded schools. But wide scale adoption may take many years at the current rate of progress. To accelerate adoption of the ANSI standard in school construction a major effort is required to educate stakeholders. ASA and its members can play an important role in this endeavor by educating stakeholders, including architects, school building officials, and parent–teacher groups through workshops, seminars and other presentations. Consistent with ASA's mission, to increase and diffuse the knowledge of acoustics and promote its practical application, ASA and its members can become ambassadors for good classroom acoustics in our communities. Some of the existing materials are described that can be used for this vital objective to help remove acoustical barriers to learning.

Session 4aAB

Animal Bioacoustics: Acoustic Monitoring of Large Whales--The Legacy of Paul Thompson I

John A. Hildebrand, Chair
Scripps Institute of Oceanography, University of California, San Diego, Mail Code A-005, La Jolla, California 92093-0205

Chair's Introduction—7:55

Invited Papers

8:00

4aAB1. Appreciated: Forty years with Paul O. Thompson. William C. Cummings (5948 Eton Ct., San Diego, CA 92122-3203, oshundbs@san.rr.com)

As longtime professional associate, ASA Fellow, and close personal friend, I have pleasured in knowing Paul O. Thompson for 40 years. In 1965 he was first to join my newly established Applied Bioacoustics Branch at NEL. He authored or co-authored 30 papers and reports on the sounds and related behavior of marine mammals in the North and South Pacific and South Atlantic oceans. Educated at St. Olaf College, MN, and USC, Paul's earliest research resulted in 36 papers and reports on psychoacoustics often in collaboration with John Webster or Bob Gales. (Along with Past President Bob Young, we all were co-workers in the Listening Division headed by Past President Gales.) Ever unassuming, Paul is noted for stubborn devotion to detail that often uncovered new findings in seemingly rote data. Unmatched were his aural memory and skills at analog sound recording and processing. Paul could ferret and dissect whale signals among the ambient like very few contemporaries. Even after ''retirement,'' his research in the Sea of Cortez markedly added to earlier findings of 20-Hz and other low frequency utterances from blue and fin whales. This presentation focuses both on Paul and his contributions that are sure to permanently mark the literature on marine mammal bioacoustics.

8:25

4aAB2. The study of whales with sound: the legacy of Paul Thompson. John A. Hildebrand (Scripps Inst. Oceanogr., Univ. Calif.—San Diego, La Jolla, CA 92093)

Passive acoustic monitoring of whales offers a complementary technique to traditional visual survey for assessment of whale populations. The development of whale acoustic monitoring techniques has proceeded over the last several decades. The work of Paul O. Thompson stands out for his pioneering contributions to the development of acoustic monitoring by: (1) identifying the variety of sounds produced by marine mammals and their mapping by species, (2) recognizing temporal patterns of sound production both on diel and seasonal cycles, and (3) demonstrating the value of long-term monitoring. The contributions of Paul Thompson will be discussed, and their connection to on-going work will be illustrated. The legacy of Paul Thompson's work is that acoustic population estimation techniques are now being developed that offer the potential for efficient and economical monitoring of whale populations.

8:50

4aAB3. Recording blue whales in the North Pacific: Building on the legacy of Paul Thompson. Kathleen M. Stafford (Natl. Marine Mammal Lab, 7600 Sand Point Way NE, Seattle, WA 98115)

All of the well-documented blue whale call types for the Pacific were first recorded and described by Paul Thompson. Early work by Thompson and colleagues W. C. Cummings and W. A. Friedl, in particular, laid the foundation for what we know today about blue whale sounds. Among these contributions is the first published confirmation of blue whale sounds recorded off Chile [W. C. Cummings and P. O. Thompson, J. Acoust. Soc. Am. **50**, 1193–1198 (1971)]. These sounds have since been recorded seasonally in the eastern tropical Pacific, suggesting a possible migration route. Thompson's early work with Naval data not only provided the first description of northeast and northwest Pacific calls, but also documented diel calling patterns [P. O. Thompson, US Naval Elec. Lab. Rep. (1965); P. O. Thompson and W. A. Friedl, Cetology **45**, 1–19 (1982); W. C. Cummings and P. O. Thompson, J. Acoust. Soc. Am. **95**, 2853 (1994); P. O. Thompson *et al.*, Marine Mammal Sci. **12**, 288–292 (1987)]. Based on these descriptions, long-term seasonal and geographic studies of blue whale calls have been used to examine the Pacific basin-wide occurrence of calling blue whales. Lastly, Thompson and colleagues first remarked that blue whale sounds show distinct geographic variation, an observation that has promoted acoustics as a tool for distinguishing among populations of blue whales worldwide.

9:15

4aAB4. A comparison of optimized methods for the detection of blue whale sounds. David K. Mellinger, Sara Heimlich, and Sharon Nieukirk (Cooperative Inst. for Marine Resources Studies, Oregon State Univ., 2030 SE Marine Sci. Dr., Newport, OR 97365, david.mellinger@oregonstate.edu)

Paul Thompson and colleagues published one of the first long-term studies of mysticete sounds [Thompson and Freidl, Cetology **45**, 1–19 (1982)]. Thompson analyzed sounds manually, finding and tallying vocalizations to arrive at a view of seasonal occurrence. Today the detection and counting tasks are often done by computer, using various methods for pattern recognition. Here we examine and compare three such methods for detecting the sounds blue whales: matched filtering, which may work well because of the

stereotypy of blue whale vocalizations; spectrogram correlation, which may work well for the same reason and also because of the noise removal that can be done with it; and a neural network, which has worked well in other contexts for detecting right whale calls. The methods are configured using optimization procedures specialized for each method, and the results are compared for vocalizations recorded at different signal-to-noise ratios. The optimized detectors are applied to SOSUS data to detect sounds characteristic of blues whales in the northeast Pacific.

9:40

4aAB5. Blue and fin whales in the Southern California Bight: Using long-term acoustic records and acoustic recording tags. Erin M. Oleson (Scripps Inst. Oceanogr., Univ. Calif. San Diego, La Jolla, CA 92093), John Calambokidis (Cascadia Res. Collective, Olympia, WA 98501), Sean M. Wiggins, and John A. Hildebrand (Univ. Calif. San Diego, La Jolla, CA 92093)

Blue and fin whales have been the subject of acoustic investigation for decades, but recent advances in technology, including acoustic tags, autonomous recorders, computer automated signal detectors, and genetic sex-identification allow us to examine fine scale patterns in acoustic behavior. Northeast Pacific blue whales are known to produce at least five separate call types, and North Pacific fin whales at least three. Since August 2000, we have been acoustically monitoring blue and fin whales in the Southern California Bight using acoustic recording packages (ARPs), acoustic recording tags, and through genetic sampling of vocal whales. There are patterns of occurrence of particular call types with nonacoustic behaviors, and sex-biases exist in the production of certain calls. Acoustic sampling from several sites in the region indicates there are significant changes in the detection of various call types both within and between years, and among sites. Hourly detection patterns also vary, potentially serving as an indicator of feeding activity. An understanding of the temporal and spatial patterns of occurrence of individual call types, and the behavior associated with each, will lead to more efficient methods to monitor habitat, distribution, and abundance of these species.

10:05

4aAB6. Remembrances of Paul O. Thompson at the Naval Ocean Systems Center. Sue E. Moore (Natl. Marine Mammal Lab, 7600 Sand Point Way NE, Seattle, WA 98115) and Sam Ridgway (Naval Command, Control, and Ocean Surveillance Ctr., RDT&E Div., San Diego, CA 92152)

Paul Thompson spent 35 years working at the Naval Electronics Laboratory (later the Naval Ocean Systems Center) in San Diego and Hawaii. His research included psychoacoustics and animal bioacoustics. He authored 66 papers and reports, gaining the distinction of Fellow of the ASA. Paul was noted for painstaking data analysis, excellent aural memory, and outstanding skills at sound recording and processing. We present remembrances of Paul as a scientist, colleague, and friend during his years studying marine mammals for the Navy.

10:30–10:40 Break

Contributed Papers

10:40

4aAB7. Estimates of the percentage of sperm whales missed on combined visual and acoustic surveys in the eastern Pacific Ocean. Jay Barlow and Shannon Rankin (Southwest Fisheries Sci. Ctr., Natl. Marine Fisheries Service, NOAA, 8604 La Jolla Shores Dr., La Jolla, CA 92037, jay.barlow@noaa.gov)

Abundance estimation from line-transect surveys commonly assumes that all animals on the trackline are detected. For long-diving species, such as sperm whales, this assumption is commonly violated. Marine mammal surveys were conducted in the eastern tropical Pacific in 2000–2002, with visual observers searching by 25× binoculars and acousticians aurally monitoring two elements of a towed hydrophone array. Visual and acoustic teams were independent and did not learn of detections made by the other team until after the animals had passed abeam. Data from 186 visual and acoustic detections were used to estimate the fraction of whales missed by each method. Many of the 178 acoustic detections were at greater distances than could be seen by the visual observers. Based on Palkas' direct-duplicate method, the percentages of missed trackline groups were 38% and 21%, respectively, for visual and acoustic methods. Assuming these two detection methods are independent, the percentage missed by both groups is 8%. Acoustic detections greatly reduce the fraction of missed animals, but by itself this method cannot reliably estimate group sizes and still misses some groups. A combination of acoustic and visual methods is recommended for sperm whale surveys.

10:55

4aAB8. Effectiveness of acoustic localization in documenting spatial and temporal patterns in autumn migration of bowhead whales in the Alaskan Beaufort sea. Susanna B. Blackwell, Robert G. Norman, Charles R. Greene, Jr., Miles W. McLennan (Greeneridge Sci., Inc., 1411 Firestone Rd., Goleta, CA 93117, susanna@greeneridge.com), Trent L. McDonald (WEST Inc., Cheyenne, WY 82001), and W. John Richardson (LGL Ltd., King City, ON L7B 1A6, Canada)

During September 2001, 2002 and 2003, westbound bowhead whales were localized acoustically using directional autonomous seafloor acoustic recorders (DASARs). Each year, ~10 500–45 600 calls were detected over 23–36 days by an array of 11 DASARs deployed 6.5–22 km northeast of the oil-production island Northstar. Locations of origin for the 29% (2001) and 75% (2002) of the calls detected by two or more DASARs were determined by triangulation. Peak call detection rates occurred near 20 September in 2002 and 2003, but early in the month in 2001. In 2003, peak call detection rates exceeded 550 calls/hour. Upsweeps, downsweeps and constant-frequency calls made up 66%–68% of the calls each year. The migration corridor was similar in 2001 and 2002, but closer to shore by ~10 km in 2003. Comparisons with yearly aerial censuses by the Minerals Management Service showed that the two methods both documented migration timing and the offshore distance of the migration corridor. Aerial surveys covered a larger area and continued after onset of freeze-up. The acoustic method revealed more temporal and spatial details because it operated continuously over long periods independent of weather and darkness, and provided far more detections of whales. [Work supported by BP.]

11:10

4aAB9. Acoustic localization of bowhead whales near a Beaufort Sea oil development, 2001–2003: Evidence of deflection at high-noise times? W. John Richardson (LGL Ltd., Environ. Res. Assoc., P.O. Box 280, King City, ON L7B 1A6, Canada), Trent L. McDonald (WEST Inc., Cheyenne, WY 82001), Charles R. Greene, Jr., Susanna B. Blackwell (Greeneridge Sci. Inc., Goleta, CA 93117), and Bill Streever (BP Exploration (Alaska) Inc., Anchorage, AK 99519)

Most bowhead whales migrating west along Alaska's north coast during autumn pass 10–75 km seaward of BPs Northstar oil production island. Our objective was to determine the extent to which underwater sound from Northstar deflected the closest whales farther offshore. Whale calls were localized, via 11 DIFAR-based seafloor recorders, on 23–36 days per year while underwater sound was measured continuously 450 m from Northstar. Totals of 1121, 2057 and 14 304 calls were localized within a 20×26 km analysis area seaward of Northstar. Quantile regression was used to relate the 5th quantile distance-from-shore to underwater sound. A permutation method allowed for lack of independence of closely spaced calls. In 2001, the closest calling bowheads tended to be slightly farther offshore at times with high industrial sound, which were mainly due to vessels near Northstar. The displacement effect was maximal if sound was averaged over 50–80 minutes preceding calls. In 2002, high sound levels were rare and results were inconclusive. In 2003, offshore displacement was not significant despite high sample size and a migration corridor unusually close to shore. The upper 90% confidence bound on the percentage of bowheads displaced 2+ km did not exceed 1.6% in any year. [Work supported by BP.]

11:25

4aAB10. Acoustic surveys of killer whales during winter along the Pacific Northwest Coast. Thomas Norris (Sci. Application Intl., San Diego, CA), Brad Hanson, Dawn Noren, and Linda Jones (Northwest Fisheries Sci. Ctr., Seattle, WA 98112-2097)

An acoustic and visual survey was conducted along the outer coast of the Pacific Northwest (Washington to Central California) between 26 February and 14 March 2004 to determine the winter-time distribution and behavior of the southern resident ecotype population of killer whales (SRKWs). SRKWs are vocally active with a vocal repertoire that is well documented. A towed hydrophone array was deployed on 17 of the 20-day cruise resulting in an average of approximately 19 h of acoustic effort per day. Eleven of the 49 unique detections of marine mammal made were attributed to killer whales (22.5% of the total). Three of these killer whale detections occurred at night. Two of the three known SRKW pods (J, K, and L) were detected acoustically and positively identified as the J and L pods. Interestingly, the remaining nine unique acoustic encounters with killer whales were of either the offshore form or the transient form, considered to be vocally less active relative to the resident form. The hydrophone array substantially increased the number of detections of killer whales that would have been made using visual methods alone. Determining estimates of group size using acoustics remains problematic. Planned improvements to the acoustic system will be discussed. [Work funded by NOAA/NWFSC-Seattle, WA.]

11:40

4aAB11. Range-depth tracking of humpback whales using autonomous acoustic recorders. Aaron M. Thode, Peter Gerstoft, Melania Guerra, Dale M. Stokes (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093), William C. Burgess (Greeneridge Sci., Inc., Goleta CA 93117), Michael J. Noad (Univ. of Queensland, St. Lucia, Queensland, Australia), and Douglas H. Cato (Defence Sci. and Technol. Organization, Sydney, NSW, Australia)

Matched-field processing (MFP) is a technique for tracking an acoustic source in range and depth by comparing the output of an ocean acoustic propagation model with measured acoustic data collected across multiple hydrophones [e.g., Thode *et al.*, J. Acoust. Soc. Am. **107**, 1286–1300 (2000)]. In October 2003 a MFP experiment was conducted using humpback whale sounds recorded during the spring migration off the Sunshine Coast in Queensland, in conjunction with a larger experiment conducted by the Humpback Acoustic Research Collaboration (HARC). Humpback whale sounds between 50 Hz to 1 kHz were recorded on a five-element vertical array deployed in 24-m-deep water. The vertical array consisted of a set of flash-memory autonomous recorders attached to rope with an anchor at one end, and a subsurface float at the other. Acoustic data were simultaneously collected and monitored on five sonobuoys deployed over approximately 2-km range. Using the range estimates from the sonobuoys as bounds on the matched-field processing, a geoacoustic inversion using the calls was performed on the vertical array data using a genetic algorithm. Preliminary results of the inversion and resultant 3-D position fixes are presented. [Work supported by ONR.]

11:55

4aAB12. Model-based tracking of marine mammals. Ahmad T. Abawi, Paul Hursky, Michael B. Porter (Ctr. for Ocean Res., Sci. Applications Intl. Corp., San Diego, CA), John Hildebrand, Sean Wiggins, and Allan Sauter (Scripps Inst. of Oceanogr., La Jolla, CA)

Data from the August 2003 experiment conducted by the Scripps Institution of Oceanography in the Southern California Offshore Range (SCORE) are used to track marine mammals. SCORE is a naval training area near the island of San Clemente located in relatively shallow water. The water depth where the experiment was conducted is around 360 meters. Data were recorded on a 100-m, eight-element vertical line array (VLA) deployed from the floating instrument platform (FLIP) and four bottom-mounted hydrophones deployed in an area covering approximately 3 square kilometers. During the course of the 7-day experiment continuous recording of the ocean environment was made. The recordings contain calls from various marine mammals, particularly blue and fin whales. Data recorded on the bottom-mounted hydrophones and the VLA are used to track singing marine mammals using two entirely different model-based tracking techniques: The animals are tracked by comparing the predicted (using a propagation model) and measured time difference of arrival recorded in pairs of bottom-mounted hydrophones. Additionally, matched-field processing on the VLA data is used to estimate the range and depth of the animals.

Session 4aEA

Engineering Acoustics: Transduction, Radiation, Ultrasonics, General Engineering Acoustics

Kim C. Benjamin, Chair
Naval Undersea Warfare Center, Newport, Rhode Island 02840

Chair's Introduction—8:25

Contributed Papers

8:30

4aEA1. Single-crystal ferroelectrics for sonar devices. E. A. McLaughlin and H. C. Robinson (NUWC-UTMR, Code 15232, 1176 Howell St., Newport, RI 02841)

Single-crystal ferroelectrics or piezocrystals were recently introduced into the electroactive materials community. The 33-mode electromechanical coupling factor of piezocrystals is typically greater than 0.90, which is significantly larger than typical values for piezoelectric ceramics (0.62–0.74). For sonar projector applications this large, k33 has been responsible for more than doubling the bandwidth of active sonar arrays over what is currently achievable with ceramics. More recently, a crystal grower produced a cut of lead magnesium niobate-lead titanate (PMN-PT) single crystal with piezoelectric shear coefficient values of 7000 pm/V and shear coupling factors of 0.97. (For PZT5H, d15 is 730 pm/V.) This piezocrystal d15 coefficient implies significantly improved sensitivity and signal-to-noise ratio for accelerometers and hydrophones, while the high coupling promises bandwidth increases greater than those realized in 33-mode projectors using piezocrystals. The 31-mode is also being investigated. By measuring the response of the materials to high- and low-level electrical bias and excitation fields, mechanical prestresses, over temperature, the material's effective material properties as a function of these operational variables have been determined. [This work sponsored by ONR and NUWC ILIR.]

8:45

4aEA2. Acoustic radiation from arrays on coated prolate spheroids. Jeffrey E. Boisvert (NAVSEA Newport, Newport, RI 02841, boisvertje@npt.nuwc.navy.mil) and A. L. Van Buren (Middletown, RI 02842)

An important aspect in the design of conformal sonar arrays is the effect of an outer coating on the array response. The design of the coating is subject to the competing requirements of acoustic transparency, toughness for impacts, and flow noise rejection. The present study considers an array of rectangular pistons, conformal to a rigid prolate spheroidal baffle, covered by an outer coating. The coating is modeled as a fluid, i.e., it is assumed to have a vanishingly small value of the shear modulus, typical of many elastomers. Unlike a coating on a spherical or cylindrical baffle, where a uniform thickness coating can be defined by simply increasing the value of the radial coordinate, the corresponding operation in prolate spheroidal coordinates results in a confocal layer, i.e., a layer that is not of constant thickness. In this paper a new method is described that allows for a uniform thickness coating in prolate spheroidal geometry. This method employs the spheroidal addition theorem to satisfy the boundary conditions. Comparisons of the beamformed array response for a confocal coating versus a uniform thickness coating are given for a sample array with various coating properties. [Work supported by the NUWC ILIR Program.]

9:00

4aEA3. A fast time-domain method for calculating the near field pressure generated by a pulsed circular piston. James Kelly (Dept. of Math., Michigan State Univ., East Lansing, MI 48824) and Robert McGough (Michigan State Univ., East Lansing, MI 48824)

A fast single-integral method has been derived for calculations of transient near field pressures generated by circular pistons. This new time-domain method outperforms other methods with respect to speed, accuracy, and ease of implementation. Unlike methods that apply the spatial impulse response, the new method inherently eliminates aliasing problems, which permits a significant reduction in the temporal sampling rate. As a result, the new method is approximately six times faster than the spatial impulse response for a 1% maximum numerical error. This new method converges exponentially with Gauss quadrature, whereas spatial impulse response methods evaluated with fast Fourier transforms only converge linearly with respect to sampling frequency. Thus, the new method achieves small numerical errors in very short times. Since the new fast method evaluates a single-integral expression that is readily computed with standard quadrature techniques, the implementation is greatly simplified relative to the spatial impulse response. The new method, which achieves smaller errors in less time with lower sampling rates, is also much more memory-efficient than the impulse response approach. Each of these advantages represents a significant improvement with respect to the impulse response method for computations of transient acoustic fields.

9:15

4aEA4. New ultrasonic ranging transducer using 200-KHz corrugated polyvinylidine fluoride film. Minoru Toda and Jerry Dahl (Measurement Specialties Inc., 460 E. Swedesford Rd., Ste. 3005, Wayne, PA 19087)

Corrugated PVDF transducers were originally developed for use in parametric arrays to demodulate audio sound in air from modulated high intensity ultrasound. These arrays were designed to operate at $\sim$40 KHz. PVDF has recently been used in low-cost short-range-distance ranging transducers and provides accurate distance measurements in consumer, industrial, and medical applications. For this purpose, we designed, fabricated, and tested a corrugated PVDF transducer that operates at 200 KHz. This results in a small, directional, and low-cost transducer. The transducer has an area of ~ 1 cm^2 and produces a beam width of 16°. Combined with a sensor circuit that converts the measured time of flight into distance, we have demonstrated a detection range of 17–500 mm with a resolution of 25 mm. The resonance frequency of curved PVDF film was increased to 200 KHz by reducing the curvature radius to $\frac{1}{5}(\sim 1$ mm) of that for previous devices operating at 40 KHz. The concave and convex regions of the corrugation vibrate with opposite phase and the radiated ultrasonic waves from both regions have a half-wavelength path difference that cancels the opposite vibration phase due to the half-wavelength effect.

9:30

4aEA5. Investigation of vibration of polyvinlidine fluoride cylindrical film by use of acoustic probe. Minoru Toda (Measurement Specialties Inc., 460 E. Swedesford Rd., Ste. 3005, Wayne, PA 19087)

Ultrasonic pen and white board digitizing systems are a rapidly growing market for PVDF transducers, utilizing an ultrasonic transmitter in the pen tip and two stationary ultrasonic receivers. The transmitter is constructed of piezoelectric PVDF welded into a cylindrical shape. Radius vibration (breathing mode) creates an 80-kHz airborne acoustic wave. The initial cycle is used to determine the time of flight, used to calculate the pen position by triangulation. An ideal cylindrical transmitter would have a uniform 360 horizontal radiation pattern. However, practical devices have a 65% to 85% minimum/maximum ratio for the initial pressure peak. A novel acoustic probe was devised to investigate this nonuniformity. The probe tip is a needle with 0.5-mm through hole, and the acoustic wave propagates through the needle to a conventional reference microphone. The probe was positioned using a precision XYZ stage. Investigations with this probe revealed the details of the nonuniformity in horizontal radiated pressure. It was found that (1) vibration at the welded seam of the PVDF cylinder is a maximum, opposite the expected behavior, and (2) the pressure distribution profiled along the height of the cylinder has a central plateau and decreases linearly at the cylinder ends.

9:45–10:00 Break

10:00

4aEA6. Lamb waves in bulk acoustic wave resonators: Analysis of spurious resonances and design of resonators in the UHF-VHF frequency range. Alexandre Volatier, Gregory Carruyer (ST Microelectronics, 38926 Crolles, France, volatieral@chartreuse.cea.fr), Emmanuel Defay (LETI-CEA, Grenoble, France), and Bertrand Dubus (UMR CNRS 8520, Lille, France)

Bulk acoustic wave (BAW) resonators exhibit attractive properties in terms of power handling capacity and on-chip integration to realize filters in the GHz range [K. M. Lakin, IEEE Ultrason. Symp. pp. 895–905 (1999)]. In a BAW resonator, a thin piezoelectric layer (a few mm) deposited between two electrodes is driven in its thickness extensional mode of vibration. To get a high quality factor, this structure is decoupled from the substrate by a multilayer Bragg reflector or a back-etched membrane. A current problem in the design of BAW resonators is the existence of spurious resonances close to the thickness extensional mode which generate ripple in the filter passband. In this paper, these spurious modes are analyzed in terms of Lamb waves resonances. Physical modeling of BAW resonators using finite element ATILA code is presented. The influence of lateral dimensions and electrode geometry on spurious resonances is emphasized. With specific electrode design and electrical excitation, it is demonstrated that lateral modes of Lamb waves can be used to realize resonators in the 50–250-MHz range. Experiments on an AlN piezoelectric layer between Pt electrodes on a SiN membrane are presented. [Work supported by a ST Microelectronics grant (CIFRE).]

10:15

4aEA7. Ultrasonic nondestructive inspection of interface defects in anisotropic fiber-metal-laminates. Stefan A. L. Stijlen (Dept. of Imaging Sci. and Technol., Delft Univ. of Technol., Lorentzweg 1, 2628 CJ Delft, The Netherlands, s.a.l.stijlen@tnw.tudelft.nl) and Maarten C. M. Bakker (Delft Univ. of Technol., 2629 HS Delft, The Netherlands)

Ultrasonic inspection is an essential tool for the quality control of the fiber-metal-laminate Glare, which is currently applied in the production of the huge A380 Airbus aeroplane. Nowadays, the ultrasonic inspection of Glare is performed by a transmission C-scan, traditionally, or by a more advanced pulse-echo scan. In this research, a new method of data acquisition and processing is investigated. The data acquisition consists of reflection measurements for variable source-receiver offsets. The first objective is to image delaminations at or nearby interfaces. These defects can, for instance, be associated with kissing bonds due to locally imperfect coupled interfaces. The algorithm which processes and converts the data into an image is based upon the inverse scattering theory. From the acquisition level in water to the interface levels in Glare, the elastodynamic Green functions are computed including strong anisotropy, multiple reflections and conversions, refractions, and surface waves. Subsequently, interface defects are modeled as additional contrasts in compliance. Finally, an image is obtained through their optimization by minimizing a cost-functional which is the mismatch between the measured data and the synthetic data generated with the Green functions. [Work supported by STW, the Dutch technology foundation.]

10:30

4aEA8. Experimental equalization of a one-dimensional sound field using energy density and a parametric equalizer. Micah Shepherd, Xi Chen, Timothy Leishman, and Scott Sommerfeldt (Acoust. Res. Group, Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602, mrs74@email.byu.edu)

A simple experimental method has been developed to equalize a one-dimensional sound field using acoustic energy density. Energy density is estimated using two methods: the two-microphone transfer-function method and the two-microphone finite-difference method. An equalization filter is manually configured for each using a digital parametric equalizer. The results of the two estimation approaches are compared. Equalization filters are similarly implemented for sound pressure measured at discrete points in the field and for spatially averaged sound pressure. Results of the approaches are compared and benefits of the energy density method are discussed.

10:45

4aEA9. Sound-quality analysis of sewing machines. James J. Chatterley, Andrew J. Boone, Jonathan D. Blotter (Mech. Eng. Dept., School of Eng. and Technol., Brigham Young Univ., 435 CTB, Provo, UT 84602), Scott D. Sommerfeldt, and Thomas L. Lago (Brigham Young Univ., Provo, UT 84602)

Sound quality has become an important factor in consumer product development. Sound quality analysis procedure and results for six sewing machines ranging from entry level to professional grade machines will be presented. The focus of the study was to determine what consumers feel constitutes a pleasant sounding sewing machine and what structural modifications can be made to produce those sounds. The procedure consisted of forming small (15 person) then large (50 person) jury-based tests. These tests consisted of listening to various sound bytes and ranking the sounds as well as indicating why sounds are appealing or unappealing. The sound bytes were from actual sewing machines as well as computer generated sounds and modified machine sounds, which were constructed such that the spectral features most important in achieving a desirable product from an acoustic perspective could be statistically determined. The procedures and analysis of the jury testing results will be presented and discussed. This paper presents sound localization scans, indicating machine sound hot spots and possible sources for undesired sounds. In conclusion proposed modifications to machine structure to alter machine sound signature into a more sensory pleasant sound will also be presented.

11:00

4aEA10. Production of a noise level database of power tools used in the construction industry. Charles S. HaydenII, Edward Zechmann, and Rohit Verma (Natl. Inst. For Occupational Safety and Health, 4676 Columbia Pky., C27, Cincinnati, OH 45226, chayden@cdc.gov)

This study focused its efforts on sound power levels and noise reduction of power tools used by workers at both commercial and residential construction sites. The objective was to develop a noise control technology database consisting of power hand tools used in the construction industry with respective sound power levels and workers' sound exposure level (SEL). The database and specific noise control applications information are made available to the public on a searchable web-site. Data includes a particular tool's wavfile and time series signatures to be used for future

qualitative and quantitative analysis by other acousticians accessing the database. With this information, reasonable assumptions can be made as to the potential occupational noise exposure expected for various occupations, tasks, etc. and as to potential sources of noise within the powered hand tool itself. The database captures tool types, tool models, tool manufacturers, and decibel power level. The database is expandable to include new tools and new manufacturers as the power tool population changes over time. These efforts support the reduction of noise induced hearing loss among construction workers by providing ''buy quiet'' and ''designed quiet'' information to power tool buyers and end users.

THURSDAY MORNING, 18 NOVEMBER 2004 PACIFIC SALON 3, 8:30 TO 11:45 A.M.

Session 4aMU

Musical Acoustics: Physics of Musical Instruments

George A. Bissinger, Cochair
Department of Physics, East Carolina University, Greenville, North Carolina 27858

Daniel O. Ludwigsen, Cochair
Science and Mathematics Departments, Kettering University, 1700 West Third Avenue, Flint, Michigan 48504-4898

Contributed Papers

8:30

4aMU1. Galilei experiments and common sense. R. Dean Ayers (Dept. of Phys. and Eng. [Adjunct], Southern Oregon Univ., 1250 Siskiyou Blvd., Ashland, OR 97520)

Galileo Galilei showed us by example how to put hypotheses about the natural world to the test of scientific experiments. It is very likely that his attitude of healthy skepticism toward his own ideas and those of others was instilled in him by some work in musical acoustics. His father, Vincenzo (or Vincentio), was a prominent music theorist and composer as well as an accomplished lutenist. During his adolescence Galileo assisted Vincenzo in practical studies on the behavior of vibrating strings. That research examined the influence of various parameters on the pitch of a string (and hence its frequency), disproving some incorrect guesses. We are now in an age when many of our students are unduly influenced by popularized reports on exotic and seemingly bizarre open questions in relativity, cosmology, and elementary particles. That exposure encourages them to abandon their common sense and skepticism when they take a science course. Our specialty can provide a valuable counterexample by letting students perform exploratory, quantitative studies on the strings of the original Galilei experiments or on other vibratory systems. We can all benefit by remembering the lessons learned from such work. [Work supported in part by the Paul S. Veneklasen Research Foundation.]

8:45

4aMU2. End correction for an open pipe from measured resonance frequencies. Daniel O. Ludwigsen and Brandon J. Dilworth (Dept. of Sci. and Math., Kettering Univ., 1700 W. Third Ave., Flint, MI 48504)

As briefly noted in freshman physics textbooks, resonance in open-ended pipes depends on an effective length that is greater than the actual length of the tube. The additional end correction varies with the termination at the open end. By measuring several peak frequencies in a response function (pressure in the pipe over source excitation), we have calculated the end correction for open ends of PVC tubes of varying diameters with several flanges of increasing size. Temperature effects were carefully controlled, and resonance frequencies recorded to within 0.01 Hz. Averaged results are in agreement with literature and theoretical expectations, but anomalies and unexpected results will also be presented.

9:00

4aMU3. Period doubling in free reeds coupled to pipe resonators. Evan M. Goetzman (Univ. of Minnesota, Minneapolis, MN 55455) and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

An earlier study investigated behavior of the reed–pipe combination consisting of an American organ reed installed at the closed end of a cylindrical pipe. Over a wide range of pipe lengths and playing pressures, the sounding frequency is slightly below a pipe resonance frequency and can be pulled considerably below the natural frequency of the reed. [Vines *et al.*, J. Acoust. Soc. Am. **114**, 2349 (2003)]. Additional measurements of reed vibration for these reed–pipe combinations for low frequency (48 Hz) free reeds show that as the playing pressure is increased there is typically a sudden transition to period doubling or, in some cases, tripling or quadrupling. Some pressure intervals exist in which apparently chaotic reed vibration occurs, or in which the reed will not vibrate at all. Period doubling was found to occur over varied values of pipe length, pipe diameter, and the use of two different reeds with matching resonance frequencies. General trends in the onset and steady-states of period doubling are described using measured spectra and spectrograms of both reed motion and radiated sound pressure over a continuous range of playing pressure. [Work supported by NSF REU Grant No. 0354058.]

9:15

4aMU4. The motion of harp strings. Chris Waltham (Dept. of Phys. and Astron., Univ. of Britich Columbia, Vancouver, BC V6T 1Z1, Canada, waltham@physics.ubc.ca)

A harp is an instrument with a set of plucked strings that excite the sound board directly, without the medium of a bridge. The strings are positioned at an acute angle to the plane of the sound board. The quality of the sound produced depends on the motion of the string and its interaction with the resonances of the sound board. The string motions of several small and large harps have been studied using small, fast position sensors. The results are compared to those of a simple nonlinear model based on the measured elastic properties of the string materials, and those of the sound board. The implications for the sound production are discussed.

9:30

4aMU5. Measurements of the acoustics of the kalimba. Daniel O. Ludwigsen (Dept. of Sci. and Math., Kettering Univ., 1700 W. Third Ave., Flint, MI 48504)

The kalimba is an idiophone (self-sounding instrument) of African origin, specifically a standardized type of mbira manufactured with modern facilities by Hugh Tracey. The sound production is somewhat similar to a small glockenspiel, but the free ends of the metal keys are plucked rather than simply-supported struck bars. The body of the kalimba, however, more resembles a guitar with a central sound hole in the top plate. There are additionally two smaller holes in the back plate, which a player can use to create vibrato effects with rapid covering and uncovering with fingertips. Spectral information and modal analysis are used to investigate the dynamics of sound production of the kalimba, and are compared with such measurements in the literature for guitar and mallet instruments. These measurements will inform future modeling efforts to better understand the dynamics of the instrument.

9:45

4aMU6. Inexpensive apparatus for studies of the lip reed. Mark T. McLaughlin, Lowell J. Eliason (Dept. of Phys. and Astron., California State Univ., Long Beach, 1250 Bellflower Blvd., Long Beach, CA 90840-3901), and R. Dean Ayers (Southern Oregon Univ., Ashland, OR 97520)

A circuit has been designed and built to provide real-time, precisely controlled video images of a brass player's lips during the relatively steady part of a sustained note. This should be useful for both the scientist who wishes to model the lip reed and the beginning player who is just learning how to control that driver. The circuit uses a phase-locked loop integrated circuit to track the signal from a microphone in the mouthpiece and subdivide the period for triggering a strobe light. The signal from the video camera is used for arming the trigger, so that there will be at most one strobe flash per field. This circuit is stable enough in tracking the playing frequency that a trombone slide can be moved continuously through its full range of positions, yet the live, strobed image advances through equally spaced fractions of a cycle at a rate determined by the settings of the circuit. This is important because the playing frequency is not perfectly constant, especially for beginners. [Work supported in part by the Paul S. Veneklasen Research Foundation and the CSULB Scholarly and Creative Activities Committee.]

10:00

4aMU7. Drumstick "Pitching"—A modal analysis approach. George Bissinger (Phys. Dept., East Carolina Univ., Greenville, NC 27858)

"Pitching" drumsticks by ear is too time intensive for a manufacturing process that generates many millions of drumsticks each year. Yet, it is a sought-after attribute that is worth implementing if some pitching technique can be made compatible with typical manufacturing rates. Drumstick pitch is normally determined by our evaluation of the sound produced by the nonharmonic vibrations when it is held and struck against a hard surface. Vibration analysis performed on drumsticks in a free–free support fixture provided normal mode results for over 100 solid wood and wood+plastic tip drumsticks. Modal analysis measurements up to 5 kHz were made parallel and perpendicular to the plane of the grain at 17 points along the length of a few drumsticks. A small sample of four drumsticks, pitch-matched by ear for comparison with these vibration analysis predictions, indicated a promising solution to the problem of real-time pitch matching. Even though the usual problems were encountered—orthotropic, small knots, kinked or twisted grain, density variations (dimensional consistency was within 0.1% but mass varied over 40%), and stiffness variations—the least sensitive to these problems was the lowest bending mode near 0.4 kHz, now the basis for automated pitch matching by a major manufacturer.

10:15

4aMU8. The glissando of colliding hard objects. Benjamin Frey and William M. Hartmann (Dept. of Phys. and Astron., Michigan State Univ., East Lansing, MI 48824)

When two hard objects are pressed together they alternately collide and rebound, leading to an audible, decaying sequence of pulses with a decreasing interpulse interval. Therefore, the sequence is heard as a rising pitch or glissando. Virtuoso performances of colliding hard objects (steel spheres) demonstrated a frequency rise that was accurately an exponential function of time over the span of more than three octaves. The amplitude decay also tended to be exponential, though less reproducibly so. Because successive decaying pulses were otherwise virtually identical functions of time, the spectra were harmonic. Four harmonics were clearly evident in spectrograms. Spectrograms also revealed a strong formant near 2000 Hz, presumably indicating the lowest mode of vibration of the steel spheres themselves. Formants had a Q of about 7. With the quantitative acoustical description well in hand, interest now centers on discovering the time-dependent force function employed by experienced performers of colliding hard objects when they achieve the most effective glissandos. [Work supported by the Michigan State University Professorial Assistant Program.]

10:30

4aMU9. Multiphonic model of split pressure-phase modes using the finite-element-method. Bader Rolf (Musikwissenschaftliches Institut, Neue Rabenstr. 13, 20354 Hamburg, Germany, r_bader@t-online.de)

The generation of multiphonic sounds in blown musical instruments is modeled by an eigenvalue finite-element-method (FEM) calculation. A common technique for the production of multiphonic tones is the use of complex sound-hole closing techniques, where more than one overtone spectrum appears, with fundamental frequencies being in inharmonic relations. Eigenfrequencies of a tube with both ends open are compared to eigenfrequencies of a tube with at least one sound hole open. It shows up that the tube with one open sound hole leads to additional eigenfrequencies in the tube for each node number. This is because of two possibilities of pressure curves within the tube for each node number, the pressure right and left of the hole being in phase or in opposite phase. The frequencies of equal phase form a new overtone structure which is harmonic within itself and so forms a new tone, a multiphonic. So the multiphonic model is in agreement with the Backus formula of multiphonic sounds as a combination of the tubes fundamental and multiples of the frequency difference of the zero node mode components.

10:45

4aMU10. Modeling the nonlinear string body coupled dynamics of bowed musical instruments. Octávio Inácio (Instituto Politécnico do Porto, Escola Superior de Música e das Artes do Espectáculo, Musical Acoust. Lab., R. da Alegria, 503, 4000-045 Porto, Portugal) and José Antunes (Instituto Tecnológico e Nuclear, 2686 Sacavém codex, Portugal)

Most theoretical papers on bowed-string instruments deal with isolated strings, pinned on fixed supports. In addition, the instrument body dynamics has not been accounted at all, or else by using extremely simplified models of the string/body interaction at the bridge. Such models have, nevertheless, been instrumental to the understanding of a very common and musically undesirable phenomenon—the "wolf note"—a strong beating interplay between string and body vibrations. Cellos, bad and good, are prone to this problem. In previous work we developed a modal method to deal with friction-excited strings, enabling effective simulations of such systems in which the strings were assumed decoupled from the instrument body. In the present paper our computational method is extended to incorporate the complex dynamics of real-life instrument bodies, coupled to the string motions. In this approach, the string is coupled with experimental body data—body modes or impulse response at the bridge. Our computa-

tional method is illustrated through extensive parametric computations performed on a bowed cello. These numerical simulations show some light on interesting and less-known features of wolf notes, in particular concerning the ranges of bowing-parameters leading to their emergence, as well as the dependence of the beating frequency on the playing.

11:00

4aMU11. Turbulence eddy viscosity in flutelike instruments. Rolf Bader (Musikwissenschaftliches Institut, Neue Rabenstr. 13, 20354 Hamburg, Germany, r_bader@t-online.de)

The generator area of a flutelike instrument is modeled by the finite-element-method (FEM) using a Navier–Stokes (NS) model and a k-epsilon turbulence model. A steady-state simulation with condition of zero pressure at the tube's end and a transient simulation using a back-impulse from the tube's end to the generator area was performed. Both models NS and k-epsilon agreed in the overall structure of the volume flow from the player's lips split in a part flowing into the tube and one flowing outside, common in eddy and vortex structure. But the k-epsilon model showed a much more realistic description of the relation of flows into and outside the tube. As the NS model results in a one-to-one split of the flow, the turbulence model leads to a much smaller flow part into the tube with most of the flow going outside and so to a much higher vorticity at the flute lip. The transient model also shows a sharpening of the transient pressure curve because of this sharpened lip near vorticity behavior. So as the turbulence model seems to be a good choice in formulating a discrete flute model, the overall eddy damping seems to play an important role in flute dynamics.

11:15

4aMU12. Separation of musical instrument timbre using performance information and real-time harmonic/inharmonic component filtering. Joseph A. Sarlo (Univ. of California, San Diego, 4991 Combe Way, San Diego, CA 92122, jsarlo@ucsd.edu)

Described here is a technique for the real-time separation of a musical instrument timbre from a multitimbre sound source. The separated timbre may be a timbral component of a single musical instrument timbre (e.g., the breath noise from a wind instrument) or an instrument timbre from a multi-instrument sound source. The separation is accomplished by first using performance information, such as pitch tracking or score alignment, to estimate a fundamental frequency for the source of the timbre to be separated. Real-time analysis/resynthesis filtering of the sinusoidal components based on their harmonic relationship to the estimated fundamental frequency is then employed to extract the desired timbre.

11:30

4aMU13. Synthesis of audio spectra using a diffraction model. V. Vijayakumar (FOSEE, Multimedi Univ., Jalan Air Keroh Lama, Melaka 75450, Malaysia, vijaya@mmu.edu.my) and C. Eswaran (Multimedi Univ., Cyberjaya, Selangor, Malaysia)

It is shown in this paper that the intensity variations of an audio signal in the frequency domain can be explained by using a Fresnel diffraction model. By proper choice of two parameters an input spectrum of discrete frequencies of unit intensity can thus be transformed into the known spectra of different musical instruments. Specific examples of musical instruments are considered for evaluating the performance of this method. It is found that this model yields musical spectra with a good degree of accuracy. Synthesized sounds of bells and drums are presented as real examples.

THURSDAY MORNING, 18 NOVEMBER 2004 ROYAL PALM SALON 1, 9:00 TO 11:15 A.M.

Session 4aNS

Noise and Committee on Standards: Active Noise Reduction for Hearing Protection: Technology Opportunities and Assessment of Effectiveness

William J. Murphy, Chair

National Institute of Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, Ohio 45226-1998

Chair's Introduction—9:00

Invited Papers

9:05

4aNS1. Active noise reduction hearing protectors; 50 years of development. Kenneth A. Cunefare (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

This paper will review the technology development of active noise reduction (ANR) hearing protectors. Such devices have been in development for close to 50 years. Early studies by Meeker on feedback implementation of ANR identified many of the issues and performance criteria that are still relevant today. These issues include compensation, minimum phase filter design, and out-of-band control stability. Numerous patents and publications since the early developments focused on means to address the issues discussed by Meeker, and, to introduce alternative implementations of ANR hearing protectors or headsets. Such alternatives include feed-forward designs, as well as open-back and closed-back designs (open-back designs rely solely upon active control for noise protection; closed-back designs include passive attenuation from an ear muff). Development has also proceeded on designs which include communication elements; here, the active control system must attenuate the noise, but not the communication. Fifty years on, ANR hearing protectors represent the most successful commercial application of active noise control technology.

9:25

4aNS2. Design and performance of hybrid feedforward-feedback active noise reduction for hearing protection and communication. Laura Ray (Thayer School of Eng., Dartmouth College, Hanover, NH 03755)

Since the 1990s research and development of digital feedforward active noise reduction (ANR) based on least-mean-square (LMS) algorithms has shown that significant gains in noise reduction can be achieved for circumaural hearing protectors. However, to-date digital feedforward hearing protection devices have not penetrated the market. This presentation reviews prior research in feedforward ANR for circumaural devices, along with the opportunities and challenges for commercialization of the technology. The effect of noise source characteristics, specifically stationarity and dynamic range, on stability and performance of LMS filters is examined. Research on hybrid ANR architectures combining LMS filters and digital feedback ANR is reviewed, showing that robust performance with stability margins required for a commercial product can be achieved with such systems. The role of ANR in enhancing speech intelligibility is outlined, along with research needed to develop standards for measurement of stability, ANR performance, and speech intelligibility afforded by digital feedforward ANR devices. The paper concludes with a presentation of opportunities offered by feedforward ANR in intra-aural and supra-aural hearing protection devices.

9:45

4aNS3. Development and performance of an active noise reduction earplug for use in extreme noise environments. Richard L. McKinley, John Allan Hall (Human Effectiveness, Air Force Res. Lab., 2610 Seventh St., Wright-Patterson Air Force Base, OH 45433-7901), Will Saunders, and Mike Vaudrey (Adaptive Technologies Inc., Blacksburg, VA 24060)

High-performance military fighter aircraft at high power settings can generate near-field noise levels typically between 140 and 150 dB SPL at approximately 50 ft. Providing adequate noise attenuation for maintenance personnel working in these locations is a challenging task. This paper describes the basic design of an active noise reduction earplug that, when combined with a passive earmuff, will operate in noise environments up to 150 dB SPL. The design and fabrication was the result of a contractual collaborative effort of the Air Force Research Laboratory and Adaptive Technologies Inc. The basic design includes a custom-molded deep insert earplug with an integrated error microphone and earphone. The active noise reduction controller was realized using digital signal processing. The performance of the ANR earplug and earmuff system was measured in both laboratory and field conditions using both real-ear attenuation at threshold and miniature microphone in real-ear measurement techniques. The performance of the system will be described in terms of both active and passive attenuation in 142–150-dB SPL broadband noise fields.

10:05–10:20 Break

10:20

4aNS4. Hearing protector labeling for active noise reduction devices. William J. Murphy, John R. Franks (NIOSH, Hearing Loss Prevention Team, 4676 Columbia Pkwy., MS C-27 Cincinnati, OH 45226-1998, wjm4@cdc.gov, jrf3@cdc.gov), and Alberto Behar (Univ. Toronto, Toronto, ON, M5S3G9 Canada)

The US EPA regulation 40 CFR part 211b does not specify how to label hearing protectors that use active noise reduction (ANR). Real-ear attenuation at threshold (REAT) measurements are appropriate for testing the passive noise reduction performance of the earmuff. The contribution of ANR to the overall attenuation must be measured either with an acoustic test fixture (ATF) or using the microphone in real ear (MIRE) technique. ATFs must adequately mimic the coupling of the protector with either the skin of the ear canal or the side of the head, as well as provide sufficient attenuation to eliminate effects of bone conduction. The MIRE technique uses a miniature microphone positioned in the ear canal either at the entrance or at the tympanic membrane, and can be used to measure the insertion loss between the occluded and unoccluded conditions for both the passive and active modes of the protector. ATF and MIRE measurements made for several protectors demonstrate effective performance below 1000 Hz. This paper will examine methods for combining ATF and MIRE measurements with REAT to develop effective and informative rating for the ANR class of hearing protectors. [Portions of work supported by the U.S. EPA IA 75090527.]

10:40

4aNS5. Influence of control structure and sound reproduction on the intelligibility of speech in headsets with active noise reduction. Anthony J. Brammer (Univ. of Connecticut Health Ctr., 263 Farmington Ave., Farmington, CT, 06030-2017, anthonybrammer@hotmail.com), Donald R. Peterson, Martin G. Cherniack, and Subhash Gullapalli (Univ. of Connecticut Health Ctr., Farmington, CT 06030-2017)

A primary requirement for a communication headet is to maintain speech intelligibility under all conditions of use. For headsets equipped with active noise reduction (ANR), the performance of the control system may influence the communication signal reaching the ear. Conversely, the communication signal may perturb the operation of the ANR system. The interaction between the communication and control signals will depend primarily on the control structure, and the electro-acoustic transfer function of the earphone and drive electronics of the communication channel. The effect on speech intelligibility will be described using the speech transmission index, for circumaural headsets employing feedback control with a fixed filter and analog signal processing, an approach commonly used in commercial devices, and one employing adaptive, digital feed-forward control. In each case, the active control systems attempt to control low-frequency noise, as the passive attenuation of the earmuff enclosing the ear substantially reduces noise at frequencies above 500 Hz. Results will be discussed for different environmental noise spectra, and speech-to-noise ratios.

Contributed Paper

11:00

4aNS6. An objective method for measuring the attenuation of hearing protection devices using otoacoustic emissions. Darin A. Knaus, Anthony J. Dietz (Creare Inc., P.O. Box 71, Hanover, NH 03755, dak@creare.com), and Frank E. Musiek (Univ. of Connecticut, Storrs, CT 06269)

An objective human subject test technique for measuring the attenuation performance of hearing protection devices is presented. The technique is similar to the real-ear attenuation at threshold (REAT) technique, except distortion product otoacoustic emission (DPOAE) measurements in the ear canal replace the standard audiometric threshold measurements. The primary advantages of the new technique are that the measurement is objective, relatively fast, and can be applied at supra-threshold sound levels. Microphone in real ear (MIRE) data is also available from the DPOAE probe in the ear canal. Disadvantages are the restricted frequency range in which DPOAE signals may be reliably recorded (2–5 kHz) and the need for an in-ear probe. Attenuation data for a set of earmuffs was measured on a pool of seven subjects using DPOAE, MIRE, and threshold methods. The DPOAE and threshold measurements were in agreement, validating the DPOAE method. The MIRE data were significantly higher than the DPOAE or threshold data. The difference in the attenuation levels represents sound reaching the cochlea that is bypassing the ear canal. This difference was smaller for measurements of the attenuation performance of a helmet, which provides increased shielding from bypass transmission paths through the head.

THURSDAY MORNING, 18 NOVEMBER 2004 PACIFIC SALONS 4 & 5, 7:45 A.M. TO 12:00 NOON

Session 4aPA

Physical Acoustics: Applications of Acoustic Radiation Pressure

Gregory A. Kaduchak, Cochair
Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, New Mexico 87545

Thomas J. Matula, Cochair
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698

Chair's Introduction—7:45

Invited Papers

7:50

4aPA1. Manipulation of cells in ultrasound standing waves. W. Terence Coakley (Cardiff School of Biosciences, Cardiff Univ., Cardiff CF10 3TL, UK)

The acoustic radiation force operating on a cell in a 0.5-MPa pressure amplitude 3-MHz standing wave field equals that experienced in a 250-g gravitational field. Consequently, cells may be driven rapidly to preferred points in the sound field. Acoustic radiation pressure can drive bacteria, flowing by an acoustic reflector (prepared as an immunosensor surface), onto that surface where they are then captured. Animal cells have been driven onto the fabric of wide-gauge filters to form, a basis for a supported bioreactor. Real standing wave fields do not have the ideal plane-wave form, particularly when the transducer diameter is much larger than a wavelength (typical transducer diameter of 20 mm, wavelength of 0.5 mm at 3 MHz in water). Pressure variations within a node plane then give rise to regions towards which cells migrate and may be held against flow in potential traps. These traps are exploited commercially to retain cells in fermentors in the pharmaceutical industry. Most recently, ultrasound traps have been employed to concentrate animal cells in the focal plane of a fluorescence microscope so that cell interactions and their consequences at the molecular level can be monitored.

8:15

4aPA2. Nondestructive microscopic concentration, separation, analysis of biomaterials in microfluidic space using acoustic radiation force for cellomics study. Kenji Yasuda (Dept. of Life Sci., Grad. School of Arts and Sci. Univ. of Tokyo, 3-8-1 Komaba, Meguro, Tokyo 153-8902, Japan)

A method for noncontact handling of biological materials such as cells using acoustic radiation force could be incorporated in a fully automated microfluidic analysis system for cellomics studies. Benefits of acoustic radiation force are simple and compact design without moving parts, suitability for continuous concentration, separation and mixing in micrometer-sized spaces, and ready use in microchambers allowing contamination-free handling. Since this can be an advantage in biological applications for on-chip cellomics screening, such handling techniques using ultrasound are pursued. At the meeting, the following three applications are presented: (1) mixing, concentration and fractionation of biomaterials [K. Yasuda *et al.*, J. Acoust. Soc. Am. **99**, 1248–1251 (1996); **102**, 642–645 (1997), Jpn. J. Appl. Phys. **34**, 2715–2720 (1995); K. Yasuda, Sens. Actuat. B. **64**, 128–135 (2000)]; (2) superposition method for improving concentrating efficiency [K. Yasuda, Jpn. J. Appl. Phys. **36**, 3130–3135 (1997)]; (3) separation of different materials exploiting the competition between acoustic radiation force and electrostatic force [K. Yasuda *et al.*, J. Acoust. Soc. Am. **99**, 1965–1970 (1996); Jpn. J. Appl Phys. **35**, 3295–3299 (1996), K. Yasuda, Jpn. J. Appl. Phys. **38**, 3316–3319 (1999)].

8:40

4aPA3. Acoustic levitation in the field of analytical, physical, and atmospheric chemistry. Rudolf Tuckermann (GKSS-Res. Ctr., Inst. for Coastal Res., Max-Planck-Str.1, D-21502 Geesthacht, Germany, rudolf.tuckermann@gkss.de)

In this contribution an overview of fundamentals, possibilities, and applications of acoustic levitation in the field of analytical, physical, and atmospheric chemistry investigated by Neidhart and co-workers since the 1990s will be given. Started with sample preparation procedures in acoustically levitated drops, e.g., analytical enrichment, liquid–liquid extraction, and acid-base titration [E. Welter and B. Neidhart, Fresenius J. Anal. Chem. **357**, 345–350 (1997); R. Eberhard and B. Neidhart, *ibid.* **365**, 475–479 (1999); O. Rohling, C. Weitkamp, and B. Neidhart, *ibid.* **368**, 125–129 (2000)], the work has been continued in evaporation processes of liquids [R. Tuckermann, S. Bauerecker, and B. Neidhart, Analytical Bioanal. Chem. **372**, 122–127 (2002)], crystallization, and freezing processes in acoustically levitated drops, formation and characterization of monomolecular films on the surface of levitated drops, formation and growth of particles [S. Bauerecker and B. Neidhart, J. Chem. Phys. **109**(10), 3709–3712 (1998); Science **282**, 2211–2212 (1998)] and trapping of heavy gases [R. Tuckermann, B. Neidhart, E. G. Lierke, and S. Bauerecker, Chem. Phys. Lett. **363**, 249–354 (2002)] in stationary ultrasonic fields. Examples and results of these investigations will be presented.

9:05

4aPA4. Acoustically assisted phase-separation processes. Donald Feke (Dept. of Chem. Eng., Case Western Reserve Univ., Cleveland, OH 44106)

The separation and recovery of micron- to millimeter-sized secondary phases suspended in liquids is of fundamental importance in a variety of chemical, biochemical, and materials processing applications. These secondary phases can be solid particles, immiscible liquid droplets, or gas bubbles. In our laboratory, we have developed a separation technique in which the application of acoustic radiation pressure enables the phase separation within meshes having pore size up to 100 times the size of the secondary phase being retained. For the case of droplets and bubbles, the transport and coalescence induced by the acoustic radiation forces augments the separation efficiency. As a means to understand the fundamental physics of this process, we have developed both micro- and macro-scale models of the behavior of dispersed phases in our laboratory-scale acoustic devices and have performed experiments to validate these models. Micro-scale models predict the motion of individual dispersoids in response to the applied acoustic and flow fields, while the macro-scale models predict overall collection dynamics and breakthrough times. Generally, the correspondence between model predictions and experimental observation is excellent. These models thus reveal the fundamental physics that underlies the enhancement of the separation process by the acoustic radiation pressure.

9:30

4aPA5. Manipulation of flow fields using acoustic radiation pressure. Eugene H. Trinh (Exploration Systems Directorate, NASA HQ, Washington, DC 20546)

The accurate trapping and positioning of single drops and bubbles in outer host fluids is a capability that can be utilized to study the detailed fluid motion within these particles, as well as in the outer host fluid. The understanding of this motion and the ability to control its magnitude impacts the use of the mechanisms involved in mass and energy transport across the drop and bubble boundaries. We used acoustic radiation pressure generated by an ultrasonic field between 25 and 60 kHz in degassed water to trap centimeter-size air bubbles and to observe the correlation between the onset of capillary waves on the bubble surface and the generation of internal air flow inside the bubbles. The modulation of the ultrasonic field has also allowed the observation of coupling between the resonant bubble radial oscillation mode and the shape oscillation modes for smaller diameter samples. Other applications of radiation pressure combined with electrostatic levitation of charged drops will be discussed together with some applications in the field of crystal growth and liquid surface property measurements. [Work funded through NASA.]

9:55–10:10 Break

10:10

4aPA6. Manipulation of fluid objects using acoustic radiation pressure: Beyond drops and bubbles and retrospection. D. B. Thiessen, J. B. Lonzaga, C. F. Osterhourdt, W. Wei, and P. L. Marston (Phys. Dept., Washington State Univ., Pullman, WA 99164-2814)

Some standard applications of acoustic radiation pressure include the manipulation of drops and bubbles for the study of interfacial dynamics and light scattering. After a review, this presentation summarizes recent applications including radiation pressure effects on small flames and liquid cylinders [Wei *et al.*, J. Acoust. Soc. Am. **116**, 201–208 (2004)]. The dependence of the radiation pressure on the acoustic frequency has important consequences including liquid cylinder stabilization. While the radiation pressure is normally associated with sound propagation in the outer gas or liquid, our recent experiments also show that interfacial responses to modulated radiation pressure can be significant when the sound is applied internally to the manipulated object. For example, a liquid jet in air is an acoustic waveguide having a cutoff frequency inversely proportional to the jet diameter. When modulated ultrasound is applied to the jet supply liquid with the ultrasonic carrier frequency near-to or above the cutoff frequency, the jet can be made to oscillate at the modulation frequency of the radiation pressure, thus stimulating jet breakup. [Work supported by NASA.]

10:35

4aPA7. Acoustic particle manipulation devices driven by resonantly excited, cylindrical structures. Gregory Kaduchak, Greg R. Goddard, Christopher S. Kwiatkowski, and Dipen N. Sinha (Los Alamos Natl. Lab., Electron. and Electrochem. Mater. and Devices Group, Los Alamos, NM 87545)

An overview of acoustic particle manipulation devices that use resonantly excited structures as the displacement generators will be discussed. These devices rely on coupling of the structural modes of a hollow, cylindrical tube to the acoustic modes set up in its interior cavity. For practical applications, discussed for both particles in air and in liquids, the structural vibrations are tuned such that desired modal responses of the cavity are obtained. The effects on particle manipulation due to different internal configurations and symmetry-breaking of the cross-sectional geometry will be shown in experimental data and videos. Specific designs of these devices have demonstrated that they can be engineered into power-efficient assemblies. It will be shown that different structural and cavity modifications can be made to meet the resonance matching condition for several different azimuthal mode numbers. Material and geometry considerations for nonresonantly driven devices will also be discussed. Applications include sample positioning for real-time analysis in flow cytometers, and aerosol concentration and positioning for in-line aerosol detectors.

Contributed Papers

11:00

4aPA8. Forces generated in biological tissue-like media due to second-order effects in the focal region of an ultrasonic beam. Lev Ostrovsky (Zel Technologies/Univ. of Colorado, Boulder, CO 80305), Alexander Sutin, and Armen Sarvazyan (Artann Labs., Lambertville, NJ 08530)

Several techniques of medical diagnostics have been developed in the last decade based on the application of radiation forces of focused ultrasound. The most commonly considered physical mechanism of the radiation force generation is essentially related to ultrasound beam attenuation. However, in the case of short ultrasonic pulses typical for conventional ultrasonic imaging systems, an additional mechanism of force generation exists that is caused by acoustic nonlinearity of the media and does not explicitly depend on viscous dissipation. For water-like media such as human tissue, our theoretical description of the preceding effects is based on Eckart's theory of second-order effects in fluids, although modified by taking into account the shear modulus in the rotational component. In our theory, nonlinearity in a pulsed beam produces also a longitudinal component of an averaged force and a resulting displacement. Estimates show that for short acoustic pulses generated in conventional diagnostic ultrasound systems, the latter component can make a significant and even dominant contribution to the averaged displacement in the focal area of an ultrasonic beam.

11:15

4aPA9. An investigation of the effect of sound field distribution on the stability of the levitated object in the near-field acoustic levitation. Cheol-Ho Kim (35-3, HongChonRi, IbJangMyun, ChonAnSi, (330-825) Korea), Jeong-Guon Ih (KAIST Sci. Town, Guseong-dong, Yuseong-Gu, Taejon, Korea), and Sung-Il Lee (HongChonRi, IbJangMyun, ChonAnSi, (330-825) Korea)

Near-field acoustic levitation (NFAL) is a method that levitates an object or group of particles using the ultrasonic force. It is observed that, when a circular piston type of source is employed, stability problems can occur which makes the object position unpredictable. For example, a thin, circular planar object wobbles to lateral direction like a pendulum when its center is not precisely coincident with the center of source. In this study, the motive force, which makes the levitated object swivel through the center of source, is investigated by varying the sizes of the object and the source. To this end, a visualization of the sound field in the micro gap between the levitated object and source is made by using a photo view of the motion of fine powder scattered inside. The result is compared with the calculation result by BEM. A further investigation is experimentally made on any relation between the temperature distribution on the surface of the levitated object and the acoustic force. The experimental visualizations and numerical modeling results suggest that one may have a model describing the instability of the levitated object in NFAL, if a proper simplified mathematical model is set up in the further study.

11:30

4aPA10. Optical detection of the response of liquid jets to internal modulated ultrasonic radiation pressure. Joel B. Lonzaga, Curtis F. Osterhourdt, David B. Thiessen, and Philip L. Marston (Phys. Dept., Washington State Univ., Pullman, WA 99164-2814)

A liquid jet in air is an acoustic waveguide having a cutoff frequency inversely proportional to the jet diameter. Ultrasound applied to the jet supply liquid can propagate down the jet when the carrier frequency is near-to or above the cutoff frequency. The jet response to the internal acoustic radiation pressure of amplitude modulated ultrasound was monitored along the jet using an optical pseudo-extinction method. The jet profile oscillates at the frequency of the radiation pressure modulation and, where the response is small, the amplitude was found to increase in proportion to the square of the acoustic pressure amplitude as previously demonstrated for oscillating drops [P. L. Marston and R. E. Apfel, J. Acoust. Soc. Am. **67**, 27–37 (1980)]. Small amplitude deformations initially grow approximately exponentially with axial distance along the jet. Modulated radiation pressure can be used to stimulate large amplitude deformations and the breakup of the jet into drops. [Work supported by NASA.]

11:45

4aPA11. Use of a slightly curved reflector to enhance an airborne ultrasonic standing wave in a semi-infinite rectangular channel. Michael Anderson, Andrew Cluff, E. Clark Lemmon, and Gabriel Putnam (Dept. of Mech. Eng., Univ. of Idaho, Eng./Phys. Bldg. 324O, Moscow, ID 83844-0902)

One way to separate small particles from a moving air stream is to pass the stream through an intense ultrasonic standing wave. If the standing wave propagates perpendicular to the fluid flow direction, acoustic radiation pressure will move the particles to specific locations in the stream, where they can be collected at the stream outlet. Of primary importance is to achieve very high pressure amplitude in the standing wave, in spite of the presence of openings for fluid flow. In this presentation, the use of a slightly curved reflector within a flow channel to increase the amplitude of an ultrasonic standing wave is discussed. A finite element analysis was used to predict the amplitude of the standing wave that would be excited in a semi-infinite channel. A perfectly matched layer was used to account for the semi-infinite geometry. The finite element analysis showed that a significant gain in amplification within the channel could be achieved with a surprisingly small amount of reflector curvature. Experiments show that much of this gain can be obtained in practice.

Session 4aPP

Psychological and Physiological Acoustics: Psychological and Physiological Acoustics (Poster Session)

Li Xu, Chair
School of Hearing, Speech and Language Sciences, Ohio University, Athens, Ohio 45701

Contributed Papers

All posters will be on display from 8:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon.

4aPP1. A psychophysical measure of level-dependent shifts in the peak of the traveling wave. Christopher A. Brown, Changmo Jeung, and Sid P. Bacon (Arizona State Univ., P.O. Box 870102, Tempe, AZ 85287-0102, c-b@asu.edu)

Thresholds for detecting a 50-ms pure-tone probe in the presence of a 200-ms pure-tone forward masker were measured as a function of masker level (50, 70, or 90 dB SPL). The masker frequency was 250, 1000, 4000, or 6000 Hz, and the probe frequency was 1.00, 1.10, 1.20, 1.30, 1.40, or 1.60 times the masker frequency. For subjects 1–3, a 50-ms delay was used. For subjects 4–6, subject-dependent delays were chosen to keep the amount of masking to about 10 dB or less, to restrict the influence of off-frequency listening. These delays ranged from 50 to 110 ms. Subjects 1–3 showed the greatest amounts of masking on frequency in almost all conditions, whereas subjects 4–6 tended to show the most masking at frequencies above the masker frequency, but only at the highest masker levels and only at masker frequencies above 250 Hz. Although inconclusive, these results are in general agreement with both psychophysical and physiological studies that suggest a level-dependent basalward shift in the peak of the traveling wave along the basilar membrane. [Work supported by NIDCD.]

4aPP2. A systems and neurobiological model of nucleus angularis and nucleus laminaris in the barn owl. Brian J. Fischer (Dept. of Elec. and Systems Eng., Washington Univ. in St. Louis, Campus Box 8108, 660 S. Euclid, St. Louis, MO 63110) and Charles H. Anderson (Washington Univ. in St. Louis, St. Louis, MO 63110)

The representation of binaural sound localization cues was investigated in a model of the barn owl's auditory nerve, nucleus angularis (NA), nucleus magnocellularis (NM), nucleus laminaris (NL), and superior olivary nucleus (SON) using spiking leaky integrate and fire neurons. A phenomenological model of the barn owl's auditory-nerve spike response was created to reproduce responses to tonal stimuli and exhibit two-tone rate suppression effects. The representation of stimulus level by NA was investigated by decoding the envelope of the input signal from populations of NA model neurons where the envelope was defined as in a previous computational model of sound localization in the barn owl [B.J. Fischer and C.H. Anderson, Adv. NIPS **16** (2004)]. Simulations demonstrated that inclusion of NA neurons with type IV responses increased the accuracy of the representation of the stimulus envelope. The effect of inhibition on the output of the population of NL model neurons was investigated and compared to a cross-correlation model. The output of the model SON provides inhibitory feedback to ipsilateral NA, NM, and NL model neurons. Simulations demonstrated that inclusion of inhibition of NL neurons by SON output increased peak–trough differences in ITD tuning curves.

4aPP3. The influence of spread of excitation on the temporal modulation transfer function (TMTF) for sinusoidal carriers. Rebecca E. Millman and Sid P. Bacon (Dept. of Speech and Hearing Sci., Arizona State Univ., P.O. Box 870102, Tempe, AZ 85287-0102)

The improvement in modulation detection thresholds with increasing stimulus level has been attributed to listening on the high-frequency side of the excitation pattern, where the growth of excitation is more linear, or to an increase in the number of channels via spread of excitation. TMTFs were measured for sinusoidal amplitude modulation (SAM) applied to a 1-kHz sinusoidal carrier, whose level varied between 10 and 80 dB SPL. Modulation detection thresholds improved as the carrier level increased. To minimize the use of spread of excitation at a carrier level of 80 dB SPL, various tonal restrictors were used alone and in combination. The restrictors were presented continuously and their frequencies were 501, 801, 1210, or 1510 Hz. Only the combination of the 801 and 1210-Hz restrictors resulted in an increase in modulation detection thresholds to values comparable to those obtained at 30 dB SL. These data suggest that listening on the high-frequency side of the excitation pattern alone cannot account for better modulation detection thresholds at higher levels, and that spread of excitation above and below the carrier is required for optimal performance. [Work supported by NIDCD.]

4aPP4. Auditory evoked potentials to changes in interaural cross correlation. Helmut Riedel, Helge Lüddemann, and Birger Kollmeier (Medizinische Physik, Universität Oldenburg, 26111 Oldenburg, Germany)

The objective of this study is to analyze the representation of specific binaural stimulus features in the human brain and to compare the results to psychoacoustic data. Late auditory evoked potentials to changes in the interaural cross correlation (IACC) of continuous bandpass noise (100–2000 Hz) were recorded, 1000 sweeps per stimulus condition were averaged. In experiment 1 a brief "binaural gap" with a gap IACC was interposed in the noise with a reference IACC. N1 and P2 amplitudes increased with increasing gap duration. For uncorrelated reference IACC (0/1/0 and $0/-1/0$ conditions) objective thresholds were comparable to psychoacoustic data. For uncorrelated gap IACC (1/0/1 and $-1/0/-1$ conditions) objective thresholds were about four times larger than in psychoacoustic measurements. In experiment 2 reference and test interval had the same length of 800 ms and the magnitude of change in IACC was varied. Objective IACC-JNDs are larger for an uncorrelated reference than for IACCs of ± 1. For the uncorrelated reference, test IACCs towards $+1$ elicited larger AEP components than test IACCs towards -1. In contrast to binaural difference potentials used at the brain stem level, employing changes in IACC constitutes a more direct approach to objectively investigate the binaural system in humans.

4aPP5. Psychophysical evidence for the existence of envelope information in the internal auditory representation of signal fine structure. Frederic Apoux, Rebecca E. Millman, Christopher A. Brown, Sid P. Bacon (Psychoacoust. Lab., Dept. of Speech and Hearing Sci., Arizona State Univ., Tempe, AZ 85287-0102), and Neal F. Viemeister (Univ. of Minnesota, Minneapolis, MN 55455)

This study examined the extent to which the auditory system can recover envelope information from phase information contained in signal fine structure [O. Ghitza, J. Acoust. Soc. Am. **110**, 1628–1640 (2001)]. A 2IFC paradigm was used to assess amplitude modulation (AM) detection performance for stimuli whose envelope was removed using the Hilbert transform, leaving only the fine structure. One interval contained the fine structure of an amplitude modulated stimulus, the other the fine structure of an unmodulated stimulus. The first experiment measured AM detection for a fixed modulation index ($m = 0.95$). The original signal carriers were broadband noise, harmonic tone complexes, and positive Schroeder-phase complexes modulated at rates of 5, 10, or 15 Hz. Listeners were unable to detect AM when presented with the fine structure only, suggesting little envelope recovery. The second experiment measured AM detection thresholds for modulation rates of 5, 10, or 15 Hz imposed, before combination, on one of three nonharmonically related tones. Preliminary results indicate that thresholds are comparable whether listeners are presented with the full signal or fine structure only, suggesting that the depth of the recovered AM is similar to that of the original AM for these stimuli. [Work supported by NIDCD.]

4aPP6. Effects of talker variability on vowel recognition by cochlear implant users and normal-hearing subjects listening to cochlear implant simulations. Yi-ping Chang, Qian-Jie Fu (Dept. of Biomed. Eng., Univ. of Southern California, Los Angeles, CA 90089, yipingch@usc.edu), and Qian-Jie Fu (DAIP, Los Angeles, CA 90057)

Different talker characteristics can significantly affect speech recognition, especially under conditions of reduced spectral resolution. In the current study, single- and multi-talker vowel recognition was tested with both cochlear implant (CI) patients and normal-hearing (NH) subjects listening to CI simulations. CI users were tested with their everyday processors. NH subjects were tested using four-channel acoustic processors in which the envelope cutoff frequency (20, 160 Hz) and the carrier band (noise-band, sine-wave) conditions were varied. Results showed that CI users' single-talker vowel recognition was significantly better than multi-talker recognition. Among the simulations, both single- and multi-talker vowel recognition was best with the 160 Hz/sine-wave processor. Also, single-talker performance was significantly better than multi-talker performance only with the 160 Hz/sine-wave processor; there was no difference between single- and multi-talker recognition with the 20 Hz/sine-wave or noise-band processor. While the effects of talker variability were less for the 20 Hz/sine-wave and noise-band processors, overall vowel recognition was also significantly lower. Listeners' sensitivity to talker variability may indicate their ability to register the relative vowel space for individual talkers. Without sensitivity to talker variability, multi-talker vowel space may collapse, causing cross-talker differences to be confused with phonetic differences. [Research is supported by NIDCD.]

4aPP7. Revisiting the quartic model for early identification of noise-induced hearing loss. William J. Murphy and John R. Franks (NIOSH Hearing Loss Prevention Team, 4676 Columbia Pkwy., MS C-27, Cincinnati, OH 45226-1998, wjm4@cdc.gov)

In a 1976 article by Cooper and Owen ["Audiologic profile of noise induced hearing loss," Arch. Otolaryngol. **102**, 148–150 (1976)], they advanced a quartic model to perform a least-squares fit to an audiogram for the purpose of identifying a noise-induced notch indicative of early onset of noise-induced hearing loss. This model has been revisited for the purpose of better quantification of a set of rules that will best identify the presence or absence of a noise notch. A subset of 1660 subjects' right and left audiograms collected for the National Health and Nutrition Examination Survey (NHANES IV) was fit with a quartic model. As well, several hearing scientists and audiologists within NIOSH examined and classified the audiograms as having a notch or not. Preliminary results suggest that the choice of notch criteria dramatically affects the identification accuracy when correlated with clinical judgments. This paper will present how these rules were developed and how the quartic model may be interpreted and applied in a hearing conservation program.

4aPP8. Early indicators of noise-induced hearing loss: Issues for consideration. Mark R. Stephenson, William J. Murphy, and John R. Franks (NIOSH, Hearing Loss Prevention Team, 4676 Columbia Pkwy., MSC-27, Cincinnati, OH 45226-1998, mos9@cdc.gov)

A recent article in the NHCA Spectrum [Dobie and Rabinowitz, **20**, 8–11 (2003)] considered several methods for identifying the early onset of noise-induced hearing loss (NIHL). Among the methods they examined was the Notch Index: the difference in the pure-tone average of 2, 3, and 4 kHz and the pure-tone average of 1 and 8 kHz. This statistic was expected to be positive for NIHL where noise has produced a threshold shift in the 2- to 4-kHz region but not adversely affected the thresholds at 1 and 8 kHz. The Notch Index was examined for a subset of 1660 subjects' right and left audiograms collected for the National Health and Nutrition Examination Survey (NHANES IV). The Notch Index did not prove to be effective as expected. However, modifications to the calculation, the inclusion of 6 kHz, or calculation of the area difference provided better identification of the early onset of NIHL. This paper will examine other issues such as work-relatedness and potential gender effects that might be influenced by differential trends in ear canal size.

4aPP9. Relating the acoustic vowel space to perceptual vowel space in cochlear implant simulation. Chuping Liu, Qian-Jie Fu (Dept. of Elec. Eng., Univ. of Southern California, Los Angeles, CA 90007, chuping1@usc.edu), and Qian-JIe Fu (Los Angeles, CA 90057)

Because of limited spectral resolution, the acoustic differences between important speech features may not be preserved by cochlear implants (CIs). The present study compared the acoustic vowel space of spectrally degraded speech with the perceptual vowel space of normal-hearing (NH) subjects to CI simulations in which the spectral resolution, frequency allocation, carrier bands (noise-band or sine-wave) and signal-to-noise ratios (SNRs) were varied. For each speech-processing condition, acoustic differences between vowel tokens were calculated by extracting the mel-frequency cepstrum coefficients, time-aligning the tokens, then computing the weighted Euclidean distances. Results showed that the acoustic vowel space monotonically expanded with increasing SNR and spectral resolution; sine-wave carriers also expanded the acoustic vowel space, relative to noise-band carriers, for all SNR conditions. We will compare the acoustic vowel space to NH subjects' perceptual data for these speech-processing conditions. [Research is supported by NIDCD R01-DC-004993.]

4aPP10. Comparison of acoustic features of time-compressed and natural speech. Akiko Kusumoto and Nancy Vaughan (NCRAR, Portland VA Medical Ctr., 3710 SW US Veterans Hospital Rd., Portland, OR 97207)

Time-compressed speech is used as a speech material to examine the upper limits of processing speed in young and old listeners with and without hearing loss. Generally, older normal-hearing listeners at 60% time compression perform similarly to younger normal-hearing listeners at 70% time compression on speech recognition test. This may be due to slowing of processing speed that typically accompanies aging. It has not been determined if time-compressed speech material carries spectral distortions or other acoustic changes that affect the ability to process rapidly. If so, differences in performance between younger and older listeners may be attributable in part to acoustic distortions other than temporal changes. In

this work, we examine acoustic feature changes in time-compressed speech processed using synchronized overlap-add (SOLA) method. Also, we discuss which parts of speech (consonants or vowels) are more affected by time compression. In the temporal domain, we compare the duration of consonants, vowels and voice onset times (VOT) in both 50% compressed speech and natural speech. We also examine the long-term power spectra of both signals. Possible effects on speech intelligibility will be discussed.

4aPP11. Correlation of speech recognition by cochlear implant users with metrics based on thresholds and loudness levels. Li Xu (School of Hearing, Speech and Lang. Sci., Ohio Univ., Athens, OH 45701) and Bryan Pfingst (Univ. of Michigan, Ann Arbor, MI 48109-0506)

Intersubject variability in speech recognition is a persistent characteristic of cochlear implant users, despite significant improvements in average performance. This study examines simple psychophysical measures, maximum comfortable loudness levels (C-levels) and dynamic ranges (DRs) that may serve for diagnosis of conditions that affect speech recognition. Eighteen subjects who had used Nucleus CI24 cochlear implants for a minimum of 1 year participated in the study. T and C levels were measured using the method of adjustment. Consonant, vowel, and sentence recognition were tested using the subjects everyday speech processors (SPEAK or ACE). Across-site variation (ASV) was quantified for each subject using the variance of the T and C levels across all tested stimulation sites. Mean DRs (C levels minus T levels) for all tested sites were calculated for each subject. The results showed a moderate but significant negative correlation between speech-recognition performance and ASV of the T and C levels measured with both monopolar and bipolar electrode configurations. Mean DRs also correlated with speech performance. These results support the hypothesis that the large across-site variation in detection thresholds and small dynamic range reflect conditions that are detrimental to speech-recognition with cochlear implants. [Work supported by NIDCD Grant DC03808.]

4aPP12. Effects of short-term training on the recognition of spectrally shifted vowels by non-native listeners. Tianhao Li, Qianjie Fu, and Geri Nogaki (DAIP, House Ear Inst., 2100 W. 3rd St., Los Angeles, CA 90057, tianhaol@usc.edu)

Previous studies have shown that normal-hearing (NH) native listeners can rapidly adapt (though not always completely) to spectrally mismatched speech with short-term training. However, it is unclear how short-term training might affect the performance of non-native listeners. The present study investigated whether non-native listeners' recognition of spectrally shifted English vowels could be improved by moderate short-term training. Recognition of 12 American English vowels (in c/V/c format) was trained and tested in five Chinese-speaking NH subjects; speech tokens were processed by an eight-channel noise-band vocoder in which the spectrum was upwardly shifted by one octave. After baseline testing subjects were trained with shifted speech for 5 consecutive days; daily training consisted of three to four sessions, each of which included a 5-min preview of the test tokens and a 10-min test. Results showed that for all subjects, recognition of spectrally shifted English vowels improved significantly over the 5-day training period. These results show that non-native listeners are able to rapidly adapt to spectrally mismatched speech with short-term training. The ability to fully accommodate spectrally shifted speech may depend on non-native listeners experience with the second language. [Research supported by NIDCD R01-DC-04993.]

THURSDAY MORNING, 18 NOVEMBER 2004 ROYAL PALM SALON 5, 9:00 TO 11:15 A.M.

Session 4aSA

Structural Acoustics and Vibration: Vibration of Sports Equipment

Daniel A. Russell, Chair

Science and Mathematics Departments, Kettering University, 1700 West Third Avenue, Flint, Michigan 48504-4898

Invited Papers

9:00

4aSA1. Vibration and sound radiation of solid wood and tubular metal baseball bats as a function of ball–bat location. Robert D. Collier (Thayer School of Eng., Dartmouth College, Hanover, NH 03755), Kenneth Kaliski (Resource Systems Group, Inc., White River Junction, VT 05001), and James A. Sherwood (Univ. of Massachusetts Lowell, Lowell, MA 01854)

The ''crack of the bat'' is an important part of the game of baseball played with solid wood bats. The spectral characteristics of the radiated sound depend on the location of the ball–bat impact location along the length of the barrel and the resulting bat vibration. Balls hit on the sweet spot generally are hit harder and result in a distinct and recognizable ''crack'' sound. On the other hand, balls hit in on the handle or off the end of the barrel result in different modes of vibration and low-frequency radiation, i.e., a ''thunk-like'' sound. Analytical predictions and modal analyses, supported by both laboratory and field measurements, provide a more comprehensive picture of bat vibration and radiated sound relationships. Comparisons are made with tubular metal bats which exhibit the narrow-band signature due to excitation of the cylindrical breathing (bell) modes which are essentially independent of hit location. The differences in bat vibration and sound radiation can provide important clues for baseball players in the field.

9:30

4aSA2. Cricket bats: Some performance and vibration control issues. Sabu John, Cao Jialong, Fatmir Gutaj (RMIT Univ., School of Aerosp., Mech. and Manuf. Eng., Bundoora East Campus, Bundoora, VIC 3083, Australia), and Tom Molyneaux (RMIT Univ., Melbourne, VIC 3001, Australia)

This paper will address some of the issues in an on-going research project involving the performance optimization and vibration control of Cricket bats. In many ways, the rules of the game have made our task a little easier by restricting exactly what we can modify in order to improve performance. For example, the blade or the hitting proportion of the Cricket bat has to remain as wood but

its geometry can be altered to fit within a volume-based specification. The handles specification, however, is not specified. Hence by our interpretation of the rules of the game, this lack of specificity for the handle in the rules provides an opportunity to improve the performance of the Cricket bat as a whole by making material and geometric modifications to the handle. This paper will address the investigation of bat performance from a dynamics and vibration analyses perspective. This paper will report on parameters such as bat–ball contact time and translation, clustering of distal node points (DNPs) to mimic an enhanced sweet spot. Results showing performance comparisons between traditional and composite material-handle bats will be presented along with computational and analytical results for various handle shapes.

10:00–10:15 Break

10:15

4aSA3. Dynamic characterization of golf equipment. Tom Mase (Composite Mater. and Structures Ctr., Michigan State Univ., East Lansing, MI 48824)

In the golf swing the shaft, head, and ball respond at different time scales. The shaft is swung from the top of the back swing to impact in less than a half second. In that time, the shaft loads and unloads in different ways for different shafts. Natural frequencies and mode shapes characterize different shafts uniquely making some shafts better than others for individual golfers. Shafts are benchmarked by measuring their stiffness profile along the length of the shaft. Knowing this, a shaft finite element model is generated from which the natural frequency is found. Shaft zone frequencies, an important fitting method, are cantilevered natural frequencies found from the stiffness profile. As for golf balls, their core and cover mechanical properties influence performance and feel. The loading of the golf ball occurs in 0.0005 s. Golf ball stiffness and damping properties are measured using DMA and universal testing machines. Distance balls have cores that are nominally twice as stiff as performance balls, making them feel harder to the golfer. Performance and distance balls also respond to strain rate changes differently. Ultimately, stiffness properties affect the initial velocity, spin, and launch angle which dictate the distance the ball flies.

10:45

4aSA4. The sweet spot of a hollow baseball or softball bat. Daniel Russell (Sci. and Mathematics Dept., Kettering Univ., Flint, MI 48504)

The ''sweet spot'' of a baseball bat is often defined as a region approximately 5–7 in. from the barrel end of the bat where the performance (as measured by a batted-ball speed) is the greatest and the vibrational sensation (the sting felt in the hands) is minimized. In terms of the vibrational properties of a bat, the sweet spot region includes the nodes of the first two bending vibrational modes and is often also identified as being related to the center-of-percussion (COP). This paper will discuss recent research which shows that the COP has no bearing on performance and little, if any, on the perception of feel. We will also discuss the contribution of the bending vibrational modes to the location of the sweet spot and compare results for a wide variety of wood, metal and composite baseball and softball bats. We will attempt to correlate the sweet spot defined in terms of performance and that defined in terms of vibrational sensation. Finally, we will discuss the contribution of the hoop mode, unique to hollow bats, which correlates strongly to performance, and which may also influence the perception of feel.

THURSDAY MORNING, 18 NOVEMBER 2004 SUNRISE AND SUNSET ROOMS, 8:00 A.M. TO 12:00 NOON

Session 4aSC

Speech Communication: Foreign-Accented Speech: Production, Perception and Applications (Lecture/Poster Session)

Amee P. Shah, Chair

School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec, H3V 1C2 Canada

Chair's Introduction—8:00

Invited Papers

8:05

4aSC1. Why do non-native speakers have a foreign accent? A three-dimensional perspective. Amee P. Shah (School of Commun. Sci. and Disord., McGill Univ., 1266 Pine Ave. West, Montreal, QC H3V1C2, Canada, amee.shah@mail.mcgill.ca)

A three-dimensional perspective, following the speech-chain model, is taken in arriving at the variables that influence the production and perception of foreign-accented speech. Essentially, research to date indicates the interactive role of all three communication components of the speech-chain model. First, speech-related variables, i.e., the interlanguage differences in the phonetic patterns of the speech, of L2 speakers compared to the L1 speech patterns influence listeners perception of accentedness of non-native

speech. Second, speaker-related variables (i.e., differences in age, other psychological variables) cause the non-native speakers to have difficulties in learning to map new sounds of the L2 onto their existing L1 phonetic system, thus resulting in foreign-accented speech patterns. Lastly, differences in listener-related factors (i.e., L1 of the listener, prior linguistic experience, amount of exposure, listening-conditions in which they hear the accented speech) have been found to influence the perception of foreign-accentedness of speech. Past and current findings will be brought to bear upon this issue that has implications in theoretical understanding of speech perception, as well as practical applications in accent-modification and ESL classroom training programs.

8:25

4aSC2. Understanding foreign-accented speech. Ann Bradlow (Linguist. Dept., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208)

Why is foreign-accented speech hard to understand? In one view, foreign-accented speech intelligibility is poor due to its marked deviation from the native talker norm. An alternative view claims that if the talkers and listeners shared knowledge systems are sufficiently aligned with each other, foreign-accented speech intelligibility can be good despite its deviation from the native talker norm. Results from two studies in support of this latter alternative will be presented. Study 1 investigated whether intelligibility of Chinese-accented English could be improved for native English listeners through perceptual training. This study showed that if given exposure to multiple talkers of Chinese-accented English during training, native English listeners could achieve talker-independent adaptation to Chinese-accented English. Study 2 compared intelligibility of foreign-accented and native-accented English for non-native and native listeners. Results showed that native listeners found native-accented English more intelligible than foreign-accented English, but non-native listeners found foreign-accented English as intelligible as native-accented English. Interestingly, this pattern was obtained even for non-native listeners and talkers from different language backgrounds. Taken together, these findings demonstrate that intelligibility of foreign-accented speech varies depending on the degree of talker–listener attunement rather than on the extent of its deviation from a native talker norm *per se*.

8:45

4aSC3. Theoretical models of non-native speech perception: Implications for perception of foreign-accented speech. Catherine Best (MARCS Auditory Labs., Univ. Western Sydney, Penrith NSW 1797 Australia, and Haskins Labs., 270 Crown St., New Haven, CT 06511)

Linguistic experience systematically affects speech perception, specifically for non-native phonemes and contrasts. Native language influences are strong in adults, and emerge in infancy. Several theoretical models of non-native speech perception have been proposed, including the Perceptual Assimilation Model (PAM: Best), Speech Learning Model (LSM: Flege), and Native Language Magnet model (NLM: Kuhl). Findings that infants learn statistical distributions of phoneme sequences, and phonetic details, converge on the notion that exposure to specific phonetic patterns affects the perception of phonological structure. A reasonable extrapolation is that perception of even native utterances should be influenced by their pronunciation with either familiar or unfamiliar phonetic structure. That is, if a listener is inexperienced with a given foreign accent, or another native language dialect, perception should be predictably affected when native phonemes/words are produced with those accents. This issue has received modest research attention, which has focused on sentences/words rather than phonemes. Findings suggest that unfamiliar accents do reduce recognition and intelligibility of native sentences and words. The implications of non-native speech models, particularly PAM and SLM, for understanding and studying the impact of foreign accent on perception of native phonemes and contrasts, as well as longer utterances, will be discussed. [Work supported by NIH.]

9:05

4aSC4. Bridging the gap between speech research and *L*2 pedagogy. Murray J. Munro (Dept. of Linguist., Simon Fraser Univ., Burnaby, BC V5A 1S6, Canada, mjmunro@sfu.ca)

In studies of *L*2 speech learning it is common to draw conclusions (sometimes as an afterthought) about the potential pedagogical value of research findings, particularly when they provide insights into specific pronunciation difficulties faced by *L*2 learners or when they indicate improved perception or production as a result of laboratory training. However, many important research findings have, as yet, had only a minor impact on second language teaching, and there remains a significant gap between what has been empirically established in the research laboratory and what is actually practiced in the classroom. This presentation will identify some of the reasons for this disparity through an examination of goals and priorities that are commonly accepted by pedagogical specialists. In particular, it will be argued that the implementation of research findings depends on establishing practical ways of improving *L*2 learners' speech intelligibility (as opposed to mere accent reduction) for a diverse audience of interlocutors in authentic interactive settings. In general, there is a need for research that is specifically motivated by pedagogical concerns and that is informed by an awareness of a wide range of current issues in applied linguistics. [Research supported by SSHRC.]

9:25

4aSC5. Please understand me: Toward modeling of non-native speech in automatic speech recognition Laura Mayfield Tomokiyo (Cepstral, LLC 1801 E. Carson St., Pittsburgh, PA 15203)

Non-native speech represents a fundamental challenge for automatic speech recognition (ASR). Most major advances in ASR technology assume consistency in certain features that native speakers of a language share, such as coarticulatory patterns, control of syntax, and conventions in lexical choice. Learners of a language, however, traverse very individual, and often unstable, trajectories toward spoken competence. While ASR systems can often handle highly proficient non-native speech with little or no modification of their statistical models, recognition performance can be very poor as proficiency levels and speaker fluency fall. Interest in handling foreign-accented speech in ASR has emerged from two different directions: large-scale applications such as dictation and telephony systems, and educational applications offering tutoring and diagnostics. Developers of the former have seen the need for adaptation to

non-native speech as their systems become more widely deployed in the real world. And, as speech technology has improved, it is seen as a potentially viable tool in foreign-language education. This presentation will highlight approaches to and advances in treatment of non-native speech in ASR for these two domains, examining both characteristics of learner speech and methods for improving recognizer performance.

9:45–9:55 Summary Remarks

9:55–10:05 Break

10:05–10:55 Panel Discussion

10:55–11:00 Break

Contributed Papers

All posters will be on display and all authors will be at thier posters from 11:00 a.m. to 12:00 noon in the Sunset Room.

4aSC6. Relations between native Spanish speakers' perception and production of English vowels. Kimberly A. Lydtin (Dept. of Speech, Commun. Sci., and Theatre, St. John's Univ., Jamaica, NY 11439; Grad. Ctr, CUNY; NYC Dept. Educ.), Fredericka Bell-Berti, Peggy Jacobson, and José Centeno (St. John's Univ., Jamaica, NY 11439)

This study is about the relation between native Spanish speakers' perception and production of English vowels and their age of arrival (AoA) in the U.S. The participants were also grouped by dialect, Central American/ Caribbean or South American Spanish. In a listening task, the participants identified the words in the set ''heed, hid, head, had, hod, hawed, hood, who'd, hud, heard'' and then produced four repetitions of each of the ten words for spectral and temporal analysis. Finally, they indicated their sensitivity to contrasts between the vowels of the English vowel pairs /i–I/, /I–ϵ/, and /æ-$\gtrsim$/, rating them from ''very similar'' to ''very dissimilar.'' The results of our acoustic analyses will be compared with the participants' identification data to determine how the vowel space of $L1$ (primary language) affect the production and perception of $L2$, and also with the listeners' sensitivity to the vowel–contrast pairs.

4aSC7. Why are non-native (L2) utterances longer than native-produced utterances? James E. Flege and Anders Hojen (Univ. of Alabama at Birmingham, CH19 301, 1530 3rd Ave. S., Birmingham, AL 35294-2041)

Research has repeatedly shown that utterances produced by non-native speakers are longer than those of native speakers, even when produced as rapidly as possible. The non-native speaking rate effect is known to be greater for late than early L2 learners, but may disappear for late learners when very short-duration L2 utterances are examined. This study examined the duration of English utterances produced by native English (NE) speakers and late native Spanish (NS) learners of English ($n=20$ each). For normal-rate utterances, durations were significantly longer for the NS than NE group for syntactically complex sentences (e.g., The boy saw that Gary forgot a box of pencils) but not simple sentences (Gary forgot a box of pencils) or noun phrases (a box of pencils). However, the NS-NE difference reached significance for all three utterance types at a fast rate. In Experiment 2, the NS group took longer than the NE group to produce the digit string 1 2 3 4 5 6 7 as rapidly as possible. Taken together, the results suggest that differences in both the articulation of L2 speech sounds and higher-order planning (including the formulation of syntactic representations) contribute to the nonnative speaking rate effect. [Work supported by NIH.]

4aSC8. Perceptual shift of the /d/-/t/ voice onset time boundary following exposure to French-accented English. Constance M. Clarke and Paul A. Luce (Dept. of Psychol., Univ. at Buffalo, 245 Park Hall, Buffalo, NY 14260, cclarke2@buffalo.edu)

Previous work suggests native English listeners perceptually adapt to foreign-accented English, as indicated by faster reaction times, after only brief exposure [C. M. Clarke, Proc. ICSLP, 253–256 (2002)]. The present study explored whether changes in the perceptual criteria for phonetic categories may contribute to this increased processing speed. Categorization of /d/ and /t/ was tested using a voice onset time (VOT) continuum. French-accented English was used as the exposure accent because it typically contains prevoiced /d/'s and short-lag /t/'s. Native English has short-lag /d/'s and long-lag /t/'s. Categorization was tested prior to exposure and again following 20, 40, and 60 accented sentences. Each set of 20 sentences contained 20 syllable-initial /t/'s and /d/'s. It was predicted that native English listeners categorization boundaries would shift to lower VOTs after exposure to French-accented speech. Native English speech was used in a control condition. To test the contribution of exposure to specific exemplars of /d/ and /t/ versus exposure to more general characteristics of an accent, two other conditions were included: native English speech with /d/'s and /t/'s digitally modified to be French-like, and French-accented speech with /d/'s and /t/'s digitally modified to be English-like. [Work supported by NIH.]

4aSC9. Accuracy and variability in vowel targets produced by native and non-native speakers of English. Shawn L. Nissen (Dept. of Audiol. and Speech-Lang. Pathol., Brigham Young Univ., 138 TLRB, Provo, UT 84602), Bruce L. Smith (Univ. of Utah, Salt Lake City, UT 84112), Ann Bradlow, and Tessa Bent (Northwestern Univ., Evanston, IL 60208)

The primary issue explored in this investigation concerned how accurately and consistently non-native, adult speakers learning English as a second language produce vowel targets when compared with native speakers. One possibility is that non-native speakers show average formant and duration values similar to those of native speakers but are nonetheless more variable across repetitions. Alternatively, they might be ''off-target'' but still reasonably consistent in such productions. A group of native speakers (3F and 3M) and a group of Mandarin-speaking non-native subjects who had recently come to the United States (3F and 3M) produced at least ten repetitions of each of 20 different English words embedded in a carrier phrase. The first two formant frequencies and duration were measured for seven different vowels. Preliminary findings suggest that the non-native speakers' vowel formants were generally quite similar to those of the native speakers, whereas their vowel durations deviated from those of the native speakers. Moreover, the non-native speakers exhibited greater within-speaker variability in their vowel productions (across word repetitions) than the native speakers, as well as a greater range in vowel formants and durations across speakers. Findings will be discussed in reference to their implications for second language acquisition and foreign-accented speech perception.

4aSC10. Global foreign accent in English sentences produced by native Spanish learners of English: A 5-year longitudinal study. Satomi Imai, James E. Flege (Div. of Speech and Hearing Sci., Univ. of Alabama at Birmingham, Birmingham, AL 35294), Ratree Wayland (Univ. of Florida at Gainesville, Gainesville, FL 32611), and Surya P. Talla (Univ. of Alabama at Birmingham, Birmingham, AL 35294)

Second language acquisition theorists posit that the performance of adult second language (L2) learners improves up to a certain point but may then fossilize, that is, show little further improvement. However, few empirical studies have tested for changes in L2 production by individual adult L2 learners over an extended period of time. In the present study, sentences produced by adult native Spanish (NS) learners of English over a 5-year period will be evaluated for changes in pronunciation. Fifteen NS adults (mean age = 34 years when first tested) who had lived in Birmingham, AL and three native English (NE) controls produced four English sentences at 6-month intervals for 5 years. NE listeners will rate the sentences for foreign accent. One-way repeated measures ANOVAs will be carried out to determine if individual NS participants, or the group as a whole, became less foreign-accented over the study interval. Individual differences in demographic variables that might have affected degree of foreign accent (e.g., age of arrival in the U.S., chronological age of testing, percent English and Spanish use, family status) will be discussed in relation to any improvement in L2 production. [Work supported by NIH.]

4aSC11. Perceived accents and comprehensibility of native and nonnative English. Ayako Ikeno (Robust Speech Processing Group, The Ctr. for Spoken Lang. Res., Dept. of Linguist., Univ. of Colorado at Boulder, CO 80309-0295), John H. L. Hansen, and Barbara Fox (Univ. of Colorado at Boulder)

This study investigates the relation between perceived accentedness and comprehensibility of utterances spoken by native and nonnative English speakers. Native and nonnative English listeners heard six types of stimuli: single words, 2–10-word utterances from native and nonnative English speakers. Listeners classified the stimuli as either native or nonnative English speech and transcribed some utterances. Overall, results from the classification and transcription tasks indicate that perceived native-ness or nonnative-ness does not directly correspond to good or poor comprehensibility, as measured by transcription scores. Some nonnative utterances were classified as native speech by both native and nonnative listeners, although the transcription accuracy was not necessarily high. In other cases, native speech was correctly identified as native; however, dialectal pronunciation variation seemed to contribute to low transcription accuracy in some cases. On the other hand, although the transcriptions were accurate, native speech was sometimes classified as nonnative. In other cases, nonnative utterances were correctly identified as nonnative, and the speech was transcribed accurately. These results suggest that familiarity with certain dialects or accents influences listeners' perception and comprehension for both native and nonnative speech, and that perceived nonnative-ness may not be the main cause for nonnative speech to be less comprehensible.

THURSDAY MORNING, 18 NOVEMBER 2004 ROYAL PALM SALONS 3 & 4, 8:00 A.M. TO 12:00 NOON

Session 4aSP

Signal Processing in Acoustics, Underwater Acoustics, Animal Bioacoustics, Noise, Acoustical Oceanography and Engineering Acoustics: Time Delay Estimation, Localization and Tracking in Acoustics Part I

Christopher O. Tiemann, Chair
Applied Research Laboratory, University of Texas at Austin, P.O. Box 8029, Austin, Texas 78731-8029

Chair's Introduction—8:00

Invited Papers

8:05

4aSP1. An introduction to time-delay estimation, localization, and tracking in Acoustics. Leon H. Sibul (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804-0030)

Acoustic source localization and tracking has a long, productive history in both passive and active sonar signal processing. More recently, microphone arrays have been used to localize and track air acoustic noise sources. An overview of basic localization and tracking techniques of acoustic radiating sources and scatters is presented. For a wide range of applications, many *ad-hoc* passive and active source localization techniques have been investigated. Concepts from the statistical estimation theory, such as maximum likelihood estimators and Cramér-Rao lower bounds (CRLB), are useful benchmarks for evaluation of these practical *ad hoc* systems. In both air and underwater media, bearing and range estimates are determined by time-delay estimates; hence, properties of the time-delay estimates are of critical importance. In the case the of narrow-band active sonar, localization and tracking is based on time delay, Doppler, and direction of arrival estimation. In the case of wideband sonar and tracking of high-speed air acoustic sources, tracking and range estimates are determined from time-scale dilation estimates. CRLBs for active sonar are determined by signal-to-noise ratio, signal properties of signal ambiguity functions such as effective duration and mean-square bandwidth of the transmitted signal, and array parameters. [This work has been supported by ONR.]

8:35

4aSP2. Obtaining optimal time-delay, source localization and tracking estimates in free space and in an ocean waveguide. Nicholas C. Makris, Purnima Ratilal, Michele Zanolin, and Ioannis Bertsatos (MIT, 77 Massachusetts Ave., Cambridge, MA 02139)

Asymptotic methods for obtaining optimal time-delay, source position and track estimates are derived for measurements in free-space and in an ocean waveguide using nonlinear statistical inference theory. The methods are useful for determining necessary conditions on controllable variables such as signal-to-noise ratio and sample size to attain desired design error thresholds. For free-space time-delay estimation, it will be shown both analytically and intuitively that the matched filter cannot be unbiased and attain minimum variance unless the kurtosis of the signal's energy spectrum exceeds the signal-to-noise ratio [Naftali and Makris, J. Acoust. Soc. Am. **110**, 1917–1930 (2001)]. In an ocean waveguide, it will be shown that signal-to-noise ratios and sample sizes necessary to attain practical localization error thresholds can become prohibitively large when random waveguide inhomogeneities such as internal waves degrade intermodal coherence.

9:00

4aSP3. Navigation from time differences of arrival. John L. Spiesberger (Dept. of Earth and Environ. Sci., 240 South 33rd St., Univ. of Pennsylvania, Philadelphia, PA 19104, johnsr@sas.upenn.edu)

A brief history and a few modern developments are given for hyperboliclike navigation systems in which location is estimated from measurements of the difference in the propagation time of a signal from two or more known locations. Today, hyperbolic navigation is used in locating ships, calling animals, GPS receivers, cellular telephones, and other objects. Topics discussed include the minimum number of receivers required for location, the effects of environmental fluctuations and receiver positioning errors on location accuracy, methods for obtaining probability density functions for location, and methods for allowing the speed of signal propagation to vary from path to path. Recent developments indicate that this 67-year-old navigation problem is not mature.

9:25

4aSP4. The use of acoustic multipath for localization of sperm whales. Aaron M. Thode (Marine Physical Lab., Scripps Inst. Oceanogr., 9500 Gilman Dr., La Jolla, CA 92093-0238)

Sperm whales produce impulsive sounds, or ''clicks,'' that are ideally suited for acoustic tracking. If a widely distributed passive acoustic array is available, then the relative time delays from a click's direct path can be used to derive an animal's location. There are many interesting circumstances, however, where deploying a wide-baseline 3D array geometry is either unavailable or impractical. In such situations acoustic reflections of clicks from the ocean surface and bottom can be used to effectively convert a few hydrophones into a wide-aperture array that permits localization in range, depth, and azimuth. In addition acoustic multipath, particularly surface reflections, can also simplify automated detection and tracking procedures, and permit a single hydrophone to be used to distinguish between multiple whales calling simultaneously. Reviewed here is recent research on how multipath effects detected on one or two hydrophones have been used to recover dive profiles of sperm whales in the Gulf of Alaska and the Gulf of Mexico. This information, in turn, has been used to explore whether the animals exploit multipath for their own purposes, to estimate signal directionality, and to observe how sperm whales remove fish from longlines. [Work sponsored by ONR, the Minerals Management Service, and the North Pacific Research Board.]

9:45

4aSP5. Passive detection, data association, and localization of marine mammal calls using traditional 3D hyperbolic tracking algorithms. David Moretti, Susan Jarvis, Nancy DiMarzio, and Ron Morrissey (NUWC, Bldg. 1351, Newport, RI 02841)

The Office of Naval Research Marine Mammal Monitoring on Navy Ranges (M3R) project has developed a toolset for passive detection and localization of marine mammals using the existing infrastructure of Navy undersea ranges. The tools are designed to work across a broad class of calls including clicks, sweeps, and whistles. Using vocalizations received by widely spaced omnidirectional hydrophones, the M3R tools can detect and localize marine mammals in real-time. Signal detection algorithms, data association algorithm, and tracking algorithm have been developed, and demonstrated. The tools have also been successfully fielded to conduct extended monitoring of a 500 square nautical mile area located in the Tongue of the Ocean. This presentation will present the passive acoustic monitoring and localization tools developed under M3R. It will also present results of the application of these tools to the real-time detection and tracking as well as extended monitoring of various toothed whale species at the Atlantic Undersea Test and Evaluation Center (AUTEC), Andros Island, Bahamas.

10:05–10:35 Break

10:35

4aSP6. Perception of multimodal time differences in virtual environments. Durand R. Begault and Elizabeth M. Wenzel (Human Information Processing Res. Branch, NASA Ames Res. Ctr., MS 262-2, Moffett Field, CA 94035, durand.r.begault@nasa.gov)

Presentation of multimodal information with a virtual environment human interface can enable teleoperation of remote tools, sensors, and vehicles for space applications. However, the performance advantages of providing multimodal information can be potentially reduced by intermodal timing asynchronies between different rendering systems (system–system asynchrony). Additionally, asynchronies between the rendering system and proprioceptive, vestibular, and kinesthetic cues that are produced by the actions of the human user can also be problematic (human-system asynchrony). Preliminary data on tolerable asynchrony for audio-haptic interfaces for virtual environments is presented, along with data relevant to perceptually acceptable update rates and latency to minimize computational and perceptual artifacts within virtual auditory displays. [Work supported by the Human Measurement and Performance Project within NASA's Airspace Systems Program.]

10:55

4aSP7. Dynamics of bullfrog choruses studied by time-delay localization with an acoustic array. James A. Simmons (Dept. of Neurosci., Brown Univ., Providence, RI 02912) and Andrea M. Simmons (Brown Univ., Providence, RI 02912)

Understanding the dynamics of bullfrog choruses, particularly the occurrence of acoustic interactions among the calling males, requires segregating and characterizing the sequences of vocalizations from individual frogs making up the chorus night after night for a calling season lasting almost a month. Chorusing takes place in the dark, often at ponds surrounded by vegetation, in conditions where visual observation to locate calling individuals is rarely possible. To record choruses, we used two acoustic sensor units, each consisting of four electret condenser microphones mounted on the faces of a 3-cm cube, located 5–10 m apart and aligned alongside a pond. Signals were stored on an 8-channel digital instrumentation recorder. These signals were used to estimate the horizontal directions and, from these, the locations of multiple bullfrogs (*Rana catesbeiana*) vocalizing at night. The array's mixture of short and long apertures was conceived for acoustic surveying to avoid problems caused by variations in acoustic propagation due to wind and vegetation, and to avoid ambiguities caused by multiple overlapping signals detected only over long apertures. Several kinds of acoustic interactions are evident when calls are assigned to individual animals. [Work supported by DARPA, ONR, NIH.]

Contributed Papers

11:15

4aSP8. Time-delay estimation in underwater acoustics: A global optimization approach. Zoi-Heleni Michalopoulou (Dept. of Mathematical Sci., New Jersey Inst. of Technol., Newark, NJ 07102) and Michele Picarelli (Saint Peter's College, Jersey City, NJ 07306)

Arrival times and amplitudes of individual ray paths provide significant information for source localization and geoacoustic inversion in underwater acoustics. Considered in a statistical framework, estimates for such quantities can be obtained via maximum likelihood or maximum *a posteriori* methods. These approaches, however, require function maximization on a multidimensional space, demanding extensive computations depending on the number of distinct paths. Local search methods have been proposed as computationally low-cost alternatives; such techniques, however, often get trapped in local maxima. In this work, a global optimization approach is proposed for the maximization of the posterior probability distribution of arrival times and amplitudes. To avoid an exhaustive search, an estimate of the posterior distribution is obtained employing Gibbs sampling; it is, subsequently, straightforward to obtain values for time delays and amplitudes for which the distribution attains a maximum. Simulations show that the new method provides excellent estimates that are comparable to those of the analytical maximum *a posteriori* approach. Real data results from the Haro Strait Experiment further validate the method. [Work supported by ONR.]

11:30

4aSP9. Concurrent visual and acoustic tracking of fin whales offshore Southern California. Sean M. Wiggins, Allan W. Sauter, John A. Hildebrand (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0205), Ahmad T. Abawi, Michael B. Porter, Paul Hursky (Sci. Applications Intl. Corp., San Diego, CA), and John Calambokidis (Cascadia Res., Olympia, WA)

Ocean acoustic techniques can be used to monitor whale presence within a region over long time periods. However, using the number of recorded calls to provide an estimate of the number of total whales present has not yet been realized. To help provide a transfer function from the number of acoustic calls to the number of whales present, we conducted a 10-day, concurrent visual and acoustic experiment aboard the FLoating Instrument Platform (FLIP) focusing primarily on fin whale monitoring. FLIP is a stable platform with a visual observation deck approximately 25 m above the sea surface providing excellent visual range. Suspended beneath FLIP at 90 m below the sea surface was a 105-m-long acoustic array consisting of 8 hydrophones, 15 m apart. FLIP was placed in a stationary 3-point mooring centered above three seafloor-mounted hydrophone recorders at depths between 250 and 400 m. The FLIP vertical and seafloor horizontal arrays provide excellent geometry for acoustically tracking whales via arrival-time difference and model-based techniques. These fin whale call tracks are compared to the visually observed fin whale surfacing to provide call-to-whale relationships.

11:45

4aSP10. Location and range of calling blue and fin whales off the Western Antarctic Peninsula. Ana Sirović, John A. Hildebrand, Sean M. Wiggins (Scripps Inst. of Oceanogr., 9500 Gilman Dr., MC 0205, La Jolla, CA 92093-0205), Mark A. McDonald (Whale Acoust., Bellvue, CO 80512), Sue E. Moore (Univ. of Washington, Seattle, WA 98105), and Deborah Thiele (Deakin Univ., Warrnambool VIC 3280, Australia)

Blue and fin whales were heavily whaled from the Southern Ocean during the 20th century and their populations today remain at extremely low levels. Both species produce loud, low-frequency calls that are well suited for long propagation. Eight acoustic recording packages (ARPs) were deployed off the Western Antarctic Peninsula, with instrument spacing from 100 to 180 km. Blue and fin whale calls were commonly heard and two different methods were used to determine locations and ranges to the calling whales. Blue whales were localized using the differences in call arrival times to three different instruments. It also was possible to calculate ranges to blue whales from multipath arrivals of their calls and verify the results obtained by the difference in arrival times method. Short detection ranges for fin whales precluded localization, but multipath arrival difference allowed range estimation. Source levels for both species were calculated using these ranges and the measured received levels.

Session 4aUW

Underwater Acoustics, Signal Processing in Acoustics and Acoustical Oceanography: Long-Range Acoustic Propagation Part I

Peter F. Worcester, Chair

Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0225

Chair's Introduction—8:15

Invited Papers

8:20

4aUW1. Fifteen years of long-range propagation experiments in the North Pacific. Robert C. Spindel (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, spindel@apl.washington.edu) and Peter F. Worcester (Scripps Inst. of Oceanogr., La Jolla, CA 92093)

Over the last 15 years a series of long-range acoustic propagation experiments has been conducted in the North Pacific Ocean using various combinations of low-frequency, wide bandwidth transmitters and horizontal and vertical line array receivers, including a two-dimensional array with a maximum vertical aperture of 1400 m and a horizontal aperture of 3600 m. These measurements were undertaken to further our understanding of the physics of low-frequency, broadband propagation and the effects of environmental variability on signal stability and coherence. The first experiment in this series, which was conducted in 1989, combined a broadband, 250-Hz transmitter with a 3-km-long vertical receiving array 1000 km distant. The transmissions revealed a traditional SOFAR arrival pattern of steep, raylike time fronts in the early portion building to a crescendo of near-axial arrivals that cut off abruptly. The near-axial finale was found to be highly scattered, however, and included significant energy in the geometric shadow zone. The prevailing theory of scattering from internal-wave-induced sound-speed fluctuations predicted far more resolvable arrivals than were observed. These observations stimulated an intense research effort to reconcile data with theory. In this presentation we summarize the central issues these experiments have addressed, and identify the many still remaining unanswered questions.

8:45

4aUW2. Experimental measurements of normal-mode statistics at megameter ranges. Kathleen Wage (ECE Dept., George Mason Univ., 4400 Univ. Dr., MSN 1G5, Fairfax, VA 22030, kwage@gmu.edu)

The Acoustic Thermometry of Ocean Climate, Alternate Source Test, and North Pacific Acoustic Laboratory experiments generated extensive vertical array data sets for studying normal-mode signals at megameter ranges. These data sets include receptions for 28-, 75-, and 84-Hz sources at ranges of 3.5 and 5 Mm. Recent analysis of the low-mode arrivals in these receptions produced estimates of important mode statistics, including the modal time spread, cross-mode coherence, and temporal coherence. This talk will present the experimental results and place them in context, making comparisons to theoretical predictions about long-range mode propagation and to numerical simulations. The implications for future research will be discussed. [Work sponsored by an ONR Ocean Acoustics Entry-Level Faculty Award.]

Contributed Papers

9:10

4aUW3. Analysis of multipath acoustic field variability and coherence in the finale of broadband basin-scale transmissions in the North Pacific Ocean. John Colosi (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

The statistics of low-frequency, long-range acoustic transmissions in the North Pacific Ocean are presented. Broadband signals at center frequencies of 28, 75, and 84 Hz are analyzed at propagation ranges of 3252 to 5171 km, and transmissions were received on 700- and 1400-m-long vertical receiver arrays with 35-m hydrophone spacing. The analysis focuses on the energetic ''finale'' region of the broadband timefront arrival pattern where a multipath interference pattern exists. Two-dimensional (depth and time) phase unwrapping is employed to study separately the complex field phase and intensity.

9:25

4aUW4. Spatial coherence of acoustic signals measured during the 1998–1999 North Pacific Acoustic Laboratory (NPAL) experiment. Vladimir E. Ostashev and Alexander G. Voronovich (NOAA/Environ. Technol. Lab., 325 Broadway, Boulder, CO 80305), and the NPAL Group (J.A. Colosi, B.D. Cornuelle, B.D. Dushaw, M.A. Dzieciuch, B.M. Howe, J. A. Mercer, W.H. Munk, R.C. Spindel, and P.F. Worcester)

Little experimental data have been reported in the literature about spatial coherence of low-frequency sound waves propagating over megameter ranges in the ocean. In 1998–1999, the North Pacific Acoustic Laboratory (NPAL) carried out a comprehensive low-frequency long range sound propagation experiment with the use of the receiving billboard acoustic array. By signal-processing of the NPAL data, we calculated spatial coherence of the broadband acoustic signals (with 75 Hz carrier) after their propagation of about four thousand km in North Pacific. The results ob-

tained are presented in this paper. First, the horizontal coherence of the NPAL signals is discussed. The results show that the horizontal coherence radius varies in the range 500–1000 m. This value is in a good agreement with predictions of a 3D modal theory of low-frequency long-range sound propagation in a fluctuating ocean. Secondly, the vertical coherence of the NPAL signals is presented. After averaging over several days of the experiment, the vertical coherence reveals non-monotonic dependence on the distance between hydrophones. Finally, the horizontal-vertical coherence is studied. This coherence rapidly degrades with the increase of vertical separation between hydrophones. [Work supported by ONR.]

9:40

4aUW5. Transverse horizontal spatial coherence of deep arrivals at megameter ranges. Rex K. Andrew, Bruce M. Howe, and James A. Mercer (Appl. Phys. Lab., 1013 NE 40th St., Seattle, WA 98105)

Predictions of transverse horizontal spatial coherence from path integral theory are tested for ranges from 2000 to 3000 km. A low-frequency (75 Hz) source bottom-mounted at depth 820 m near Kauai transmitted ''m-sequence'' signals over several years to bottom-mounted U.S. Navy horizontal line arrays in the North Pacific. The arrays acquired only the early, deep arrivals. Coherence lengths estimated from the datasets, on the order of 400 m, show good agreement with lengths calculated from theory. Both measurements and predictions show a dependence on path as well as path length. These coherence lengths correspond via a simple incoherent source model to about 1° in horizontal arrival angle variability. Estimates of scintillation index, log-amplitude variance, and variance of decibel intensity indicate that the fields were partially saturated. There is no significant seasonal variability in these measures. Scintillation index predictions from path integral theory also agree well, although theory identifies these fields as fully saturated. [Work supported by ONR.]

9:55–10:10 Break

Invited Paper

10:10

4aUW6. α, β, and wave-field statistics. Michael G. Brown, Francisco J. Beron-Vera, Irina Rypina, and Ilya A. Udovydchenkov (RSMAS-AMP, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

Recent analysis of measurements made in the AET experiment, in which 75 ± 15 Hz signals propagated a distance of 3.25 Mm in the eastern North Pacific Ocean, show that long-held assumptions and theoretical ideas about wave-field statistics at long range require significant revision. Progress towards developing a new theory is discussed in this talk. In contrast to the older theoretical ideas, the background sound-speed structure plays an active and critically important role in the new theory, while details of the perturbation field play only a minor role. The importance of the background sound-speed structure is quantified by the ray-based stability parameter α, or the asymptotically equivalent mode-based waveguide invariant β. [Work supported by ONR and NSF.]

Contributed Papers

10:35

4aUW7. Comparison between ocean-acoustic fluctuations in parabolic-equation simulations and estimates from integral approximations. Stanley M. Flatté (Phys. Dept., Univ. of Calif., Santa Cruz, CA 95064) and Michael D. Vera (Univ. of Southern Mississippi, Long Beach, MS 39560)

Line-integral approximations to the acoustic path integral have been used to estimate fluctuations due to internal waves. Approximations for the root-mean-square (rms) fluctuation and the bias of travel time, rms fluctuation in vertical arrival angle, and the spreading of the acoustic pulse out to 1000-km range are here compared to estimates from simulations that use the parabolic equation (PE). Integral-approximation (IA) estimates of rms travel-time fluctuations were within statistical uncertainty at 1000 km for the Slice89 profile, and in disagreement by between 20% and 60% for the canonical profile. Bias estimates were accurate for the first few hundred kilometers of propagation, but often disagreed beyond. The PE structure functions of travel time with depth were quadratic for vertical separations of 20 m or less, in qualitative agreement with the IA prediction. Pulse spreads in the PE results were much smaller than predicted by the IA estimates. Implications of these results will be discussed.

10:50

4aUW8. Eliminating unnecessary diffractive features in ocean sound-speed models for long-range acoustic propagation. Katherine C. Hegewisch, Nicholas R. Cerruti, and Steven Tomsovic (Dept. of Phys., Washington State Univ., Pullman, WA 99164-2814, khegewisch@wsu.edu)

A major focus of long-range acoustic experiments is the study of the properties of the ocean. In order to explore these properties, it is necessary to have an effective model of the ocean that is sufficient to understand the physics of the wave propagation. Large structures in the ocean cause refraction of the acoustic waves, while smaller structures cause diffraction. Since long-range time fronts are mainly due to refractive effects, adding diffractive features to a model does not significantly change the long-range wave propagation. However, such features cause important theoretical techniques based on refraction, e.g., ray methods, to be inaccurate and they also increase computational effort in the wave propagation calculations. A prescription to remove the diffractive features in an ocean sound-speed model is outlined and the effects on the time fronts is shown. This removal eliminates the ''microfolds'' in the phase space manifolds [discussed by Simmens *et al.* in J. Acoust. Soc. Am. **102**, 239 (1997)] and improves the wave/ray correspondence. Therefore, it is necessary to remove diffractive structures from ocean models before attempting to use ray methods to understand long-range propagation. [Work supported by ONR.]

11:05

4aUW9. Acoustic beam dynamics in deep ocean environments. Francisco J. Beron-Vera, Ilya A. Udovydchenkov, Irina Rypina, and Michael G. Brown (RSMAS-AMP, Univ. of Miami, 4600 Rickenbacker Cswy., Miami FL 33149)

Parabolic equation simulations of broad-band directional beams in highly structured deep ocean environments are used to illustrate the important role played by the background sound speed structure on controlling wavefield structure and stability, and to test theoretical predictions of time spreads. The predictions are based on both ray- and mode-based theoretical considerations, in which the stability parameter α and the waveguide invariant β, respectively, play important roles. [Work supported by ONR and NSF.]

11:20

4aUW10. Mean field of a sound wave propagating in an oceanic waveguide with random inhomogeneities. Alexander G. Voronovich and Vladimir E. Ostashev (NOAA/Environ. Technol. Lab., 325 Broadway, Boulder, CO 80305)

The mean field is an important statistical characteristic of a sound wave propagating in an oceanic waveguide with random inhomogeneities such as internal gravity waves. A closed equation for the mean sound field propagating in a 3D random waveguide can be obtained with the use of the diagram technique. Using the Fourier transform, this equation is reduced to a one-dimensional integro-differential equation with respect to the vertical coordinate. The latter equation determines acoustic modes which are generally different from those in a regular waveguide. For low frequency sound waves, the integro-differential equation is solved approximately. This results in an explicit formula for the mean sound field as a sum of modes which attenuate exponentially with the propagation distance. The extinction coefficient of each mode is linearly related to the spectrum of random inhomogeneities. Therefore, measuring the extinction coefficients for different frequencies and different mode numbers, one can try to reconstruct this spectrum. The mean field of a sound wave propagating in a 2D random waveguide is also calculated. The result obtained is compared with that for a 3D random waveguide. This allows us to study the range of applicability of 2D approximation.

11:35

4aUW11. Interference wave associated with a multitude of cusped caustics in long-range acoustic propagation in the ocean. Natalie S. Grigorieva and Gregory M. Fridman (Dept. of Appl. Math. and Math. Modeling, St. Petersburg State Marine Tech. Univ., 3 Lotsmanskaya Str., St. Petersburg, 190008, Russia, nsgrig@natalie.spb.su)

In long-range propagation experiments, the observed time-of-arrival patterns show early geometrical-like arrivals followed by a crescendo of energy that propagates along the sound-channel axis and is not resolved into individual arrivals. In these experiments the sound source and receiver are placed close to the depth of the axis. As a result, rays emerging from the source at sufficiently small angles intersect the axis many times and form in its vicinity a large number of caustics with caustic cusps located repeatedly along the axis. It is well known that a description in terms of geometrical acoustics is not valid in neighborhoods of cusped caustics because such a description presupposes that the waves associated with individual ray paths do not interfere with one another. These neighborhoods of interference grow with range and at long ranges they overlap. The interference of the wave fields that correspond to near-axial rays results in the formation of a complex interference wave, or axial wave, that propagates along the waveguide axis being a component of the axial crescendo. In this paper for a range-independent ocean the axial wave formation is discussed as well as its main properties in the frequency and time domains. [Work supported by ONR.]

THURSDAY AFTERNOON, 18 NOVEMBER 2004 PACIFIC SALON 2, 1:00 TO 4:45 P.M.

Session 4pAAa

Architectural Acoustics, Noise, Psychological and Physiological Acoustics, Speech Communication and Committee on Standards; Implementation of Classroom Acoustics II

Lily Wang, Chair
Architectural Engineering, University of Nebraska Lincoln, Omaha, Nebraska 68182

Contributed Papers

1:00

4pAAa1. A case study: School band and practice rooms in Harrington Park (NJ) school. Daniel R. Raichel (Eilar Assoc. and CUNY Grad. Ctr., 2727 Moore Ln., Ft. Collins, CO 80526, draichel@comcast.net)

Several years ago, the Harrington Park (NJ) Board of Education constructed a new gym/cafeteria addition and converted the old multipurpose room in the elementary school (K–8) into a band practice room with two adjacent individual practice rooms. In an attempt to allay the sound level of the band, which can reach as high as 102 dB(A), sound panels having an effective absorption coefficient of 0.74 were installed over approximately 20% of the plaster wall surfaces, resulting in little acoustic effect. An analysis indicates that using wall and ceiling panels with a higher values of absorption coefficients over the entire wall surface and the ceiling would bring the sound level down by only 3 dB, insufficient to bring the prevailing SPLs to more reasonable levels, due to the band's sound power. However, the reverberation time of 0.8 s would be reduced to 0.3 s, which would allow the music to be more discernible. The walls and the doors separating the individual practice rooms from the band room proved to be ineffective in isolating these rooms acoustically. It was recommended that modular practice rooms be purchased and installed within these two practice rooms, with ventilation ducts leading outside the music complex.

1:15

4pAAa2. Acoustical design of Kumamoto Health Science University. Masahiro Ikeda (Yamaha Corp. of America, 6600 Orangethorpe Ave., Buena Park, CA 90620, mikeda@yamaha.com)

Kumamoto Health Science University was built in 2002. The key of the project is how to include variety of facilities required for medical school within a limited cost. For this purpose, a large flat circle formed with folded metal plate was adopted. The circle roof covers the rooms, which is located on the ground, except a gymnasium. The prior acoustical concern was the noise caused by rain drops on the metal roof. Through two times experiments carried out in a design and a construction stage, three types of the details were utilized depending on the purposes of the rooms. The effect of inserted rubber, which is cost effective compared to laminated damping steel sheets, is 10 dB in the structure borne transmission loss. Regarding to the room acoustics, the ceilings of lecture rooms are designed to reflect frontal reflection from talkers effectively to the seating area. Acoustical panels are located at a higher wall part instead of the ceiling. Geometrical simulation showed the effect of the acoustical design. The interior of the gymnasium was acoustically designed for utilization of ceremony and assembly, while keeping a liveness suitable for a sports utilization. The highly designed campus was realized in low cost.

1:30

4pAAa3. Speech intelligibility in real and virtual classrooms. Wonyoung Yang and Murray Hodgson (Rm. 313, Univ. of BC, 2206 East Mall, Vancouver, BC V6T 1Z3, Canada)

In current research by the authors, auralization is being used to identify the optimal reverberation times and signal-to-noise level differences for speech in classrooms. The work presented here was initiated to validate the auralization procedure in comparison with live listening tests in real classrooms. Reverberation and noise conditions were created in both real and virtual classrooms, and speech intelligibility tests performed. Two architecturally identical classrooms which have different amounts of sound absorption were selected for real classroom speech intelligibility tests, and their acoustical parameters were measured. Speech and noise sources were placed at the fronts and backs of the classrooms. Speech intelligibility tests were performed at three positions with normal-hearing subjects. The modified rhyme test (MRT) and the noise babble were generated by loudspeakers. The classrooms and speech intelligibility test conditions were then simulated for auralization using CATT-Acoustics, based on the real classroom measurement results. Virtual speech intelligibility tests were performed on the same subjects in the laboratory using headphones. The MRT results in the real and virtual classrooms were compared.

1:45

4pAAa4. Acoustics in a high school library. Sergio Beristain (P.O. Box 12-1022, Narvarte, 03001, Mexico D. F., Mexico, sberista@hotmail.com)

A new and up-to-date library is under development within a Mexican high school educational institution. Usually, libraries are supposed to be quiet places where people can read without disturbance, but nowadays with the new data storage technologies and their capabilities for storing much more than only text, where people can interact with the material under study, for instance, read the description of the words Water Fall, see the water sparkling, and hear the sound of the fall, plus the sound of the computers and proyection equipment, libraries tend to be more and more noisy. The approach taken for this project includes separate areas and acoustic absorption material distribution for different activities within the library, such as reading, computer consultation of texts, and computing consultation of multimedia stored books either in DVD format or by the network.

2:00

4pAAa5. Articulation tests of the first and the second languages in classrooms. Coriolanus C. L. Lam, Kai Ming Li, and Stanley M. L. Cheung (Dept. of Mech. Eng., The Hong Kong Polytechnic Univ., Hung Hom, Hong Kong)

The current study is motivated by the needs of improving acoustic environments in classrooms for the primary and secondary schools in Hong Kong because one of the main education policies is to enable pupils be biliterate and trilingual. Being biliterate means that the pupils are expected to write good English and Chinese. Being trilingual requires that pupils be able to speak English, Cantonese (dialect spoken in Hong Kong) and Mandarin (dialect spoken in China). There are many related studies investigating the speech intelligibility in both English and Mandarin. However, these studies are targeted at the respective native listeners. On the other hand, there are few research studies using Cantonese as the first language of listeners. A series of articulation tests, which were based on the phonetical characteristics of these different languages, was carried out for university students in Hong Kong. Other relevant parameters, such as the speech transmission index, reverberation time and signal-to-noise ratio, were also examined. The outcomes of the present study will provide a basis for further investigations involving pupils of primary and secondary schools. [Work supported by the Research Grants Council of the Hong SAR Government and the Hong Kong Polytechnic University.]

2:15–2:45 Panel Discussion

2:45–4:45 Brainstorming Session

THURSDAY AFTERNOON, 18 NOVEMBER 2004 PACIFIC SALONS 6 & 7, 3:00 P.M. TO 5:05 P.M.

Session 4pAAb

Architectural Acoustics, Speech Communication and Committee on Standards: Speech in Architectural Spaces — Both Intelligibility and Privacy I

Kenneth P. Roy, Cochair
Armstrong World Industries, Innovation Center, 2500 Columbia Avenue, Lancaster, Pennsylvania 17604

Peter A. Mapp, Cochair
Mapp Associates, 5 Worthington Way, Colchester C03 4JZ United Kingdom

Chair's Introduction—3:00

Invited Papers

3:05

4pAAb1. An evaluation of source speakers for open-plan privacy measurements. Kenneth W. Good, Jr. and Kenneth P. Roy, Ph.D. (Armstrong Innovation Ctr., 2500, Columbia Ave, Lancaster, PA 17601)

A variety of source speakers is used today when making field measurements of interzone attenuation in open-plan spaces for the purpose of calculating the privacy index. These speakers have a variety of performance characteristics and often do not meet ASTM E 1179. The question is, do these variations matter more or less for field or laboratory measurements? In this study we look at the speaker performance related to their impact on interzone attenuation in field measurements. An omnidirectional source speaker, a speaker meeting ASTM E 1179, and another typically used speaker are compared.

3:25

4pAAb2. Sound fields and speech intelligibility in rooms. Kenneth P. Roy (Armstrong Innovation Ctr., 2500 Columbia Ave., Lancaster, PA 17604), John Stowell, Mike Walter, and Mathew Schuepbach (Univ. of Kansas, Lawrence, KS 66045)

A key objective in the architectural design of buildings is to choose building elements that provide both speech intelligibility within a space, and speech privacy between spaces. Standard measurement and evaluation methods per ASTM standards on speech intelligibility and privacy will be discussed as they relate to field measurements. Field data on the reverberant sound fall-off with distance from the source will be presented for a range of small to mid-sized rooms. The variation in reverberant sound within a space can result in an estimation error for the AI and PI between rooms. A possible approach to resolving this issue will be discussed.

3:45

4pAAb3. Assessment of Architectural speech security of closed offices and meeting rooms. Bradford N. Gover and John S. Bradley (Inst. for Res. in Contruction, Natl. Res. Council, Otttawa, ON K1A OR6, Canada, brad.gover@nrc-cnr.gc.ca)

The degree to which converstions taking place within closed offices and meeting rooms are audible or intelligible to a listener outside the room depends on the listeners hearing ability, the talkers voice level, the sound attenuation afforded by the building, and the background noise at the position of the listener. Sensible assumptions can be made about the listeners ability so that signal-to-noise measures can be used to reliably indicate audibility or intelligibility. A survey of actual meetings was conducted to acquire estimates of typical voice and noise levels, which can be used in the prediction of security. In existing buildings, the background noise can be measured, leaving only the need to measure sound attenuation through the building. A new measurement prodcedure has been developed that accurately indicates the attenuation from within a source room to a single receiving point outside the room. The results of this procedure can be used to derive a reliable indicator of the level of speech security at the receiving point. This procedure is capable of handling localized weak spots (''leaks'' or ''hot spots'') in the partitions. An overview of the measurement and assessment procedures will be discussed.

4:05

4pAAb4. Speech privacy for healthcare. Joel A. Lewitz (Lewitz and Assoc. Inc., 1505 Bridgeway, Ste. 128, Sausalito, CA 94965)

The Healthcare Insurance Portability and Accountability Act, also known as HIPAA, requires healthcare providers to implement reasonable safeguards to protect the privacy of healthcare information. This includes privacy for oral communications. Architects and facility managers are aware of this mandate, and are calling for these capabilities to be designed into current projects. HIPAA contains no objective, measurable criteria for speech privacy, leaving the design team to select specific criteria for their project. Criteria options will be discussed including Privacy Index and other descriptors that have been developed and correlated with subjective evaluation of privacy. This paper describes several recent projects, which began with evaluations of existing facilities and led to establishment of criteria and design solutions. These studies included various aspects of the healthcare facility such as offices, exam rooms and conference rooms as well as open plan reception, waiting and registration areas. Field conditions, measurement techniques, measurement results, criteria selection and design solutions will be presented.

4:25

4pAAb5. A case study of speech privacy in healthcare and other closed-plan professional spaces. Kenneth W. Good, Jr. (P.O Box 595, Maytown, PA 17550-0595)

Speech privacy is of growing importance in everyday life. Most folks assume that their conversations will be confidential when the door is closed. The need for the measurement and rating of a room's performance is driven by both consumer demand for respect of privacy, and by regulations such as HIPAA and GLBA. This case study will present field measurements of various professional spaces such as healthcare counseling, medical treatment rooms, and pastors' offices and places of counseling in ministry.

4:45

4pAAb6. Application of intelligibility metrics to privacy measurements. Steven McManus (Gold Line/TEF Box 500, W. Redding, CT 06896, steve.tech@bigfoot.com)

The scale of perceived intelligibility runs from the easily understood, down through hard to understand, to private, and finally confidential. When attempting to use STI to measure privacy, it is found that the currently published translations to Articulation Index and therefore Privacy Index do not correlate well with subjective experience at the lower end of the scale. The transition from privacy, where the aim is not to be disturbed, and confidentiality, where the aim is not to be able to eavesdrop, is very rapid. Here, a study is presented that proposes a new mapping from STI to AI and therefore PI, in the low intelligibility range using the accepted threshold of 95% PI as the boundary for confidentiality. This new mapping deviates from the mathematical relationship between STI and AI, which is based solely on signal to noise ratio.

Session 4pAB

Animal Bioacoustics: Acoustic Monitoring of Large Whales II — The Legacy of Paul Thompson

Sue E. Moore, Chair
National Marine Mammal Laboratory, 7600 Sand Point Way, NE, Seattle, Washington 98115

Contributed Papers

1:25

4pAB1. Detection of a unique calling pattern during the respiratory cycle of blue whale pairs. Catherine Berchok (Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804), David Bradley, Thomas Gabrielson (Penn State Univ., State College, PA 16804), and Richard Sears (MICS Inc, 285 rue green, St. Lambert, QC J4P1T3, Canada)

Although blue whales are known for their long (10–20 s) low-frequency (<20 Hz) vocalizations that occur in regularly patterned series, in coastal waters they produce short (1–4 s) higher-frequency (30–150 Hz) unpatterned calls just as often. Paul Thompson was among the group of scientists who first attributed these short calls to blue whales. They found the majority of these calls to be downswept in frequency but noted the presence of others that were not. This paper describes a special patterning of these non-downswept sounds that was detected in the St. Lawrence estuary in both 2000 and 2001. This patterning is variable but normally includes both regularly and irregularly spaced components. In at least a dozen cases, these mixed-pattern bouts were detected in the presence of a pair of blue whales at the surface for their respiratory cycle. Characterization of these bouts and their daily, seasonal, and spatial distributions will be presented and compared with the distributions of the short downsweeps and the long low-frequency calls. [Work supported by the Applied Research Lab (Penn State), NSF, and the American Museum of Natural History.]

1:40

4pAB2. Insights into underwater behavior and vocalizations of blue and fin whales from attachments of acoustic recording tags. John Calambokidis (Cascadia Res. Collective, Olympia, WA 98501), Erin M. Oleson (Scripps Inst. of Oceanogr., La Jolla, CA 92037), William C. Burgess (Greeneridge Sci., Inc., Goleta, CA 93117), Mark A. McDonald (Whale Acoust., Bellvue, CO 80512), and John A. Hildebrand (Scripps Inst. of Oceanogr., La Jolla, CA 92037)

The underwater behavior and vocalizations of blue and fin whales in the northeast Pacific were examined using three types of electronic archival tags attached with suction cups. The bulk of the attachments were of the "Bioacoustic Probe," a commercially available acoustic recording tag developed by Greeneridge Sciences under ONR support. We completed more than 30 deployments of this tag between 2002 and 2004 at multiple locations off California and in the Sea of Cortez, Mexico. Acoustic data recorded by the tags were used to investigate: (1) vocalizations made by the tagged subject whale and nearby animals; (2) anthropogenic noise to which the subject was exposed, including airguns and vessel noise; and (3) the variability of flow noise, which indicated the relative speed of the subject through the water. Results to date have revealed a number of new findings on blue whale vocalizations, including that only a small proportion of blue whales are producing regular calls, that more blue whales may be producing calls irregularly than regularly, and that callers are males. The new data have also clarified the behavioral context in which calling takes place. [Work supported by ONR, SERDP, CNO N45.]

1:55

4pAB3. Techniques for separating interfering blue and fin whale calls. Paul Hursky, Ahmad T. Abawi, Michael B. Porter (Ctr. for Ocean Res., SAIC, San Diego, CA 92121), John A. Hildebrand, Sean M. Wiggins, and Allan W. Sauter (Univ. of Calif. in San Diego, La Jolla, CA 92093)

Recordings of blue and fin whales were made at the Southern California Offshore Range (SCORE) in August 2003. In analyzing this data, previously successful tracking algorithms, based on time differences of arrival (TDOA) measured on geographically distributed hydrophones, were frustrated by the mutual interference between the shorter, broadband, but higher amplitude fin whale calls and the longer duration, narrow-band, blue whale calls. We present a series of progressively more powerful detectors designed to isolate these two types of calls, by zeroing in on specific features of these calls. Because the limited bandwidth of both types of calls presents a challenge to time delay estimation, we also present a comparison of a number of TDOA estimators, including correlation of suitably isolated time-series and spectral features. We apply these techniques to data from different years and geographic sites to assess previous anecdotal evidence of strikingly similar calls within groups of blue whales, despite the presumably wide range of physical characteristics among individuals within each group.

2:10

4pAB4. Vocalizations of blue whales in the Antarctic: Implications for the use of passive acoustics for population studies on baleen whales. Shannon Rankin (Southwest Fisheries Sci. Ctr., NMFS, NOAA, 8604 La Jolla Shores Dr., La Jolla, CA 92037, rankin_shannon@hotmail.com), Don Ljungblad (Ljungblad Assoc., Elk Mountain, WY 82324), Chris Clark (Cornell Univ., Ithaca, NY 14850), and Hidehiro Kato (Natl. Res. Inst. of Far Seas Fisheries, Shimizu, Japan)

Stereotypical calls of baleen whales can be a diagnostic field characteristic for determining species and stocks. Repeated series of these calls may be used for long-term population studies using passive acoustics. Vocalizations were recorded from 15 groups of blue whales (consisting of 42 individuals) during the 2001–02 and 2002–03 International Whaling Commission-Southern Ocean Whale and Ecosystem Research (IWC-SOWER) surveys in Antarctic waters north and east of the Ross Sea. Long-duration vocalizations attributed to Antarctic blue whales included 28-Hz "tones," 28 Hz "downsweeps," and the more complex "3-unit" calls (considered to be song units). The 28-Hz component was stable for all three calls, regardless of signal strength. The 3-unit calls varied in their 2nd and 3rd units and were heard in only 10 out of the 15 groups. There was no evidence of repeated patterns (song phrases) of 3-unit calls. The results of this study suggest that the 28-Hz peak frequency may be used as a diagnostic feature for acoustic identification of Antarctic blue whales in the field. Additional research on the gender, group size, and associated behavior of these animals is necessary to ground-truth long-term acoustic population studies of baleen whales.

2:25

4pAB5. Studies of waveguide propagation effects on frequency properties of sperm whale communication signals. Natalia A. Sidorovskaia (Phys. Dept., Univ. of Louisiana at Lafayette, UL BOX 44210, Lafayette, LA 70504, nsidorovskaia@louisiana.edu), George E. Ioup, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

The Littoral Acoustic Demonstration Center (LADC) conducted a series of passive acoustic experiments for studying ambient noise environment and marine mammal (namely, sperm whales) acoustic phonations in the Northern Gulf of Mexico by long-term autonomous recordings of broadband acoustic signals on a single bottom-moored hydrophone. It was previously reported that extensive spectrogram analysis of sperm whale phonations revealed persistent null pattern in animal spectral signatures that could be associated with either waveguide propagation effects or with a uniqueness of an animal phonation apparatus. The previous modeling addressed the propagation effects on echolocation signals of deep-water foraging animals. It was demonstrated that the null patterns in the analyzed spectrograms could not be explained by the propagation effects for an animal echolocating at foraging depths. This paper presents the results of modeling waveguide propagation effects on spectral properties of sperm whale near-surface communication signals, codas, that could be associated with distinctive sequences of clicks repeatedly present in the LADC recordings. The possibility of the individualization of sperm whale phonations based on the frequency properties of click recordings on a single hydrophone is discussed. [Research supported by ONR.]

2:40

4pAB6. Variation in acoustic detection distances of delphinid whistles using a towed hydrophone array in several geographic areas. Julie N. Oswald (Scripps Inst. of Oceanogr., Univ. of California San Diego, La Jolla, CA 92093, joswald@ucsd.edu), Shannon Rankin, and Jay Barlow (NOAA, La Jolla, CA 92037)

Dolphins were detected using a towed hydrophone array during cetacean abundance surveys in the eastern tropical Pacific (ETP) and in U.S. west coast waters up to 300 nmi offshore (USWC). Cross-correlation algorithms were used to estimate bearing angles to vocalizing dolphins. The convergence of bearing angles as the ship traveled indicated the probable location of dolphins. The study areas were stratified based on sound speed profiles, sea surface temperature, and depth of the thermocline to examine the effect of oceanographic conditions on acoustic detection distances (the greatest distance at which bearing angles can be determined) for delphinid whistles. Significant differences in mean detection distances were found between offshore ETP (2.9 nmi) and USWC waters (1.5 nmi, $p < 0.001$), and between USWC waters and waters off Peru and Baja pooled (2.9 nmi, $p = 0.005$). Ray trace models confirm these empirical differences in acoustic propagation. Results suggest that the effectiveness of acoustic methods during cetacean abundance surveys varies with oceanographic conditions. Regional variability in whistle propagation may partially explain why whistles are seldom or never used by some northern delphinid species but are commonly used by virtually all tropical delphinid species.

2:55

4pAB7. Automatic species identification of odontocete calls in the Southern California Bight. Marie Roch (Dept. of Computer Sci., San Diego State Univ., 5500 Campanile Dr., San Diego, CA 92182-7720), Melissa Soldevilla, and John Hildebrand (The Univ. of California at San Diego, La Jolla, CA 92093)

In this work, a method for the automatic classification of odontocete species from their vocalizations is presented. We consider the vocalizations of three dolphin species which were recorded in a pelagic environment of the southern California Bight over a period of approximately three years. 57B naval sonobuoys sampling at 48 kHz were used to record the calls of common (*Delphinus delphis*), Risso's (*Grampus griseus*) and Pacific white-sided (*Lagenorhynchus obliquidens*) dolphins. In this study, approximately 3 h of calls were selected, with a disproportionate number of common dolphin calls appearing. An emphasis has been made on classifying calls across recording sessions, thus illustrating the system's ability to perform classification in different acoustic environments with a high likelihood of different individuals performing the vocalizations. Cepstral feature vectors are extracted from manually endpointed call data and classified using a Gaussian mixture model system. The performance of the automatic classification compares favorably with human performance.

3:10

4pAB8. Brydes whale (*balaenoptera edeni*) sounds collected from autonomous hydrophones in the eastern tropical Pacific, 1999–2001. Sara L. Heimlich, Sharon L. Nieukirk, David K. Mellinger, Robert Dziak, Haru Matsumoto, and Matthew Fowler (Cooperative Inst. for Marine Resources Studies, Oregon State Univ., Newport, OR 97365)

An array of seven autonomous hydrophones, deployed within 12 N 8S and 95 110W, continuously recorded the frequency band 1–110 Hz from November 1999 to November 2001. Four low frequency sound types were observed, similar to Brydes whale vocalizations previously reported [E. Oleson *et al.*, Mar. Mammal. Sci. **19**, 407–419 (2003)]. Two averaged 2.5-s duration and were commonly high amplitude sounds that exceeded the sampling frequency. Both had a distinct tone averaging 18.6 and 25.4 Hz, respectively, several frequency-modulated higher tones that resembled harmonics but were not, and often an initial short broad band pulse which overlapped the tonal components. Two were lower amplitude sounds that typically began with a single constant-frequency upper tone, averaging 36.6 Hz, 1.1-s duration, and 28.7 Hz, 1.1-s duration, respectively. Sometimes, an overlapping lower tone followed 0.7 s later, lasting 1.6 s and averaging 21.4 and 15.3 Hz, respectively. Occasionally the upper and lower tones alternated two or more times, with decreasing amplitude on each repetition. Differences in seasonal occurrence and geographic distribution amongst the four types were found. These results support the use of passive acoustics as a technique for examining existing and proposed stock definitions.

3:25–3:40 Break

3:40

4pAB9. Acoustic observations of longline depredation by sperm whales. Aaron M. Thode (Marine Physical Lab., Scripps Inst. Oceanogr., UCSD, La Jolla, CA 92093-0238), Janice M. Straley (Univ. of Alaska Southeast Sitka Campus, Sitka, AK 99835), Tory M. O'Connell (Alaska Dept. of Fish and Game, Sitka, AK 99835), and William C. Burgess (Greeneridge Sci., Inc., Goleta, CA 93117)

Sperm whales have learned to take sablefish off longline gear in the Gulf of Alaska. Over the past two years the Southeast Alaska Sperm Whale Avoidance Project (SEASWAP) has collected biopsy and photo-ID data concerning longline depredation, via collaboration with local fishermen in Sitka, AK. In 2004, in collaboration with SEASWAP, sets of autonomous acoustic recorders were attached to the anchor lines of several longline deployments, effectively converting the fishing gear into a vertical acoustic array deployed 200 m beneath the surface. The fishing vessels then left the area, leaving the acoustic instruments behind to monitor the region. Several hours later, typically the next morning, each vessel would return and begin hauling in the gear, a procedure that was also recorded by the instruments. In May 2004 an interaction of two sperm whales with the F/V Cobra was recorded. It was found that the animals started producing sounds within 20 min of the longline recovery and continued being extremely vocally active throughout the recovery. Acoustic multipath was exploited to estimate the range and depth of the whales. [Work sponsored by the North Pacific Research Board.]

3:55

4pAB10. Source levels of free-ranging killer whale (*orcinus orca*) social vocalizations. Val Veirs (Phys. Dept., Colorado College, Colorado Springs, CO 80903)

An array of four hydrophones spaced over 200 m near the shore of San Juan Island, WA, has been used to localize the underwater vocalizations of Southern Resident orca whales. The hydrophones (ITC) have home-built pre-amps and amplifiers and signals are digitized simultaneously by two computers. The system has been calibrated with an underwater speaker (J-9). Automatic call detection is used to activate data logging on one computer and real-time localization on a second that directs a video camera to record the water's surface at the predicted source location. The system requires about 1 s to detect a call, locate it and redirect the video camera. This is the first report of source levels of social vocalizations of free-ranging Southern Resident orcas. To date (July 2004), 160 orca calls have been localized within 200 m of the center of the array. The average acoustic intensity of these calls is 145 dB re 1 μPa@1 m. The range of intensities is from 137 to 157 dB re 1 μPa@1 m. Complete results including intensity angular distributions with respect to the probable direction of whale travel will be presented. [Work supported by 32 undergraduate researchers over 5 years and the Colorado College Physics Department.]

4:10

4pAB11. Calculation of repetition rates of the vocalizations of killer whales. Judith C. Brown (Phys. Dept., Wellesley College, Wellesley, MA 02481 and Media Lab., MIT, Cambridge, MA 02139), Andrea Hodgins-Davis (Wellesley College, Wellesley, MA 02481), and Patrick J. O. Miller (Univ. of St. Andrews, St. Andrews, Fife KY16 9QQ, UK)

A large number of whale sounds recorded from the Captive Killer Whale Population at Marineland of Antibes, France have recently been classified into call types using acoustic input and visual examination of spectral patterns [A. Hodgins-Davis, ''An Analysis of the Vocal Repertoire of the Captive Killer Whale Population at Marineland of Antibes, France,'' thesis, Wellesley College, 2004]. The repetition rate of the pulsed component of an example of each call type has been calculated using algorithms which have proven successful for pitch tracking of musical signals [J. C. Brown and B. Zhang, J. Acoust. Soc. Am. **89**, 2346–2354 (1991), J. C. Brown, J. Acoust. Soc. Am. **92**, 1394–1402 (1992)]. A new method based on pattern recognition of the Fourier transform has been the most successful of these methods as it best circumvents the problem of background pump noise. Results will be presented and compared. [Sound recording supported by grants from WHOI's Ocean Life Institute and the Royal Society.]

4:25

4pAB12. Consistency in received echo characteristics with an echolocating *Tursiops truncatus*. Stuart Ibsen, Whitlow Au, Paul Nachtigal, and Marlee Breese (46-007 Lilipuna Rd., Kaneohe, HI 96744, stui1@aol.com)

In this experiment the dolphin was asked to perform a discrimination task between two simulated targets. The echoes from these targets were created using a phantom echo generator which replicated the actual echo characteristics of solid spherical stainless-steel or brass targets. The system provided full knowledge of each click produced by the dolphin and the corresponding echo received by the dolphin. During the discrimination task the dolphin used a range of clicks that differed widely in peak frequency and bandwidth. However, the echoes that were returned to the dolphin showed significantly less variability in these characteristics. The dolphin in this study carefully maintained consistency in just a portion of the full bandwidth of its clicks, the region that was both within the dolphin's hearing range and had the best reflectivity off the target. The rest of the click bandwidth outside this region showed a higher degree of variability. The dolphin most likely demonstrated less control over these frequency regions because they were outside its hearing range or did not provide the dolphin with nearly as much information.

4:40

4pAB13. The spatial context of free-ranging Hawaiian spinner dolphins producing acoustic signals. Marc O. Lammers, Michiel Schotten, and Whitlow W. L. Au (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734)

An important requirement for understanding the function of animal signals is the ability to determine the context in which they occur. To improve our understanding of how dolphins use acoustic signals in the wild a three-hydrophone towed array was used to investigate the spatial distribution of Hawaiian spinner dolphins (*Stenella longirostris*) relative to each other as they produced whistles, burst pulses and echolocation clicks. Groups of 10 to 60+ animals were recorded while they traveled and socialized in nearshore waters off Oahu, HI. Broadband recordings up to 62.5 kHz were obtained using a PC-based multi-channel system with simultaneous analog-to-digital converters. Signaling animals were localized using time of arrival difference cues on the three channels. Localized whistles produced closely in time typically originated from individuals spaced widely apart (10–100+ m), supporting the hypothesis that whistles play a role in maintaining contact between animals in a dispersed group. Multiple burst pulses, on the other hand, usually came from individuals spaced closer to one another (5–20 m), suggesting they function as a more intimate form of signaling between adjacent individuals. The spacing between echolocating animals was variable.

4:55

4pAB14. Will acoustically enhanced nets reduce odontocete gillnet bycatch because they are better sound reflectors? T. Aran Mooney, Whitlow W. L. Au, and Paul E. Nachtigall (Marine Mammal Res. Program, Hawaii Inst. of Marine Biol., 46-007 Lilipuna Rd., Kaneohe, HI 96744, mooneyt@hawaii.edu)

Tens of thousands of odontocetes are caught annually as bycatch in gillnet fisheries worldwide. Of particular interest are bottlenose dolphins and harbor porpoises, both caught in U.S. waters. One method to potentially reduce this incidental-take is to use nets that are easier for echolocating marine mammals to detect. These acoustically enhanced gillnets are designed to reflect odontocete sonar signals better than regular nylon nets. They have shown to reduce bycatch. Whether or not this is due to increased acoustic reflectivity remains to be determined. We measured the target strength of cod and monkfish gillnets, including those enhanced with barium sulphate, an iron oxide, using dolphin- and porpoise-like echolocation signals. From the target strengths we then predicted the detection distances of the nets by the respective animals. It was determined that at reasonable distances bottlenose dolphins should be able to detect all nets. But because of lower probable source levels, harbor porpoises may not be able to detect any gillnet before it becomes entangled. Stiffness of the nets was also quantified and experimental nets proved to be greater in stiffness than regular nets, thus likely a significant factor in reducing odontocete bycatch.

Session 4pBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Light and Sound Imaging in the Body

Ronald A. Roy, Cochair
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Todd W. Murray, Cochair
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Invited Papers

2:00

4pBB1. Ultrasound-aided high-resolution biophotonic imaging. Lihong V. Wang (Dept. of Biomed. Eng., Texas A&M Univ., College Station, TX 77843-3120, lwang@tamu.edu)

Using non-ionizing electromagnetic and ultrasonic waves, we develop novel biophotonic imaging for early-cancer detection, a grand challenge in cancer research. Unlike ionizing x-ray, non-ionizing electromagnetic waves such as optical waves are safe for biomedical applications and reveal new contrast mechanisms and functional information. For example, our spectroscopic oblique-incidence reflectometry can detect skin cancers based on functional hemoglobin parameters and cell nuclear size accurately. Unfortunately, electromagnetic waves in the non-ionizing spectral region do not penetrate biological tissue in straight paths as x rays do. Consequently, high-resolution tomography based on non-ionizing electromagnetic waves alone is limited to superficial tissue imaging as demonstrated by confocal, two-photon, and optical-coherence microscopic technologies. Ultrasonic imaging, on the contrary, furnishes good imaging resolution but has poor contrast in early-stage tumors and in addition has strong speckle artifacts. We developed ultrasound-mediated imaging modalities by combining electromagnetic and ultrasonic waves synergistically. The hybrid modalities yield speckle-free electromagnetic-contrast at ultrasonic resolution in relatively large biological tissue, where the spatial resolution is scalable with the ultrasonic frequency. In ultrasound-modulated optical tomography, a focused ultrasonic wave encodes diffuse laser light in scattering biological tissue. In photo-acoustic (thermo-acoustic) tomography, a low-energy laser (rf) pulse induces ultrasonic waves in biological tissue due to thermoelastic expansion.

2:25

4pBB2. Acousto−photonic imaging: Modulation of light by ultrasound for medical imaging. Charles DiMarzio (Northeastern Univ., 440 Dana, 360 Huntington Ave., Boston, MA 02115)

Dual-wave technologies hold the promise of new medical imaging techniques. One such combination is acousto−photonic imaging, in which light waves are ''tagged'' by ultrasound waves. The promise is the spectroscopic capability of optical imaging, to measure hemoglobin, with the resolution of ultrasound. A further promise is the ability to measure both optical and acoustic properties in a single measurement. The mechanisms of interaction include modulation of the optical index of refraction and modulation of the distance between optical scatterers, both resulting in phase shifts of the detected light. Because these phase shifts are random, they are difficult to detect. Detectors average over many speckles and thus do not enhance the signals. The use of a photo-refractive crystal permits large detectors to be used, and also provides resolution in the axial direction, leading to the possibility of using the ultrasound to introduce virtual sources of diffusive waves, which can be treated with processing techniques employed in diffusive optical tomography. This talk will provide an overview of the field, including various source and detector configurations, measurement techniques, and suggestions for practical implementation.

2:50

4pBB3. Dynamic holographic imaging of acoustic motion. Ken L. Telschow and Vance A. Deason (Idaho Natl. Eng. and Environ. Lab., Idaho Falls, ID 83415-2209, telsch@inel.gov)

Optical measurement of acoustic motion at a surface is usually accomplished with a scanning point probe. A full-field view acoustic imaging microscope is described that records subnanometer displacement amplitude and phase over a surface without scanning. Images are recorded at video frame rates and heterodyne principles are used to allow operation at any frequency from Hz to GHz. Fourier transformation of the acoustic displacement image provides a mapping of excited mode wavenumbers at any frequency. This method readily measures (a) the acoustic displacement of transducers, (b) acoustic bulk wave reflection from subsurface objects in solids and liquids, (c) resonant ultrasound vibration modes of complex shapes, (d) surface acoustic waves, and (e) plate wave modes in isotropic and anisotropic materials. These and other examples are presented along with discussion and analysis for characterizing material properties, such as physical dimensions, material elastic properties, crystal orientation, electromechanical coupling, and the identification and visualization of defects.

3:15

4pBB4. Schlieren imaging of sound at the turn of the century. Claudio I. Zanelli and Sam M. Howard (Onda Corp., 592 Weddell Dr., Ste 7, Sunnyvale, CA 94089, cz@ondacorp.com)

The mystery of sound fields lies largely in that, despite the complex presence of diffraction effects, at a human scale, they cannot be seen. We sample an acoustic field two points at a time, and mapping it completely, even in two dimensions, requires a constant field and a lot of patience. Vision allows capturing two-dimensional images, and even mental three-dimensional images. Yet light can only be seen when it is interfered with, and scattered into, our retinas. Like sound, we must destroy its field in order to sense it. Quite possibly, the lack of sensory perception of acoustic field maps is responsible for our slowness in understanding the nature of waves. The limitations of point-to-point mapping of fields leads many to take a few samples and then work directly with mathematical models. Schlieren is an ancient yet rediscovered, fascinating technique to generate instantaneous maps of acoustic fields based on the interaction of light and sound. Its value as a guide to intuitively connect with the underlying physics is undeniable, and new developments have turned it into a quantitative method, even where no other mapping methods are possible.

3:40

4pBB5. A novel approach to high-frequency laser-based acoustic microscopy. Todd W. Murray and Oluwaseyi Balogun (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

A laser-based acoustic microscopy system has been developed which uses a modulated continuous wave (cw) laser source for acoustic wave generation. An amplified electro-absorption modulated diode laser source is used for narrow-band excitation of acoustic waves generation at frequencies up to 200 MHz. The resulting surface displacement is measured using a stabilized Michelson interferometer fed into an rf lock-in amplifier. The detection bandwidth reduction afforded by this technique allows for a significant improvement in signal-to-noise ratio (SNR) over systems using pulsed laser excitation, and displacement sensitivity in the femtometer range is obtained with only minimal surface heating. The source modulation frequency is scanned over the bandwidth of interest and the transient response of the specimen reconstructed from the frequency domain data. In addition, the use of cw excitation allows for resonant excitation of micro- and nanoscale structures for materials characterization or sensor applications. Experimental results demonstrate that the system is well suited for thin-film characterization, and for the nondestructive analysis of nanometer scale electromechanical devices.

Contributed Papers

4:05

4pBB6. Further investigation of acousto-optical imaging in highly diffusive media using a photorefractive crystal based detection system: Sensitivity and contrast. Lei Sui, Emmanuel Bossy, Ronald A. Roy, and Todd W. Murray (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215, suilei@bu.edu)

Acousto-optical imaging takes advantage of good optical contrast and high ultrasonic resolution to facilitate imaging and characterization of biological tissues. By detecting ultrasound modulated optical signals, the local opto-mechanical properties can be extracted. Among the state-of-the-art detection techniques, there is a trade-off between the axial resolution (or ultrasound bandwidth) and signal-to-noise ratio. Recently, a photorefractive crystal (PRC) based system was shown to enhance the detection of ultrasound-modulated optical signals [Sui *et al.*, Proc. SPIE **5320**, 164–171 (2004)]. This system achieved improved axial resolution by using short-duration acoustic pulses [Sui *et al.*, J. Acoust. Soc. Am. **115**, 2523 (2004)]. Previous experiments were done in tissue-mimicking phantoms whose optical diffusivities were less than that of most biological tissues. In this paper, the PRC-based detection system is applied to acousto-optical imaging in highly diffuse media and in biological tissues *in vitro*. Both transmission and right-angle detection geometries are studied. The relationship between the amplitude of the detected signals and acoustic driving pressure is determined. The ability of this system to detect both optical absorption contrast and optical scattering contrast is explored. [Work supported by the Center for Subsurface Sensing and Imaging Systems via NSF ERC Award No. EEC-9986821.]

4:20

4pBB7. Combining acousto-optical imaging with diagnostic ultrasound: resolution, contrast and speed. Emmanuel Bossy, Lei Sui, Todd W. Murray, and Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215)

Ultrasound modulation of light can be used to image optical properties in highly diffusive media such as biological tissues. In a previous study [Bossy *et al.*, J. Acoust. Soc. Am. **115**, 2523 (2004)], a commercial ultrasound scanner was combined with a photorefractive crystal (PRC)-based optical detection scheme to generate simultaneous images of acoustic and optical properties of inclusions embedded in diffusive tissue phantoms. In this paper, we quantitatively investigate the performance of the system in terms of resolution, contrast, and acquisition time. Measurements are performed on gel-based highly diffusive tissue phantoms, with embedded targets possessing optical contrast relative to the surrounding medium. The resolution of the acousto-optical images turns out to be given by the ultrasound beamwidth for the lateral resolution, and by the pulse duration for the axial resolution. Phantoms and targets with different optical contrasts are imaged to assess the feasibility using this technique to delineate tissue types *in vivo*. The time needed to acquire acousto-optical images depends on the desired quality of the image (signal-to-noise ratio) and ranges from 1 to 30 s per line. [Work supported by the Center for Subsurface Sensing and Imaging Systems via NSF ERC Award Number EEC-9986821.]

Session 4pEA

Engineering Acoustics: Microphones and Loudspeakers

Gilles A. Daigle, Chair
Institute of Microstructural Science, National Research Council, Ottawa, Ontario K1A 0R6, Canada

Chair's Introduction—1:25

Contributed Papers

1:30

4pEA1. Precision calibration of measurement microphones. Gunnar Rasmussen (G.R.A.S. Sound & Vib. A/S, Staktoften 22D, DK-2950 Vedbaek, Denmark, gr@gras.dk)

The calibration of measurement microphones is often carried out as a rough check using a comparison between the measurement microphone and another microphone. A calibration with lower uncertainty involves either a complete reciprocity calibration which is hard to carry out in the field or the use of a pistonphone, where the calibration is related to well known physical quantities easy to measure. The largest contribution to the uncertainty is often the barometric pressure and the use of less accurate barometers. The use of a precision barometer built into the pistonphone greatly reduces the uncertainty due to human error and due to varying atmospheric conditions independent of altitude. Normally the ambient temperature at which the calibration takes place is also reported. Practical experience shows that it may be more relevant to report the actual temperature of the transducer under calibration. The final working temperature for a microphone is reached several minutes after the start of the calibration. It is therefore practical to measure and report the temperature of the actual calibration coupler and microphone combinations in order to minimize the uncertainty. Direct data transfer of all relevant data calibration level, barometric pressure, and temperature to a data system to ensure immunity to human error is also practical.

1:45

4pEA2. Design and performance of a micro-probe microphone. Gilles A. Daigle and Michael R. Stinson (Natl. Res. Council, Ottawa K1A 0R6, Canada)

It is often necessary to measure sound fields in confined spaces where minimum disturbance of the sound field is important. The B&K 4170 is a horn-coupled probe microphone for which a 1.25-mm-i.d. probe is attached to the narrow end of the horn and a 1/2-inch microphone is connected to the wide end. In applications where the confinement of the space is extreme, this probe is too large. The probe section of the B&K 4170 has been redesigned using smaller diameter probes. Simulations were made for probes having outside diameters ranging from 1.25 mm down to 0.2 mm. It was found that acceptable sensitivities could be expected from a probe with 0.2 mm o.d. (0.1 mm i.d.) if the length is less than 1 cm. The micro-probe is connected to a larger cylindrical tube with the correct o.d. for mounting on the B&K 4170 casing. Viscous resistance due to the small diameter provides an almost flat response. Several micro-probes were constructed and their performance measured and compared to the simulated results.

2:00

4pEA3. Harsh environment windscreen analysis and design. Jeffrey Hill, Jonathon Blotter, and Timothy Leishman (Brigham Young Univ., 435Q CTB Provo, UT 84602, jblotter@byu.edu)

This paper presents the analysis and design for a multimicrophone probe placed outdoors in relatively harsh environments including high temperature, high humidity, rain, wind, and soil particles. The windscreen was designed to minimize the insertion loss as well as maintain the phase relationship in a circular 5-probe microphone array. A numerical CFD model, which accounts for wind, soil particles, the windscreen material, and transient pressure fluctuations, will be presented. The parametric design using this numerical model will be presented and discussed, along with experimental results. The experimental testing consists of measuring the wind attenuation, the insertion loss, as well as any errors in the phase relationship. Testing is done in an anechoic chamber using two methods: a stationary microphone array, and attaching the microphone system to a rotating arm to simulate the wind. Measurement results which correlate the density and thickness of the foam with the insertion loss and the drop in wind noise level caused by the foam will be presented. Also, the insertion loss due to different weather-proof coverings is reported.

2:15

4pEA4. Reduction of airflow noise in telephone handsets. Michael R. Stinson, Gilles A. Daigle, and John F. Quaroni (Inst. for Microstructural Sci., Natl. Res. Council, Ottawa, ON K1A 0R6, Canada, mike.stinson@nrc-cnrc.gc.ca)

The bursts of airflow that accompany plosive sounds such as a ''p'' or ''t'' can cause significant noise in a telephone handset when the talker's mouth is close to the mouthpiece of the handset. Measurements of the generated noise have been made on a series of modified handsets, using various arrangements of ports and tubes between microphone and the mouthpiece exterior. The use of one, two and more exterior holes, and the introduction of different geometrical coupling between these holes and the interior microphone were explored. The noise could be reduced by over 20 dB by using an array of small (less than 0.5 mm diameter) exterior holes coupled with an offset to a sound tube leading to the microphone. [Work supported in part by Mitel Networks.]

2:30

4pEA5. Acoustical nonlinearities in compression chambers and horns of horn drivers. Alexander Voishvillo (JBL Professional, 8400 Balboa Blvd., Northridge, CA 91329)

A horn driver is an essential part of sound reinforcement systems. Principles of horn driver operation are associated with generation of acoustical nonlinear distortion. Distortion is generated in a compression chamber by nonlinear compression, and by modulation of air stiffness, mass, and viscous losses, and a nonlinear relationship between particle

velocity and sound pressure. Distortion is also generated in the phasing plug and horn by propagation of high amplitude waves. The effect of compression chamber air mass and viscous losses modulation and nonlinearity has not been researched previously. The analysis of modulation and nonlinear effects in the compression chamber is based on analysis of variation of the compression chamber's air moving mass, viscous losses, and compliance as a function of the diaphragm's instantaneous position and level of acoustical pressure. The pressure and diaphragm position-dependent parameters are incorporated into the system of nonlinear differential equations governing movement of the diaphragm and oscillation of air in compression chamber. Analysis of nonlinear propagation effects in the phasing plug and horn is based on numerical solution of an implicit nonlinear equation for propagation of finite amplitude sound waves. The work results in the methodology of simulation dynamic performance of horn drivers and provides comparative analysis of different acoustical nonlinear and parametric mechanisms.

2:45

4pEA6. Additional modeling details for the suspension of a moving-coil loudspeaker. Timothy Leishman and Gordon Dix (Acoust. Res. Group, Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602, tim_leishman@byu.edu)

The suspension of a typical moving-coil loudspeaker incorporates both a surround and a spider of dissimilar construction. Nevertheless, classical lumped-element models simplify this composite system by representing it with a single mechanical compliance and a single mechanical resistance. This paper presents a more detailed lumped-element model of the suspension to account for individual contributions of the surround and spider, their effective masses, and their primary resonances. It subsequently proposes an expanded equivalent circuit for a loudspeaker and discusses experimental methods to characterize new circuit elements.

THURSDAY AFTERNOON, 18 NOVEMBER 2004 PACIFIC SALON 3, 1:00 TO 4:30 P.M.

Session 4pMUa

Musical Acoustics: Musical Instruments of the Asia-Pacific Region

James P. Cottingham, Chair
Department of Physics, Coe College, Cedar Rapids, Iowa 52402

Chair's Introduction—1:00

Invited Papers

1:05

4pMUa1. The acoustics of the khaen, bawu, and gourd pipe. James P. Cottingham (Phys. Dept., Coe College, Cedar Rapids, IA 52402, jcotting@coe.edu)

Mouth-blown instruments using a free reed coupled to a pipe resonator have a long history in China, Japan, and throughout Southeast Asia. The khaen and the gourd pipe are unusual wind instruments, employing a free reed mounted in the side of a cylindrical tube open at both ends, although in the gourd pipe the player has the option of closing the bottom of the pipe so that the reed is near the closed end. The bawu is a closed cylindrical pipe with a free reed at one end, in which the effective acoustical length is varied by the use of tone holes. The playing frequency of each pipe of the khaen or the gourd pipe is typically slightly above both the resonant frequency of the pipe and the natural frequency of the reed. In the bawu, on the other hand, the pipe resonance pulls the sounding frequency well above the natural reed frequency, resulting in a striking difference in tone quality. Acoustical measurements made on these instruments include studies of reed vibration and impedance measurements of the pipes, with particular attention to the coupling of the reed vibration with the pipe resonator.

1:25

4pMUa2. Expressive modeling of Chinese folk wind instruments. Lydia Ayers and Andrew Horner (Dept. of Computer Sci., Hong Kong Univ. of Sci. and Technol., Clear Water Bay, Kowloon, Hong Kong, layers@cs.ust.hk, horner@cs.ust.hk)

Chinese folk wind instruments, including the *paixiao, xun, koudi, bawu, hulusi*, and *lusheng*, have distinctive timbres but are less well known than the other Chinese wind instruments. The *paixiao* (a panpipe) and the *xun* (ocarina) are ancient instruments. The *koudi* is a small bamboo whistle/flute, which can play exciting birdlike melodies. The *bawu, hulusi*, and *lusheng* have different shapes and playing techniques but have similar embedded metal plate reeds that produce their sounds. This talk will give brief descriptions of the musical and spectral characteristics of these instruments, and then describe the additive synthesis design template used to capture their subtle timbral characteristics using score data in Csound. In addition, the design is expressive and successfully uses spectral interpolation for the idiomatic ornamental phrasing for each instrument. We use a special amplitude/frequency function modulation method to simulate trills, which are also tied to varying numbers of grace notes, as tested on traditional Chinese scores. [Work supported by RGC Grant No. HKUST 6020/02H, with special thanks to Chan Chun Nin for recording instrument samples and consultation on playing the instruments, and to Alfred Wong and Jenny Lim for preparing examples.]

1:45

4pMUa3. The timbre space of the Chinese membrane flute (dizi): Physical basis and psychoacoustical effects. Chen-Gia Tsai (Grad. School of Folk Culture and Arts, Taipei Natl. Univ. of Arts, 1 Hsuen-Yuan Rd., Peitou, Taipei 112, Taiwan, tsai_chen_gia@yahoo.com.tw)

The Chinese membrane flute (dizi) has a membrane covering a hole in the wall of the instrument between the mouth hole and the uppermost finger-hole. The membrane is driven by the acoustic pressure in the pipe, producing rich harmonics through its nonlinearity. The dizi timbre is multidimensional and varies with pitch. Inhomogeneity of the dizi timbre is well known by musicians. It reflects a general aesthetic tendency in traditional Chinese music that a skillful musician should produce a variety of timbre with a solo instrument. The major spectral features of dizi tones, i.e., (1) formants, (2) the predominance of odd harmonics, and (3) harmonic-to-subharmonics ratio at high frequencies, can be explained by a Duffing oscillator model of the membrane. They constitute the three dimensions of the dizi timbre space, where the ''trajectory'' of the changing timbre of a dizi melody is embedded. Psychoacoustical effects associated with them are discussed, including (1) spatial effects due to abrupt brightness changes, (2) multi-pitch effect due to the predominance of upper odd harmonics, and (3) roughness induced by subharmonics.

2:05

4pMUa4. Investigation on tonal and playability qualities of didgeridoos: Correlations with physical properties. Rene E. Causse (Ircam CNRS UMR 9912, 1 place Igor Stravnsky, 75004 Paris, France)

The quality of an instrument, perceived by a player, is usually separated into two fields, tonal and given possibilities by the instrument in terms of playability. For the didgeridoo, tonal quality is first linked to the spectral content of the emitted notes, a result of the oscillation of a complex system composed of the instrument, the lips and the player's vocal tract. Is also related to the character of spatialization of the sounds which one sometimes associates with the semantic attribute of sound width or sound envelopment. The playability is attached to the playing facility but also to the various aptitudes of the instrument. Among those one finds the aptitude for the modulations, for the rhythmic effects, for the overblowing, for vocalization, the cries, etc. The air column resistance, the sound efficiency, and the sound dynamics are also linked to the playability. The study is carried out on a series of eight instruments chosen in a corpus of around 50 instruments. The study includes tests with instrumentalits, acoustical and geometrical measurements and analysis of the emitted sounds. Some correlations between the physical properties of the same instrument, the nature of the emitted spectrum and the quality will be presented.

2:25

4pMUa5. Conch-shell and bamboo flute: Spiritual and musical expressions of acoustics in Vedic Hinduism. M. G. Prasad (Mech. Eng. Dept., Stevens Inst. of Technol., Hoboken, NJ 07030, mprasad@stevens.edu)

In Vedic Hinduism rooted in India, the field of acoustics plays an important role. The multi-faceted expressions of acoustics include Vedic mantras (chants), music, yoga, philosophical aspects of language, spiritual aspects of sound, etc. A large number of musical instruments such as woodwind, percussion, string, etc. are used in Hindu culture and religion. Conch-shell and bamboo flute are among the earliest musical instruments. A conch-shell is used to evoke the spiritual effects in the environment during worship rituals at temples and homes. It has a high quality factor. The music played through a transverse bamboo flute is very melodious. It is a major instrument in orchestra music and dance performance. Concerts with bamboo flute as the lead instrument are also very common. This paper deals with spectral characteristics and tonal effects of these two important instruments. Also, the presentation includes a discussion of the importance of acoustics in life as viewed in Vedic Hinduism.

2:45–2:55 Break

2:55

4pMUa6. Model experiments on the "sawari" mechanism: Vibrational behavior of a string with a moving boundary. Akiko Fujise and Shigeru Yoshikawa (Dept. of Acoust. Design, Grad. School of Kyushu Univ., Fukuoka, 815-8540, Japan)

The ''*sawari*'' mechanism, seen in some plucked string instruments like Japanese *biwa* and *shamisen*, produces the vibration of the string with a moving boundary and results in its unique sustaining tone, whose elaborate tuning is essential in Japanese traditional music. Since a slight change of the *sawari* surface's shape strongly affects the tone quality, it is important to examine the relationship between the boundary shape and the consequent vibration in reality. In this presentation, some results of model experiments on a string with the *sawari* mechanism are reported and discussed. By using a high-speed digital video camera and a piezo-electric pickup, the overall string motion and the local interaction between the string and the boundary surface are experimentally measured under several *sawari* conditions. The results are analyzed by applying image processing to video data and also time-frequency analysis to acceleration signals, then compared between basic shapes (plain and curved), widths, and shaven depths of the constraint.

3:15

4pMUa7. Acoustics of Chinese, Japanese, and Korean bells. Thomas D. Rossing (Phys. Dept., Northern Illinois Univ., DeKalb, IL 60115, rossing@physics.niu.edu)

The earliest known Chinese bells were small clapper bells of clay and copper from the third millennium BC. Beginning with the early Bronze Age (ca. 2000 BC), however, bronze became the material of choice. We discuss early Chinese two-tone bells including nao, yongzhong, niuzhong, and bo and illustrate their vibrational behavior using holographic interferometry. We discuss the effect of various physical parameters on the spacing between the two tones. We compare temple bells from China, Japan, and Korea, including such famous bells as the the Yong-le bell in Beijing and the King Songdok bell in Korea.

3:35

4pMUa8. Vibrational modes of pyen-gyoung (Korean chime stone) and pyeon-jong (Korean chime bell). Junehee Yoo (Dept. of Phys. Education, Seoul Natl. Univ., Seoul 151-742, Republic of Korea) and Thomas D. Rossing (Northern Illinois Univ., DeKalb, IL 60115)

Pyeon-gyoung, chime stone, and Pyeon-jong, chime bell, have been allocated as a set of instruments in Korean traditional court music and Confucian ceremony. Pyeon-gyoung is a set of 16 L-shaped chime stones covering one and one-third octaves and pyeon-joung is a set of 16 almond-shaped bronze chime bells, one octave lower than pyeon-gyoung. The nominal frequencies increase with the thickness of the stones and bells. The vibrational mode frequencies in a pyeon-gyoung and pyeon-jong have been studied. The frequency ratios of the modes in stone chimes are smaller than those of marimbas. The modal shapes of the lowest and highest stones, mapped by TV holography, by scanning with an accelerometer and animated by STAR, are shown.

3:55

4pMUa9. Acoustic matching of the bars and resonators of the Indonesian gender. Paul Wheeler (Utah State Univ., 2140 Old Main Hill, Logan, UT 84322)

The gender is a 14-bar metallophone with bamboo resonators used in the Indonesian gamelan. The 14 bronze bars (with stiffening ridges on top of the bars) are carefully tuned to one of three pentatonic scales. The bars are strung together and spaced above 14 bamboo resonator tubes housed in a wooden frame. The bamboo tubes are roughly the same length and diameter, however the knot in the bamboo pipe is in different locations, creating closed pipes of different lengths. Fine tuning of the tubes is achieved by placing rocks or sand in the bottom of the tubes. This paper describes an acoustic study of the gender barung. Tuning of the bars across the octaves was within 15 cents (with one exception), showing great care in the construction and tuning of the bars. Tuning of the resonator tubes, however, was much more divergent, showing less consideration for the actual frequencies of the resonators. Tuning of the resonators to match the bars is left for the player by adding materials to the resonator tube.

Contributed Paper

4:15

4pMUa10. Additional modes in transients of a Balinese Gender dasa plate. Bader Rolf (Musikwissenschaftliches Institut, Neue Rabenstr. 13, 20354 Hamburg, Germany, r_bader@t-online.de)

The initial transient of a Balinese bronze plate of a Gender dasa (metallophone) struck by the instrument's hammer is examined in detail. The spectrum of the radiated sound as well as that of measurements by an accelerator consist of many more eigenmodes compared to the theoretical values for the bending, longitudinal, axial and torsional wave modes. It shows up that the trapezoid shape of the plate leads to these additional frequencies in the transient spectrum, which disappear in the quasi-steady state. As to each bending mode, there exist at least two additional modes. There seems to be a scattering of the initial impulse at the geometrically unsteady boundaries of the trapezoid, which define new internal borders, where new modes can arise from. The low energy of these modes results in a fast decay, while the fundamental mode is still sustained by the bamboo resonator of the Gender. Computer models show the reasoning for such additional spectra modes in principle using transient impulse simulations. These modes can occur just in the case of a hard struck as the typical playing mode of Bali music, which is known to be more lively than Japanese music and corresponds to the initial transient found here.

THURSDAY AFTERNOON, 18 NOVEMBER 2004 SAN DIEGO ROOM, 5:00 TO 6:00 P.M.

Session 4pMUb

Musical Acoustics: Concert by Christopher Adler: New and Traditional Music for Khaen, the Free-Reed Mouth Organ of Laos and Northeast Thailand

James P. Cottingham, Chair
Department of Physics, Coe College, Cedar Rapids, Iowa 52402

Chair's Introduction—5:00

The *khaen* is a free-reed bamboo mouth organ prominent among peoples of Lao ethnicity in Laos and northeast Thailand, and likely an ancestor to more well-known East Asian mouth organs as well as to Western free-reed instruments. Dr. Christopher Adler has been researching and performing the traditional music of the *khaen* for a decade and composing and performing new compositions for *khaen* almost as long. This performance will include five traditional improvisations in each of the five melodic modes available on the *khaen* as well as two original compositions by Dr. Adler which make extensive use of non-traditional techniques. The performance will also include demonstrations of two single-pipe free-reed instruments, the *saneng keo* and the *pii phuthai*, and a duo performance of a classical Thai composition with Supeena Insee Adler playing the three-string Thai zither *jakay*.

Christopher Adler is a composer, improviser and performer living in San Diego, California. In addition to working with the traditional musics of Thailand and Laos, his compositions draw upon his background in mathematics and computer modeling. He received Ph.D. and Master's degrees in composition from Duke University and Bachelor's degrees in music composition and in mathematics from the Massachusetts Institute of Technology. He is currently Director of the Music Program and an Assistant Professor at the Unviersity of San Diego. His work may be heard on 9 Winds Records, Artship Recordings, and WGBH's Art of the States, and three of his compositions for *khaen* may be heard on *Epilogue for a Dark Day*, a newly released solo CD on Tzadik Records. (www.christopheradler.com).

Session 4pNS

Noise: Test Chambers, Sound Sources, Noise Reduction

Brian Anderson, Chair
Graduate Program in Acoustics, Pennsylvania State University, State College, Pennsylvania 16802

Contributed Papers

1:30

4pNS1. Development and measurement of a variable acoustic test chamber. Colin Barnhill, Robert Ng, James E. West, and Ilene J. Busch-Vishniac (Elec. and Computer Eng. and Mech. Eng. Depts., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218)

We have designed and constructed an acoustic test facility capable of varying from largely anechoic to largely reverberant. The chamber measures 18.25×13.5×8 ft., so it has a total volume of 1971 ft^3 and a total surface area of 1000.75 ft^2. The walls are lined with removable fiberglass panels covered in fabric; a hard ceiling has a removable drop ceiling of tectum panels located below it; and the linoleum floor on slab is covered by removable thick carpeting. The reverberation time in the chamber can be varied by over an order of magnitude (i.e., from 150 ms to 1.5+ s) for frequencies above 200 Hz as measured by traditional means and with a special directional microphone system. All measurement tools excluding sound sources and microphones are located outside of the room so as not to influence the sound characteristics.

1:45

4pNS2. Towards aural opacity—across a visually transparent medium. Michael Ermann, Matt Lutz, Chris Jackson, Brooke Karius (Virginia Tech School of Architecture + Design, 201 Cowgill Hall (0205), Blacksburg, VA 24061, mermann@vt.edu), David Kilper, Indhava Kunjara Na Ayudhya, Thomas Pettin, Tom Pritts, Damian Seitz, Rebecca Stuecker, and Joe McCoy (Virginia Tech School of Architecture + Design)

Virginia Tech students undertook a study and participated in a design competition to provide the campus physical plant building with a wall that allows for natural light and views while maintaining a suitable aural barrier. The wall, which replaces a large garage door, will be constructed as part of an effort to renovate a wood/metal shop into an office. It must provide for adequate transmission loss to account for the noise generated by adjacent utility trucks, shop activities and an *ad hoc* basketball court. Unable to obtain published data for glass vestibules, students measured the transmission loss across another campus building's glass entry vestibule. Currently there are plans for the winning students design entry to be constructed as part of the renovation.

2:00

4pNS3. S-shaped glass also stands for soundless. Sound insulation measurements and implication for building practice. Beatriz M. Pinto (AFassociados, Projectos de Engenharia S.A., Dept. of Acoust., Cais do Lugan 224 P-4400-492 V. N. Gaia, Portugal, beatriz@beatrizpinto.com), Laurentius C. J. van Luxemburg (DHV-Dorsserblesgraaf, 5616 vb Eindhoven, Netherlands), Antonio P. O. Carvalho (U. of Porto, P-4200-465 Porto, Portugal), and Heiko J. Martin (Eindhoven U. of Technol., 5600 mb, The Netherlands)

This paper brings to discussion the sound insulation of a special type of glazing by studying a S(inusoidal)-shaped curved glass. This is been used in a concert hall (under construction in Portugal) with a forward-thinking Rem Koolhaas's architecture requiring nonstandard solutions: glazing with very large dimensions (22 m by 15 m) and each ''S-glass panel'' with a wavelength of about 1.0 m and amplitude of about 0.35 m will be applied as a double facade. The restrictions in ISO 104-3 laboratory measurements (standard test opening for glazing 1.25 by 1.50 m and niche depth of 0.45 m) were not appropriate and not completely fulfilled due to differences in dimensions between small test sample, and the real required use on site as well as the silicone joints that connect the glass elements. For these reasons, the standard test opening of 10 m^2 was used. These aspects are described and analyzed. The tests results (showing satisfactory sound insulation) were essential to understand the effect of the shape and the efficiency of the glass. This study aspires to be helpful to building designers who can, in this way, lose dependency on production process constraints and bring into being all kinds of shapes.

2:15

4pNS4. Visualizing automobile squeals and their vibration modes. Manmohan Moondra and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., Detroit, MI 48202)

Automobile brake squeals have always been a major problem for the auto industry because of the high pitch and intensity level. Currently, the vehicle brake squeal diagnostics is carried out by using a scanning laser synchronized with squeals. If squeals occur at multiple frequencies, then filtering must be used so that the scanned out-of-plane vibration corresponds to a single frequency. This process is very time consuming, especially when squeals contain multiple frequencies. Moreover, the results thus obtained are valid at measurement locations only. In this paper, we use Helmholtz equation least-squares (HELS)-based near-field acoustical holography to reconstruct squeal sounds via a custom-designed 48-microphone array conformal to the shape of a brake assembly. This microphone array allows for taking a snapshot of 3D acoustic image on the surface of the brake assembly and produces 3D mappings of all acoustic quantities. In particular, we can identify the out-of-plane vibration modes that are responsible for squeals and reveal the locations where squeals originate from. There is no need for synchronization and filtering. The whole measurement can be done in seconds, and the reconstructed results are valid not only on source surface, but also in the field. [Work supported by NSF.]

2:30

4pNS5. Sound, as we see it—A better way to diagnose noise sources. Ravinder Beniwal (221 Lewiston Rd., Grosse Pointe Farms, MI 48236) and Sean Wu (Wayne State Univ., Detroit, MI 48202)

In the noise control community, it is often required to diagnose noise sources and their transmission paths. Usually, noise diagnosis consists of taking sound pressure, power, and intensity measurements around suspected sources to learn noise characteristics. If these are not enough to pinpoint the sources, then source ranking and identification of contributions from individual components are carried out. Naturally, success in this type of diagnostics requires many years of practical experiences. Even so, these tests are time consuming and *ad hoc* in nature, and the results thus obtained are valid at measurement locations only. In this paper, a new approach to noise diagnostics is presented, which can significantly reduce the time and complexities involved in diagnostics. This approach is based on the Helmholtz equation least-squares method to visualize the entire sound field generated by a sound source. It allows for 3D conformal measurements of the acoustic pressure and produces an accurate mapping of

the acoustic field on 3D surfaces and surrounding medium. These results enable one to trace sound back to its source and separate airborne from structure-borne sounds. The process can be completed in minutes and the acoustic images thus obtained speak for themselves. [Work supported by NSF.]

2:45

4pNS6. New development of two-dimensional sound index for the design of brand sound in the passenger cars. Sang-Kwon Lee, Byung-Og Jo (Dept. of Mech. Eng., Inha Univ. 253, Young Dong, Nam Gu, Inchon, Korea), and Byung-Soo Kim (Hyundai Motor Co., Whasung City, Korea)

In automotive engineering, the production sound quality becomes an important design requirement. Especially, the brand sound is one of important advantage strategy in a car company. In the paper, the subjective evaluation for booming sound and rumbling sound, which are professional words used by NVH engineers in automotive technology, are adopted as the descriptive word for a car sound. In the stage of development of a car, the trained NVH engineers evaluate the booming sound and rumbling sound subjectively in car companies. The two sound indexes, which are objective evaluation for booming sound and rumbling sound, are developed in the paper. These indexes are used for the objective evaluation of sound quality for 14 passenger cars, which are samples of famous cars in the world. The results of objective evaluation are expressed in the two-dimensional sound indexes. The brand sounds of famous cars in the world are very well represented in two-dimensional indexes. In the paper, a production car is modified using theses two-dimensional indexes and its sound quality is improved. In future, the two-dimensional sound indexes are very useful for the development of brand sound. [Work supported by BK21 projects and partly supported by Hyundai Motor in Korea.]

3:00

4pNS7. Validation of the cylindrical microphone array for measuring sound power levels of noise sources. Matthew A. Nobile (IBM Hudson Valley Acoust. Lab., M/S P226, Bldg. 704, Boardman Rd. Site, 2455 South Rd., Poughkeepsie, NY 12601)

Recent papers have (*i*) introduced the concept of the cylindrical microphone array for the measurement of sound power levels of noise sources as an alternative to the parallelepiped and hemispherical arrays [M. A. Nobile and J. A. Shaw, Proc. Inter-Noise 99], (*ii*) set forth a detailed procedure for using the cylindrical array that could readily be incorporated into international standards [Nobile *et al.*, Proc. Noise-Con 2000], and (*iii*) presented the results of measurements using that procedure on a 1.75-m-high noise source using a variety of microphone positions to draw conclusions regarding the number of microphones, equal versus unequal areas, and traversing versus fixed-point implementations [Nobile *et al.*, Proc. Inter-Noise 2002]. The cylindrical array has since been incorporated into ISO 7779 and is being proposed for inclusion in the upcoming revision of ISO 3744. In an effort to gain experience with this new array and to help validate its use, a prototype rotating cylindrical array was constructed, and measurements were taken on a variety of noise sources. The results are compared to those taken with the well-established hemispherical microphone array, using 40 coaxial circular paths.

3:15

4pNS8. Optimal configurations for active control of axial fans. Brian B. Monson and Scott D. Sommerfeldt (Brigham Young Univ., N283 ESC, Provo, UT 84602, brianbmonson@byu.edu)

The minimum power radiated from a single monopole noise source surrounded by four symmetrically spaced control sources has been shown analytically. A similar analysis has also been completed for a monopole noise source surrounded by three symmetrically spaced control sources. At Brigham Young University, a multichannel active noise control system has been developed for a small axial cooling fan employing four symmetrically spaced miniature loudspeakers as control actuators. This arrangement of control loudspeakers has exhibited good global control of the fan's tonal noise, achieving close to the predicted ideal for some harmonics of the blade passage frequency. According to theory, a three-source control configuration should be able to achieve nearly the same power reduction as a four-source configuration. While this is so, increasing the number of control sources to greater than four gives little to no increase in predicted power attenuation. A three-source control implementation, then, appears to be a comparable solution for the fan noise control problem. The experimental results of a three-source control and fan configuration will be discussed. An assessment will be made of the three- and four-source arrangements, including also a comparison of the experimental results versus the theoretical ideal.

3:30

4pNS9. Protection against noise produced by open lines of the underground on the example of residential area Butovo in Moscow. Igor L. Shubin, Irina L. Sorokina, and Sergey I. Kryshov (NIISF 21 Lokomotivnay line, Moscow, 127238, Russia, niisf@ipc.ru)

In Moscow (Russia) there is actively conducted construction of new lines of the underground. Nowadays we use the model of the easy underground with an open system of lining for the economy of means for construction in designs of the underground. Till now in Moscow only 7% of the length from the general extent of the underground (276 km) was open. There are plans for construction of the underground with open lining for more than 100 km. One of the most difficult problems of designing lines of the underground with easy lining is the decision on the problem of protection against noise and vibration of adjoining residential areas. In the present work there is written the accepted decisions on protection against noise and vibration of a line of the underground constructed in the microdistrict of Butovo in Moscow. The accepted technical and building-acoustic decisions are resulted.

Session 4pPP

Psychological and Physiological Acoustics: General Topics in Psychological and Physiological Acoustics

Lori L. Holt, Chair
Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Chair's Introduction—1:30

Contributed Papers

1:35

4pPP1. Auditory categorization: Cue weighting and dimension bias. Lori L. Holt (Dept. Psych., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213) and Andrew J. Lotto (Boys Town Natl. Res. Hospital, Omaha, NE 68131)

The ability to integrate and weight information across dimensions is central to perception. The present experiments investigate this issue by training participants to categorize sounds drawn from two input distributions in a two-dimensional acoustic space defined by frequency-modulated sine waves center frequency (CF) and modulation frequency (MF). These dimensions were psychophysically matched to be equally discriminable and, in the first experiment, were equally informative for accurate categorization. Nevertheless, listeners' category responses reflected a bias for use of CF. This bias was moderated when the informativeness of CF was decreased by shifting distributions to create more overlap in CF. A reversal of weighting (MF over CF) was obtained when distribution variance was increased for CF. These results demonstrate that even when equally informative and discriminable, acoustic cues are not necessarily equally weighted in categorization; listeners exhibit biases when integrating multiple acoustic dimensions. Moreover, drastic changes in weighting strategies can be affected by changes in input distribution parameters. This methodology provides potential insights into acquisition of speech sound categories, particularly second language categories. One implication is that ineffective cue weighting strategies for phonetic categories may be alleviated by adding variance to noninformative dimensions in training stimuli. [Work supported by NIH.]

1:50

4pPP2. Overriding the signal: Perception of default versus non-default interconsonantal vowels in Japanese. Yuki Hirose (Univ. of Electro-Commun, 1-5-1 Chofugaoka, Chofu, 182-8585, Japan) and Emmanuel Dupoux (CNRS)

Consonant clusters are illegal in Japanese and in producing foreign words with clusters, the vowel [u] is used as a default epenthetic vowel. Our previous study revealed that Japanese speakers perceive a vowel [u] between consonants in VCCV stimuli, even when the vowel originally present was removed, thereby creating illusion. Our experiment 1 further shows the illusion can override the existing acoustic information supplied by other vowels. The duration of word medial vowels in VCuCV and VCiCV nonwords were manipulated in eight stages from stimuli full to 0 (removed). The Japanese subjects identified the word medial sound by choosing from [a], [i], [u], [e], and ''None.'' For the stimuli derived from VCuCV, majority of the Japanese subjects reported [u] even when most of the vowel information was removed. Other vowels were hardly reported. In contrast, for the stimuli made from VCiCV, [u] was still reported in 60% of the cases, [i] less than 10%, and ''none'' about 30% when the original [i] was largely removed but still existed. Together with the results of experiment 2, an ABX discrimination study, our data suggest that epenthesis is a perceptual phenomenon that could sometimes override existing acoustic information.

2:05

4pPP3. Models of commodulation masking release. Bruce G. Berg and Ying-Yee Kong (Dept. of Cognit. Sci., Univ. of California, Irvine, 3151 Social Sci. Plaza, Irvine, CA 92697-5100)

Three models of commodulation masking release (CMR) are investigated in the context of a task in which masking and flanking bands each consist of five equal-intensity tones spanning 20 Hz. Two are multiple-channel models [J. Hall *et al.*, J. Acoust. Soc. Am. **76**, 50–56 (1984); S. Buus, J. Acoust. Soc. Am. **78**, 1958–1965 (1985)] and one is a single-channel model [B. G. Berg, J. Acoust. Soc. Am. **100**, 1013–1023 (1996)]. Thresholds do not provide an adequate test of the models because all three outperform listeners, thus reducing the solution to arbitrary assumptions about internal noise. However, the models can be discerned with a molecular psychophysical technique. A small intensity perturbation is added to each component of the stimulus. Correlating the trial-by-trial, simulated responses to the perturbations of each component yields a set of spectral weights. It is shown that the three models predict highly distinct patterns of spectral weights.

2:20

4pPP4. A linear mechanical-electrical-acoustic coupled physiological model of the cochlea: Response to acoustic and electrical excitation. Sripriya Ramamoorthy and Karl Grosh (Dept. of Mech. Eng., Univ. of Michigan, 3124 GG Brown, 2350 Hayward St., Ann Arbor, MI 48109-2125, grosh@umich.edu)

A linear physiologically based finite element model is developed for analyzing the global mechanical-electrical-acoustic (active) filtering in the mammalian cochlea. The model consists of a two-duct fluid-filled rectangular geometry, the micro-mechanical structural network interacting with the fluid; the electrical circuit equivalent of cells and fluid in every cross-section is connected by longitudinal cables representing the conductivity of the cochlear fluids and includes the mechano-electrical and electro-mechanical transduction at the OHC. The acoustic pressures, structural displacements, and electrical potentials are determined numerically and compared with experiments. Reducing the amount of activity present in the model reduces the gain and lowers the frequency of peak BM velocity response compared to a fully active model, in accordance with experimental data. This model also possesses near-invariance to click induced noise at different gain levels. Using the same model parameters, the predictions of the local BM velocity response to electrical stimulation match available experimental data, providing an independent test of the model capability. Predictions of electrically evoked otoacoustic emissions are found to match experimental results as well. Roughness introduced into BM stiffness is found to result in fine structures in a fully active model and have little effect on a model with reduced activity. [Research funded by NIH–NIDCD.]

2:35

4pPP5. Active linear modeling of cochlear biomechanics using Hspice. Soon-Suck Jarng and You-Jung Kwon (Dept. of Information Control and Instrumentation, Chosun Univ., 375 Seoseok-Dong, Dong-Ku, Gwang-Ju, South Korea)

This paper shows one and two dimensional active linear modeling of cochlear biomechanics using Hspice. The cochlear biomechanical modeling is based on the deductive approach of others but is processed using Hspice which is the most popular analog electric circuit simulator among electrical and electronic engineers. Because Hspice is so generally used for active filter design, the shortage of the present cochlear model may be supplemented with analog active filters added to the present model. This extra filtering may be arbitrary so as to be called second filters. The advantage of the Hspice modeling is that the cochlear biomechanics may be implemented into an analog IC chip. This paper explains in detail how to transform the physical cochlear biomechanics to the electrical circuit model and how to represent the circuit in Hspice code. The comparison between one and two dimensional models is figured.

2:50–3:05 Break

3:05

4pPP6. Extracting the frequencies of the pinna spectral notches from measured head-related impulse responses. Vikas C. Raykar, Ramani Duraiswami (Perceptual Interfaces & Reality Lab., Dept. of Computer Sci. and Inst. for Adv. Comput. Studies, Univ. of Maryland, College Park, MD 20742), and B. Yegnanarayana (Indian Inst. of Technol. Madras, Chennai-600036, Tamilnadu, India)

Experimental data are available for the Head-related impulse response (HRIR) for several azimuth and elevation angles, and for several subjects. The measured HRIR depends on several factors such as reflections from body parts (torso, shoulder, and knees), head diffraction, and reflection/diffraction effects due to the external ear (pinna). Due to the combined influence of these factors on the HRIR, it is difficult to isolate features thought to be perceptually important (such as the frequencies of pinna spectral notches) using standard signal-processing techniques. Signal-processing methods to extract the frequencies of the pinna spectral notches from the HRIR are presented. The techniques are applied to extracting the frequencies of the pinna spectral notches from the publicly available HRIR databases. A brief discussion relating the notch frequencies to the physical dimensions and the shape of the pinna is given. [The support of NSF Award ITR-0086075 is gratefully acknowledged.]

3:20

4pPP7. A study of the pinna anthropometry and the spectral notch frequencies. Vikas C. Raykar, Ramani Duraiswami (Perceptual Interfaces & Reality Lab., Dept. of Computer Sci. and Inst. for Adv. Comput. Studies, Univ. of Maryland, College Park, MD 20742), and B. Yegnanarayana (Indian Inst. of Technol. Madras, Chennai-600036, Tamilnadu, India)

The head-related transfer function (HRTF) varies significantly between individuals due to the different sizes and shapes of different anatomical parts like the pinnae, head, and torso. Applications in the creation of virtual auditory displays require individual HRTFs for perceptual fidelity. A generic HRTF would not work satisfactorily, since it has been shown that nonindividual HRTFs result in poor elevation perception. While good geometrical models exist for the effects of head, torso, and shoulders, a simple model for the pinna that connects the anthropometry to the features in the HRTF does not exist. One of the prominent cues contributed by the pinna are the sharp spectral notches, which encode the elevation of the sound source. Recently, we proposed a method to extract the frequencies of the pinna spectral notches from the HRTF. In this paper we study the relationship between the extracted notch frequency and pinna anthropometry. Based on these observations new approaches for HRTF customization can be developed. [The support of NSF Award ITR-0086075 is gratefully acknowledged.]

3:35

4pPP8. Comparison of auditory localization performance with various loudspeaker rendering systems. Paul D. Henderson, Ioana N. Pieleanu, Evelyn Way, and Rendell R. Torres (Program in Architectural Acoust., Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, hendep2@rpi.edu)

Sound-field rendering systems based on multiple loudspeakers provide significant benefits as auditory displays for many applications in virtual audio, multimedia, teleconferencing, and auralization. In contrast to headphone-based binaural reproduction, the performance of loudspeaker rendering systems is independent of the HRTFs of the listener, and the listener is physically decoupled from the rendering system. In addition, many rendering techniques such as wave-field synthesis (WFS) are capable of simultaneously presenting an accurate perceptual experience to multiple listeners. However, the sound-field rendering technique, loudspeaker configuration, and number of sources play an important role in both the physical behavior of the rendering system and the associated localization performance of listeners. In this study, we examine the *in situ* localization performance of listeners exposed to various loudspeaker rendering systems using virtual sources. Specific reproduction methods investigated are WFS, first- and second-order Ambisonics (including a periphonic implementation), and stereophonic rendering techniques. In addition to virtual sources in the simulated free field, the effect of adding spatialized reverberation is also reviewed. The results are compared with a study of baseline auditory-system localization performance with real sources. [Work supported by the National Science Foundation and Rensselaer Polytechnic Institute.]

3:50

4pPP9. Directivity pattern simulation of the ears with two pairs of hearing aid microphone arrays by BEM. Soon-Suck Jarng and You-Jung Kwon (Dept. of Information Control & Instrumentation, Chosun Univ., 375 Seoseok-Dong, Dong-Ku, Gwang-Ju, South Korea)

The noise reduction of the ITE (in-the-ear) hearing aid (ha) can be achieved by an array of microphones. Each of the right and the left ears is assumed to be fixed with two HA microphones. This two HA microphone array produces particular patterns of directivity by some time delay between two microphones. The directivity pattern geometrically increases the S/N ratio. The boundary element method (BEM) was used for the three dimensional simulation of the HA directivity pattern with the two pairs microphone arrays. This numerical analysis may then be applied for the calculation of the time delay parameter of the digital hearing aid DSP chip. The separation between two microphones was fixed to 10 mm. The time delay between the two microphones was changed to produce the most narrow directivity pattern in the forefront of the head. The variation of the time delay was examined in accordance with input frequencies.

4:05

4pPP10. Skull simulator for design of hearing protection systems for bone-conducted sound. B. Scott May, Anthony J. Dietz, Odile Clavier, and Steve Bilski (Creare, Inc., P.O. Box 71, Hanover, NH 03755)

A skull simulator was developed for an investigation of bone-conducted sound transmission, which can be a significant source of sound in extreme noise environments. The simulator was built from a human skull, with silicon gel used to model internal organs and silicon and latex used to model the skin. Accelerometers attached to the skull bones were used to measure the skull vibration response. Mechanical point impedance measurements on the simulator were compared to results reported in the literature for humans, cadavers, and skulls. Reasonable agreement with these results served to validate the simulation. The link between skull

vibration measured on the skull simulator and bone-conducted sound at the cochlea of a human subject was determined by comparing the response at the mastoid measured on the skull with similar measurements on a human subject. The response of the skull simulator in a sound field was then measured. The frequency response peaked from 1 to 3 kHz, which is where the bone-conduction limit to hearing protection attenuation is a minimum. The simulator was used to determine the effectiveness of various hearing protection components in reducing skull vibration. The results are compared to attenuation measurements of these components on human subjects.

THURSDAY AFTERNOON, 18 NOVEMBER 2004 ROYAL PALM SALON 5, 2:00 TO 3:45 P.M.

Session 4pSA

Structural Acoustics and Vibration: Measurement Methods

Scott D. Sommerfeldt, Chair
Department of Physics, Brigham Young University, Provo, Utah 84602

Contributed Papers

2:00

4pSA1. On the accuracy of modal parameters identified from exponentially windowed, noise contaminated impulse responses for a system with a large range of decay constants. Matt S. Allen and Jerry H. Ginsberg (Georgia Inst. of Technol., Atlanta, GA 30332-0405)

Multi-degree of freedom dynamic systems may contain dynamic modes with a wide range of decay rates. In such a case, collecting data over a time span long enough to capture slowly decaying modes serves to magnify the effect of noise on quickly decaying modes because their response at later times is dominated by noise. An exponential window may be applied to the data to reduce the effect of noise on quickly decaying modes or minimize leakage errors. Some practitioners may instead use a Hanning or similar window for the same purpose. This study investigates the effect of windowing on the accuracy of modal parameters identified using an experimental modal analysis technique. The data for this evaluation are noise contaminated analytical impulse response samples from which frequency response functions are obtained via FFTs. It is seen that accurate results may be obtained for the slowly decaying modes of the system even when their decay constants are many times smaller than the decay constant of the exponential window. The role of exponential windowing in reducing the effect of noise on quickly decaying modes is also demonstrated. In contrast, other windows bias the results. [Work supported by a National Science Foundation Graduate Research Fellowship.]

2:15

4pSA2. Development of a near-field acoustic holography system based on pressure and velocity measurements. Michael C. Harris, Jonathan D. Blotter (Dept. of Mech. Eng., Brigham Young Univ., Provo, UT 84602), and Scott D. Sommerfeldt (Brigham Young Univ., Provo, UT 84602)

Since its conception in 1980, NAH has become an accepted analysis tool with a number of commercial packages now available. One limitation of NAH is that the acoustic field is reconstructed through a spatial sampling. Therefore, this technique becomes less efficient at higher frequencies where microphone array spacing must be reduced. Analytical models have indicated that a pressure and velocity-based NAH method will provide the same reconstruction resolution of current NAH methods with significantly fewer measurement locations. This would lead to a considerable savings in data acquisition time for scanning array systems and reduce the inefficiency at high frequencies. The pressure and velocity-based reconstruction method will be introduced. Experimental work towards validation of the analytical models will be presented.

2:30

4pSA3. Reconstructing transient acoustic field resulting from an explosion. Huancai Lu and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., Detroit, MI 48202)

Generalized Helmholtz equation least squares (HELS) formulations for reconstructing transient acoustic radiation from an arbitrary object subject to arbitrarily time-dependent excitations are developed. Laplace transform is employed to build a transfer matrix for a multi-input multi-output system and reconstructed signals are obtained using convolution with respect to the input data in the time domain directly. Stability of solutions is realized by controlling the pole-residue locations in the Laplace domain. To facilitate calculations of the poles and residues, ztransform is utilized. To validate this general transient HELS formulation, the reconstructed pressure histories are compared with those measured by pressure sensors at the same locations. Space-time tomography resulting from different charges is produced, and the acoustic pressure distribution over the entire time record is reconstructed. It is interesting to notice that this transient HELS formulation enables one to trace time histories of blast pressures back to the explosion location. In other words, it is possible to reconstruct an explosion back in space and time based on the time-domain signals measured at a later time. [Work supported by NSF.]

2:45

4pSA4. Method of superposition applied to patch holography. Angie Sarkissian (Naval Res. Lab., Washington, DC 20375)

The method of superposition is applied to reconstruct the field on a patch on the surface of a radiating structure from measurements made on a nearby surface. In patch holography the measurement surface need not surround the entire structure. It need only be approximately as large as the patch on the structure surface where the reconstruction is required. The surfaces may be planar or curved. The method of superposition is applied where the field on and near the measurement surface is approximated by the field produced by a large number of sources placed on a nearby surface inside the radiating structure. The source strengths are determined by applying boundary conditions on the measurement surface. Because of the ill-posed nature of the problem, this procedure requires the inversion of an ill-conditioned matrix. Singular value decomposition is applied to invert it. Once the source strengths are determined, the field on the partial structure surface may be approximated by the field produced by the sources. [Work supported by ONR.]

3:00

4pSA5. Assessment of accumulated structural damage with the help of linear and nonlinear acoustic techniques. Yulian Kin, Chenn Zhou, Eric Roads, Bernard Parsons (Purdue Univ. Calumet, Hammond, IN 46323), Alexander Sutin (Stevens Inst. of Technol., Hoboken, NJ), Pinakin Shaubal, and Frank Huang (Ispat Inland Steel, East Chicago, IN)

The paper presents applications of linear and nonlinear acoustic techniques for reliable detection of early stages of fracture, fatigue, and other damage, as well as investigation of structure-sensitive acoustical signatures for fundamental material research. The experimental investigations were conducted on steel samples, carbon bricks, and pavers under different types of loading (static and cycling regimes) in laboratory conditions and, in field, to detect progressive changes in a blast furnace wall. A special effort was made to investigate the changes of acoustical response after aging of materials, but before a warning of damage (necking, crack initiation, crazing, disbonding, surface flow, etc.) is physically observed in service life. Measurements were made periodically as structural damage accumulated, but no visible surface crack formation was observed during the tests. The described techniques can be used for the assessment of structure integrity, remaining life, and adjustment of maintenance schedules in applications as diverse as reactor containment wall damage, inspection of damage in buildings, bridges, furnaces, pipelines, aging aircraft, etc. [Support of Indiana 21st Century Fund is greatly appreciated.]

3:15

4pSA6. Acoustic inspection of friction-stir weld. John W. Doane (GTRI and Dept. of Mech. Eng., Georgia Inst. of Technol., 771 Ferst Dr., Atlanta, GA 30332, john.doane@gtri.gatech.edu)

Friction stir welding is gaining popularity as a joining technique in many manufacturing arenas. As it makes the transition from prototype evaluation status into mainstream, large-scale usage, there is an increasing need for comprehensive, efficient, nondestructive inspection methods for qualifying friction stir welded joints. Two of the most common defects in this type of weld are cavities inside the weld, called wormholes, and weld depth excursions. Each of these defects presents a potentially serious reduction in joint strength and must be found soon after part production. Present inspection methods, including penetrant dyes and x-ray imaging, are either ineffective or not practical. In this paper a nondestructive inspection method, using high-frequency acoustic waves, is used to detect these defects by measuring scattered signals using a laser-based acoustic receiver system. The experimental configuration, methodology, and sensitivity of the technique to the defects of interest are discussed as well as its advantages and disadvantages compared to the present methods. [Work supported by Lockheed-Martin Co.]

3:30

4pSA7. Active vibration damping for optical tables. Vyacheslav Ryaboy (Newport Corp., 1791 Deere Ave., Irvine, CA 92606, vryaboy@newport.com)

The paper considers theoretical background, analysis and experimental verification of an active vibration damping system for optical tables. Optical table is a generic name for a stiff platform supporting vibration-sensitive equipment. A typical optical table is a sandwich structure consisting of two faceplates and a lightweight (usually honeycomb) core. Due to low weight-to-stiffness ratio, these platforms found wide application in optical research and high-precision manufacturing, usually in conjunction with soft pneumatic vibration isolators. Although excellent isolation from floor vibration can be achieved in these systems, the platform deviates from the ideal rigid-body behavior at natural frequencies of its flexural vibrations. The theoretical analysis and experimental data reported in this paper show that a properly designed active damping system consisting of a small number of collocated sensor-actuator pairs provides robust and universal means of attenuating these unwanted vibrations.

THURSDAY AFTERNOON, 18 NOVEMBER 2004 CALIFORNIA ROOM, 2:00 TO 5:00 P.M.

Session 4pSC

Speech Communication: Sounds, Speaking and the Vocal Tract (Poster Session)

Marianne Pouplier, Chair

Biomedical Sciences, University of Maryland Dental School, 666 West Baltimore Street, Baltimore, Maryland 21201

Contributed Papers

All posters will be on display from 2:00 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 2:00 p.m. to 3:30 p.m. and contributors of even-numbered papers will be at their posters from 3:30 p.m. to 5:00 p.m.

4pSC1. Production and perception of clear speech in Croatian and English. Rajka Smiljanic and Ann Bradlow (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208)

This paper presents a cross-language comparison of clear speech, a distinct mode of speech production intended to enhance intelligibility. Specifically, it explores whether language specific phonological properties guide the articulatory modifications of clear speech in Croatian and English, and examines whether clear speech enhances intelligibility in both languages. Five Croatian and five English speakers read 20 nonsense sentences in Croatian and English, respectively, in conversational and clear speaking styles. In a sentence-in-noise perception test, listeners recognized words more accurately in clear than in conversational speech in their native language, establishing that the clear speech intelligibility effect is not English-specific. Acoustic analyses showed that both phonemic (e.g., the vowel length contrast in Croatian) and allophonic (e.g., the tense versus lax vowel duration difference in English) contrasts are enhanced in clear speech. Furthermore, speakers enhance contrasts even when categories seem to be sufficiently distinct and there is little chance of confusability. For example, the vowel space is expanded equally in English (with 11 vowels) and in Croatian (with just five vowel qualities). The results suggest that clear speech production is driven by the phonological properties of the language, but at the same time all segments are hyperarticulated regardless of inventory based considerations.

4pSC2. Talker differences in clear and conversational speech: Preliminary data with monosyllabic words. Sarah Hargus Ferguson (Dept. of Speech-Lang.-Hear., Univ. of Kansas, Dole Ctr., 1000 Sunnyside Ave., Rm. 3001, Lawrence, KS 66045, safergus@ku.edu)

Several studies have shown that when a talker is instructed to speak as though talking to a hearing-impaired person, the resulting clear speech is significantly more intelligible than typical conversational speech. A recent study of 41 talkers [S. H. Ferguson, J. Acoust. Soc. Am., in press] demonstrated that for normal-hearing listeners identifying vowels in noise, the amount of improvement talkers achieve by speaking clearly varies widely. The current pilot experiment begins to extend this line of research to meaningful materials using two talkers from the Ferguson database, one male and one female. While the clear speech vowel intelligibility benefit was similar for these two talkers, overall vowel intelligibility was greater for the female talker than for the male talker. Monosyllabic words produced in clear and conversational speech by these talkers were excised from the sentences in which they were recorded and presented in noise for identification by young, normal-hearing listeners and elderly listeners with hearing impairment. These data will provide important information about the relationship between vowel intelligibility and overall speech intelligibility. Results will be discussed in this light as well as in terms of expected talker variability in the clear speech effect for meaningful materials. [Database development supported by NIHDCD-02229.]

4pSC3. The influence of speaking rates and utterance units on segmental duration of Mandarin speech. Jing-Yi Jing (Dept. of Special Education, Natl. Taiwan Teachers College, Taiwan, jjeng@ipx.ntntc.edu.tw)

The main purpose of this study is to investigate the effect of five different speaking rates and five utterance units on segmental durations of Mandarin. Thirty normal speakers participated in the speech sampling. The speakers read the speech material at five speaking rates, including the fastest, faster, median, slower, and slowest rate. There were five utterance units for each target syllable, including monosyllable, disyllable, trisyllable, four syllable phrases, and sentences. These segments measured include VOT of stops, noise duration of fricatives and affricates, nasals and vowel duration, as well as silence intervals, or pause intervals. Results showed that the consonant and vowel durations were all significantly influenced by speaking rates and utterance units. At five kinds of speaking rates, the durations of vowels changed more than those of consonants. Fricative durations expanded more than the durations of other consonants, while the duration of unaspirated stops kept constant at five speaking rates. Vowels in monosyllable had longer duration than those in other utterance units. In addition to speaking rates and utterance units, the syllable structures and tones also greatly influenced the vowel and syllable durations of Mandarin.

4pSC4. Perception of Thai tones in citation form and connected speech. Elizabeth C. Zsiga and Rattima Nitisaroj (Dept. of Linguist., Georgetown Univ., Washington, DC 20057, zsigae@georgetown.edu)

*F*0 contour shapes of the five tones of Thai differ from citation form to connected speech [Potisuk *et al.*, Phonetica **54**, 2242 (1997)]. "Falling" tones have a rise-fall contour in citation, but may be realized as a rise in fluent speech. This study investigates whether Thai listeners can reliably distinguish tones in connected speech, and whether any cues to tone identity remain stable across contexts. In four experiments, ten Thai listeners identified naturally produced and digitally altered tones in a forced-choice task. In experiment one, listeners identified naturally produced "falling" tones with 100% accuracy in citation forms and 96% accuracy in sentences, despite differing contours. In experiments 2–4 (replicating and extending Abramson [Lg&Sp **21**, 319325 (1978)]), *F*0 onset and offset, peak height, and peak alignment were systematically modified on syllables in citation and sentence contexts. In all contexts, tones identified as high or low had an *F*0 peak or valley aligned to the right edge of the syllable. Mid tones had no *F*0 inflection. Tones identified as falling or rising necessarily had a pitch inflection at the syllable midpoint (end of the first mora). These findings support the view that tones are aligned with moras in Thai.

4pSC5. The production of the stop voicing distinction in French at normal and rapid speaking rates. Nassima Abdelli-Beruh (Dept. of Speech-Lang. Pathol. and Audiol., New York Univ., 719 Broadway, Ste. 200, New York, NY 10003 and Dept. of Speech and Hear. Sci., City Univ. of New York, 365 Fifth Ave., New York, NY 10016)

This study examined how monolingual Parisian French speakers produced the stop voicing distinction at two different speaking rates (normal, rapid) in two different phonetic contexts (between vowels, between voiceless-fricatives). Results showed that all the acoustic correlates of the voicing distinction under study (syllable-initial: VOT, closure duration and closure voicing; syllable final stops: preceding vowel duration, closure duration, closure voicing and release) were shorter in the rapid speaking rate condition than in the normal speaking rate condition. Furthermore, voicing-related differences were also smaller in the rapid speaking rate condition than in the normal speaking rate condition to a degree that varied with each acoustic correlate and each phonetic context. Overall, similar patterns of voicing-related duration differences and closure voicing differences were found in the normal and rapid speaking rate conditions, but the perceptual salience of small voicing-conditioned duration differences remains to be investigated.

4pSC6. A cross-language study of the influence of duration and intensity on perception of lexical stress. Cheryl L. Moran, Amy E. Coren, Jessica F. Hay, and Randy L. Diehl (Dept. of Psych., Univ. of Texas, 1 Univ. Station A8000, Austin, TX 78712-0187, cmoran@mail.utexas.edu)

The present study examines perceived lexical stress across five languages which differ in their prosodic characteristics: English, French, German, Japanese, and Turkish. Native speakers from each language were presented with a two-syllable nonsense word, (/bibi/), and were asked to judge whether the first or second syllable was more prominent. The relative durations and intensities of the two syllables were varied orthogonally across ranges, which were determined on the basis of analyses of natural utterances from the five languages. The key issue addressed is whether judgment of syllable prominence is cross-linguistically invariant or dependent on exposure to language-specific lexical stress patterns. [Work supported by NIDCD.]

4pSC7. Envelope offset asynchrony as a cue to voicing in final english consonants. Hanfeng Yuan, Charlotte M. Reed, and Nathaniel I. Durlach (Res. Lab. of Electron., MIT, Cambridge, MA 02139)

Previous research [Yuan *et al.*, J. Acoust. Soc. Am. **113**, 2291 (2003)] has demonstrated that a simply derived, real-time measurement of the envelope-onset asynchrony (EOA) of two different bands of speech serves as a reliable cue to syllable-initial voicing. The current study was undertaken to determine the applicability of a similar type of temporal cue to voicing of final consonants. This cue is based on the time asynchrony between the offsets of amplitude-envelope signals derived from two different bands of speech (a low-pass-filtered band at 350 Hz and a high-pass-filtered band at 3000 Hz). Measurements of the offset-timing difference between the low- and high-frequency envelopes (EOFA) were obtained on a set of 16 final consonants represented through multiple tokens of C1VC2 syllables. The EOFA measurements were fit to Gaussian distributions and signal-detection theory was applied to determine the discriminability of pairs of voiced–voiceless cognates by an ideal observer. These results indicate that the EOFA cue provides a reliable and robust cue for distinguishing voicing in syllable-final position regardless of manner

or place of production. This cue has applications to the design of sensory aids for the deaf as well as to automatic speech recognition. [Research supported by NIDCD Grant R01-DC00126.]

4pSC8. Place of articulation effects on the acoustic correlates of the stop voicing distinction in French. Nassima Abdelli-Beruh (Dept. of Speech-Lang. Pathol. and Audiol., New York Univ., 719 Broadway, Ste. 200, New York, NY 10003 and Dept. of Speech and Hearing Sci., City Univ. of New York, 365 Fifth Ave., New York, NY 10016)

This study examined the influence of place of articulation on some acoustic correlates of the voicing distinction in syllable-initial (VOT, closure duration and closure voicing) and in syllable final stops (preceding vowel duration, closure duration, closure voicing and release) embedded in different French sentence contexts (between vowels, between voiceless-fricatives). Results indicated that, as in English, VOT increases as place of articulation moves back in the mouth, labial closures are longer than alveolar and velar closures and preceding vowel durations are shorter before alveolars than before labials and velars. Finally, /b/'s are significantly more often phonated than /d/'s or /g/'s in the between voiceless-fricatives context only. The effects of place of articulation on closure and preceding vowel durations are mediated by the factors of voicing category, local phonetic context and syllable-position.

4pSC9. Aerodynamic reduction in American English flapping. Sidney Martin and Joaquin Romero (Universitat Rovira i Virgili, Pl Imperial Trraco 1, 43005 Tarragona Spain, smm.predoc@estudiants.urv.es, jrog@fll.urv.es)

Are all tongue tip gestures for /t,d/ the same? Acoustic and articulatory studies have revealed that /t,d's show a gradient behavior, in which invariancy is nonexistent. These results undermine phonological rules which represent leniting processes, such as flapping, as a rule where /t,d/ is replaced by an invariant flap. Adding to more evidence for gradiency, an aerodynamic experiment was designed to investigate the nature of alveolar tongue tip gestures, specifically in flapping contexts. Three native speakers of American English were recruited and each yielded 200 tokens, which contained medial /t,d/ in different phonetic contexts (stressed-syllable onset, unstressed-syllable onset, n+/t,d/, l+/t,d/ and r+/t,d/) and in two speaking rates. For each target consonant, duration ratio and airflow rate were obtained. In general, both duration ratio and airflow rate showed gradiency across rate and context. Tokens at a normal speech rate yielded higher ratios and airflow rates than tokens at a faster speech rate. However, rate did not affect each context equally. Accordingly, /t,d's reveal a gradient behavior which ranges from a canonical stop to complete deletion.

4pSC10. Greek fricatives: Inferring articulation from *F*2 at vowel onset. Chao-Yang Lee and Georgia Malandraki (School of Hearing, Speech and Lang. Sci., Ohio Univ., Athens, OH 45701)

Fricative consonants are produced with a narrow constriction in the vocal tract. According to the acoustic theory of speech production (Fant, 1960), fricatives are characterized by spectral prominences corresponding to the natural frequencies of the cavity anterior to the constriction. The spectral characteristics may also be modified by acoustic coupling of the noise source to the cavity behind the constriction (Stevens, 1998). In addition to the defining acoustic attributes, other articulatory gestures may be added to enhance the perceptual saliency of the fricatives (Stevens, 1989). This study examines the second formant frequency ($F2$) at the onset of a vowel following a fricative as an indication of tongue body position, a potential enhancing gesture for fricatives. $F2$ measurements were obtained from a database of 1950 tokens including all ten Greek fricatives (five places of articulation, two voicing categories) produced in five vowel contexts by 25 speakers. The results indicate highest $F2$ for palatals, followed by alveolars, dentals, labiodentals, and velars. Despite some variability by vowel context, the $F2$ values are reliably distinct among the five places of articulation. These findings are discussed in terms of the acoustic theory and the enhancement account.

4pSC11. Developmental processes in production of Korean /s/ and /s*/. Hyunjoo Chung and Anna M. Schmidt (Kent State Univ., School of Speech Pathol. and Audiol., Kent, OH 44242, hchung@kent.edu)

The purpose of this study was to investigate development of Korean children's productions of tense and lax Korean s. Korean alveolar fricatives are acquired late; /s*/ is generally later than /s/. However, little is known about developmental processes in acquiring the contrast between tense and lax Korean alveolar fricatives. Twenty typically developing monolingual Korean children between 4 and 8 and ten monolingual Korean adults were recruited in Korea. All children fell within normal limits on both a Korean articulation test and a standardized Korean vocabulary test. A picture-naming picture task elicited productions of initial fricatives, alveolar stops, and affricates followed by /a, i, u/ in monosyllabic words. Production error patterns will be analyzed to describe general developmental trends in production. More importantly, the contrast between productions of /s/ and /s*/ will be presented based on acoustic parameters including frication durations, centroid frequency, and harmonic-related amplitude measurements. Finally, similarities and differences in each acoustic parameter will be compared between adults and children. Results will discuss which acoustic parameter(s) effectively describe the distinction between productions of /s/ and /s*/ as well as differences between productions of adults and children between age 4 and 8.

4pSC12. The status of vowels in Jordanian and Moroccan Arabic: Insights from production and perception. Jalaleddin Al-Tamimi, Egidio Marsico (Laboratoire Dynamique du Langage, Institut des Sci. de l'Homme, 14, Av. Berthelot, 69007 Lyon, France, jalal-eddin.al-tamimi@etu.univ-lyon2.fr), and Rene Carre (ENST, Paris, France)

Due to the morphological structure of Arabic, vowels only occur in consonantal environment, never in isolation. The issue addressed here is whether static acoustic cues are sufficient to characterize Arabic vowels or dynamic ones are needed. This study is based on Jordanian and Moroccan Arabic (JA & MA), with eight and four vowels, respectively. Ten speakers per dialect were tested on both production and perception. In production, speakers produced vowels in words, syllables, and in isolation (in alveolar context). In perception, they categorized isolated vowels on an $F1/F2$ plan. Results from production show that, both in JA and MA, (a) vowels in isolation have greater variation than in syllables or words, (b) vowels in a word context are more centralized (than in syllables or in isolation), and that (c) MA vowels are more centralized than JAs. In perception, a high degree of variability within and between dialects is found. A new experiment of vowel perception in syllable context (CV) (i.e., dynamic trajectories) was elaborated to study the effect of contextual information on vowel perception in Arabic. Results on five JA speakers show that the variability has been lessened and that dynamic information seems to be more useful in CV categorization.

4pSC13. Phonetic properties of three Ayuuk (Mixe-Zoquean) dense vowel systems. Avelino Heriberto (Dept. of Linguist., Univ. of California at Berkeley, avelino@berkeley.edu)

Several theories of vowel dispersion have proposed that the vowel systems of languages tend to distribute evenly in the phonetic space to maximize the perceptual distinctness of the vowels (Liljencrants and Lindblom, 1972; Lindblom, 1986; Lindblom and Maddieson, 1988; Schwartz *et al.*, 1997). This study investigates the acoustic properties of three Ayuuk (Mixe) languages spoken in Mexico, possessing dense vocalic systems, which in addition have a series of normal and the corresponding set of long vowels: Santa Mara Ocotepec Ayuuk (SMOA) with nine vowels

/i, I, u, e, E, o, A, a, O/, Tlahuitoltepec Ayuuk (TA) with eight vowels /i, I, e, E, o, A, a, O/, and Alotepec Ayuuk (AA) with seven vowels /i, u, e, E, o, a, O/. The inventories of the Ayuuk languages are typologically interesting; the SMOA and AA systems are rather uncommon (Maddiesson, 1984), in contrast with the system of TA, which conforms to the main trends of vowel systems. Preliminary results lend support to a dispersion/focalization hypothesis in that the vowels are evenly dispersed in the available space and remain focalized without significant overlap with adjacent vowels. The phonetic properties analyzed in this paper suggest that two principles, maximal dispersion and maximal distinctness, shape the overall distribution of the vowel inventory of Ayuuk languages.

4pSC14. Diphthongization as cue for the automatic identification of British English dialects. Emmanuel Ferragne and Franois Pellegrino (Laboratoire Dynamique Du Langage, ISH, 14 avenue Berthelot, 69007 Lyon, emmanuel.ferragne@univ-lyon2.fr)

The degree to which a given speaker of British English diphthongizes her/his vowels has been known for decades to be a good indicator of this speaker's dialectal origin: for instance, some Northern varieties tend to have variants perceived as monophthongs where Standard Southern English has a perceived diphthong in the word GOAT. A portion of the recently released Accents of the British Isles database (14 dialects recorded throughout the British Isles, 20 speakers per dialect) is investigated to determine whether this view still holds at least acoustically for contemporary dialects. Several metrics (based on formant-frequency estimation) are proposed in an attempt to capture a ''degree'' of diphthongization. In a preliminary and rather oversimplified study, the standard deviation of $F2$ was measured over the whole duration of the vowel in the word BOAT for all speakers and dialects. A nonparametric one-way analysis of variance was performed to test the effect of ''dialect'' on diphthongization. The effect turned out to be significant at the 0.01 level. More elaborate procedures (including all first three formants) will be shown in the final presentation.

4pSC15. Nasalization, neutralization, and merger in English front vowels. Jerry Liu (Dept. of Linguist., USC, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693, jerryliu@usc.edu)

Before a velar nasal, a tense/lax contrast in the front vowels is generally considered to be conflated to only the set of lax vowels in American English. However, Ladefoged (2001, A Course in Phonetics) states that many younger Americans pronounce ''sing'' with a vowel closer to that in ''beat'' rather than to that in ''bit.'' This paper examines front vowels produced by nine young Californians before the velar nasal to document whether or not the vowel, in fact, raises. Two outcomes are possible: (1) a front and raised vowel intermediate between its tense and lax counterparts (which we call neutralization) or (2) a lax vowel (which we call merger). Further, the pattern for both high- and midfront vowels is examined. The results demonstrate a consistent pattern of neutralization (i.e., an intermediate vowel) for both the high- and mid-vowels, the latter a previously unattested environment for this pattern of raising before velar nasals.

4pSC16. An ultrasound and tagged cine MRI study of German vowels. Marianne Pouplier (Vocal Tract Visualization Lab, BMS, UMAB Dental School, 666 W Baltimore St, Rm 4-D-20, Baltimore, MD 21201; mpoup001@umaryland.edu), Adam Buchwald (Johns Hopkins, Baltimore, MD 21218), and Maureen Stone (UMAB Dental School, Baltimore, MD 21201)

German vowels contrast along the tense–lax as well as the rounding dimension, rendering a relatively complex (front) vowel system. The difference between tense and lax vowels is often described in terms of more extreme articulator movement and greater muscular tension for tense vowels, manifest especially in tongue root displacement and tongue height. These factors are assumed to correlate in the Germanic languages. This study uses ultrasound data to investigate global tongue shape and in particular tongue root differences for German monophthongs in stressed position. Preliminary results for one speaker indicate that some vowel pairs are nearly identical in tongue height, but differ in tongue root displacement. Rounded front vowels are more retracted and lower than their unrounded counterparts. The /a: a/ pair, which has been described as contrasting in duration only, exhibits a clear qualitative difference as well. In addition, tagged cine MRI data were collected for a subset of high and mid vowels. This technique allows inferring expansion and contraction of the tongue musculature and thus can inform about differences in muscular tension between tense and lax vowels. [Supported by NIH Grant RO1-DC01758.]

4pSC17. 3D tongue segmentation for a minimal pair sentence. Melissa A. Epstein and Maureen Stone (Biomed. Sci., Rm. 5A12, Univ. of Maryland Dental School, 666 W. Baltimore St., Baltimore, MD 21201)

The tongue appears to have as many as five articulatory control regions. If this is true, each of these regions of the tongue, or segments, could be independently controlled. What does it mean to be independently controlled? Independent control would be manifested as low correlations between segments, because independent segments could move asynchronously in speech. Previous studies for midline data have supported this and have shown that segments can aggregate or deaggregate dependent on the phonemic content of the speech sample. However, the tongue is a 3D structure, and its muscles are more localized lateral-to-medial than anterior-to-posterior. Therefore, additional lateral units of control are possible [Stone, J. Acoust. Soc. Am. (1990)]. This study looks at correlations between front to back segments of midline points, lateral points, and 3D shape data to determine whether lateral segments need to be considered independently of midline segments. We will also consider what features besides functional independence might cause low correlations for midline or lateral segments.

4pSC18. An acoustic and perceptual study of voiced geminates in Japanese. Shigeto Kawahara (Dept. of Linguist., UMass, Amherst, MA)

In Japanese, voicing in geminates is subject to dissimilation while voicing in singletons is not. This paper advances a hypothesis that voicing in geminates is not as perceptually salient as voicing in singletons, and therefore losing voicing in geminates is accepted as an instance of ''perceptually tolerated articulatory simplification'' (Kohler, 1990), while devoicing singletons is too perceptually conspicuous. This paper reports an acoustic and perceptual experiment that supports this hypothesis. The acoustic study shows that closure voicing and closure duration are significantly weakened in geminates: Voicing ceases relatively early during the closure, and the difference between voiced and voiceless geminate closure length is smaller than the closure difference in singletons. The second experiment is a perceptual study, which is an identification task under a noisy environment. The results directly support the hypothesis that voicing in geminates is hard to hear, while voicing in singletons is reliably perceived. Taken together, the results of the two experiments provide support for the view that voicing in geminates may be prone to neutralization because it is poorly perceived. This conclusion, more generally, lends support to the view that contrasts signaled by weak phonetic cues are prone to phonological neutralization (Kohler, 1990; Steriade, 1997).

4pSC19. Movement duration and movement stiffness in long and short consonants. Anders Lofqvist (Haskins Labs., 270 Crown St., New Haven, CT 06511; lofquist@haskins.yale.edu)

This study examined the relationship between movement duration and movement stiffness in speech. Stiffness was measured as the interval from movement onset to peak velocity. The speech material consisted of long and short Japanese consonants produced by native speakers. A magnetometer was used to track articulatory movements. For the lower lip, the

velocity of vertical closing movement for bilabial consonants was measured, while the speed of the tongue tip was measured for dental stops and fricatives. Results from six subjects show that the lower lip movement duration was longer for the long consonants, which also had a lower stiffness than their short cognates. The results also show higher correlations between duration and stiffness for the short consonants. The results for the tongue tip showed more variability across and within subjects, although the overall trend was similar to the one found for the lower lip. In addition, all correlations between tongue tip movement duration and stiffness were high. These results extend earlier work on the same relationship, where it has mostly been studied across variations in speaking rate. They also suggest potential differences between articulators and between stops and fricatives. [Work supported by NIH.]

4pSC20. Perturbation and compensation in speech acoustics using a jaw-coupled robot. Mark K. Tiede (R.L.E. Speech Comm. Group, MIT, Cambridge, MA, tiede@speech.mit.edu, and Haskins Labs, New Haven, CT), Frank H. Guenther (Boston Univ., Boston, MA), Joseph S. Perkell, Majid Zandipour (MIT, Cambridge, MA), Guillaume Houle, and David J. Ostry (McGill Univ., Montreal, PQ, Canada)

Observations were made in three speakers of compensation in formant trajectories in response to jaw perturbations during utterances with the general form /siyCVd/, as in ''see red.'' Custom dental prostheses were used to help immobilize the head (upper jaw) and couple a computer-controlled robotic device (lower jaw). A 3-Newton perturbation force was applied to the jaw during one out of every five repetitions, selected at random, with half of the perturbations applied downward and half upward. Perturbations were triggered from jaw opening (for CV) exceeding a threshold relative to clench position. Audio (at 10 kHz) and jaw position (at 1 kHz) were recorded concurrently. Individual tokens were extracted using the perturbation threshold for alignment. Formants computed over these intervals show initial deviation from control trajectories and then compensation that begins 60–90 ms after perturbation. Since jaw position does not recover its unperturbed trajectory, compensation presumably is effected through modified tongue movements. The observed behavior is compatible with the function of the DIVA model of speech motor planning, in which corrective motor commands are computed in response to errors between anticipated and produced sensory (auditory and somatosensory) consequences. [Research supported by NIDCD.]

4pSC21. Effects of jaw perturbation on cortical activity during speech production. Jason A. Tourville, Frank H. Guenther, Satrajit S. Ghosh, and Jason W. Bohland (Dept. of Cognit. and Neural Systems, Boston Univ., 677 Beacon St., Boston, MA 02215, jtour@cns.bu.edu)

Functional magnetic resonance imaging (fMRI) was used to investigate cortical activation caused by introducing a perturbation to the jaw during speech. The DIVA model of speech production [e.g., Guenther and Ghosh (2003), Proceedings of the ICPhS. Barcelona, ICPhS Committee] predicts that jaw perturbation will cause a mismatch between learned target (expected) sensations and actual sensory inputs in higher-order somatosensory cortical areas in the supramarginal gyrus, and that this mismatch will drive compensatory activations of motor cortical areas. To test this hypothesis, we performed an fMRI experiment in which articulation of pseudo-words was perturbed by the rapid inflation of a small, stiff balloon held between the subjects molars. The perturbation condition occurred pseudo-randomly on $\frac{1}{7}$ of the trials. After controlling for brain activation due to normal speech production (no perturbation) and due to the sensation of balloon inflation alone (no production), two regions responded preferentially to the perturbed speech condition: one region along the boundary of the angular gyrus and posterior supramarginal gyrus in inferior parietal cortex, and another in the ventral posterior inferior frontal gyrus extending into the frontal operculum and anterior insula. These results provide evidence for the type of speech motor control processes hypothesized by the DIVA model.

4pSC22. Sound propagation through the large front-cavity volume of American English rhotic sounds. Zhaoyan Zhang, Carol Espy-Wilson (Dept. of Elec. and Computer Eng., Univeristy of Maryland, College Park, MD 20742), Suzanne Boyce (Univ. of Cincinnati, Cincinnati, OH 45202), and Mark Tiede (Haskins Labs, New Haven, CT 06511)

Rhotic sounds in American English (AE) are produced with a large front-cavity volume between the palatal constriction and the lip constriction which is the source of the characteristic low $F3$ of /r/'s. Depending on its shape, this volume may be modeled as (a) a single tube or (b) a single tube with a side branch. Both models produce a low $F3$, but through different mechanisms that involve differences in wave propagation direction. In this paper, we describe finite element studies of the acoustics of the front cavity, using dimensions derived from MRI studies of different subjects producing AE /r/. In particular, we discuss the case of a front cavity with minimal sublingual space. The sound-pressure distribution in the front cavity and the possible sublingual cavity was calculated at various harmonic excitations. The results show that at low frequencies (up to 2000 Hz, depending on specific dimensions) the front cavity is better modeled as a single tube, and at higher frequencies as a single tube with a side branch. The implications of accounting for wave propagation patterns (and thereby acoustically critical points along the vocal tract) on extraction of dimensional measures from MRI data will be discussed.

4pSC23. A formant-to-area conversion technique based on acoustic sensitivity functions. Brad Story (Univ. of Arizona, Dept. of Speech and Hearing Sci., P.O. 210071, Tucson, AZ 85721-0071)

A technique for transforming formant frequencies to vocal tract area functions is developed by using sum and difference combinations of acoustic sensitivity functions to perturb an initial vocal tract configuration. First, sensitivity functions [e.g., Fant and Pauli, Proc. Speech Comm. Sem., Vol. 74 (1975)] are calculated for a neutral area function, at its specific formant frequencies. The sensitivity functions are then multiplied by scaling coefficients that are determined from the difference between a desired set of formant frequencies and those supported by the current area function. Then, the scaled sensitivity functions are summed together to generate a perturbation of the area function. This produces a new area function whose associated formant frequencies are closer to the desired values than the previous one. This process is repeated iteratively until the coefficients are equal to zero or are below a threshold value. Results show that vowel area functions are generated that are comparable to those measured with imaging techniques. The performance of the technique with time-varying formant frequencies will also be discussed. [Work supported by NIH R01-DC04789.]

4pSC24. Using phonetic constraints to improve inversion. Blaise Potard and Yves Laprie (LORIA, Campus Scientifique - BP 239, 54506 Vandoeuvre-ls-Nancy Cedex, France)

The goal of this work is to recover articulatory information from the speech signal by acoustic-to-articulatory inversion. One of the main difficulties with inversion is that the problem is under-constrained and inversion methods generally offer no guarantee on the phonetical realism of the inverse solutions. A way to adress this issue is to use additional constraints: in this work, contraints on phonetically relevant characteristics of the vocal tracts. Knowledge of the phonetic characteristics of most French vowels enables the derivation of reasonable articulatory domains in the space of Maeda parameters: given the formants frequencies ($F1$, $F2$, $F3$) of a speech sample, and thus the vowel identity, an ''ideal'' articulatory domain can be derived. The space of formants frequencies is partitioned into vowels, using either speaker-specific data or generic information on formants. Then, to each articulatory vector can be associated a phonetic score varying with the distance to the ''ideal domain'' associated with the corresponding vowel. Inversion experiments were conducted on two different speakers (male and female), using different implementations of

these phonetical constraints. The articulatory trajectories were compared with those obtained without using these constraints and to articulatory trajectories measured from x-ray data.

4pSC25. The role of principal component analysis in vowel acquisition research. Krisztina Zajdó (Univ. of Wyoming, Div. of Commun. Disord., Dept. 3311, 1000 E. Univ. Ave., Laramie, WY 82071, kzajdo@uwyo.edu), Jeannette van der Stelt, and Ton G. Wempe (Univ. of Amsterdam, Amsterdam, 1016 CG, The Netherlands)

Traditional formant analysis techniques have yielded questionable results, due to a high fundamental frequency in young children's speech. Therefore, an alternative approach for researching vowel acquisition appears desirable. A band-filtering analysis method was developed that minimizes the dependence of the results on $F0$ to measure the spectral envelopes in children's utterances automatically. In PRAAT scripts, first the criteria are set with regard to sound quality (to avoid clipped fragments) and $F0$ (to avoid vowel fragments with $F0$ above 425 Hz). Secondly, the spectral envelopes covering a range from 0 to 7 kHz, resulting in 40 filter values per spectrum, are calculated. A matrix is produced, representing intensity in each of the 40 filters. Data reduction is achieved via principal component analysis. Previous research (Pols, 1977) suggests that the first two principal components are related to the $F1$ and $F2$ values. Analyses of unlabeled vowel productions in 2-year-old Hungarian- and Dutch-speaking children show that the first two eigenvectors account for 55% of the variability. By using these two eigenvectors, a reference plane is created. Mapping labeled vowel measures onto the plane suggests that children produce language-specific vowel categories at this young age. [Work supported by UW and UvA/ACLC.]

THURSDAY AFTERNOON, 18 NOVEMBER 2004 ROYAL PALM SALONS 3 & 4, 2:00 TO 4:05 P.M.

Session 4pSP

Signal Processing in Acoustics, Underwater Acoustics, Animal Bioacoustics, Noise, Acoustical Oceanography and Engineering Acoustics: Time Delay Estimation, Localization and Tracking in Acoustics Part II

George E. Ioup, Cochair
Department of Physics, University of New Orleans, New Orleans, Louisiana 70148

Christopher O. Tiemann, Cochair
Applied Research Laboratory, University of Texas at Austin, P.O. Box 8029, Austin, Texas 78731-8029

Chair's Introduction—2:00

Contributed Papers

2:05

4pSP1. Video/acoustic-array studies of swarming by echolocating bats. James A. Simmons, Kyler M. Eastman, Gregory Auger (Dept. of Neurosci., Brown Univ., Providence, RI 02912, james_simmons@brown.edu), Michael J. O'Farrell (O'Farrell Biological Consulting, Las Vegas, NV 89131), Alan D. Grinnell (UCLA, Los Angeles, CA 94720), and Donald R. Griffin (Harvard Univ., Bedford, MA)

Echolocating bats (*Pipistrellus, Myotis*) sometimes fly in dense swarms for 1–2 h over desert water sources. Swarming is different from streaming at caves or roosts because ~10–100 bats staying within a small volume (~1000 m^3) for an extended time rather than transiting. We study swarming using thermal infrared video cameras and a three-microphone acoustic array. Swarming bats emit a steady stream of short (~2 ms) FM sonar sounds with nearly identical sweeps and a collective duty-cycle approaching 100%. Surprisingly, given the tendency to think in terms of bats in close proximity deliberately varying their signals for jamming avoidance, the signals of swarming bats are very similar. Video/acoustic tracks reveal no sign of mutual interference or collisions. It is not obvious whether bats gather to drink or to feed on insects, which also congregate near the water, or whether social behavior predominates. Some drinking occurs throughout the swarming period, but the occurrence of captures during swarming is negligible, at least judging from the absence of feeding buzzes. The bats may use multistatic sonar, in which each bat takes advantage of the sounds of other bats as helpful additional ensonification rather than interference. [Work supported by NIMH, DoD, Brown University.]

2:20

4pSP2. A generalized Cramer–Rao lower bound for line arrays. E. J. Sullivan (EJS Consultants, 46 Lawton Brook Ln., Portsmouth, RI 02871, paddypriest@aol.com) and G. S. Edelson (BAE SYSTEMS, Nashua, NH 03061-0868)

The Cramer–Rao lower bound (CRLB) on the variance of an estimate is a consequence of the underlying likelihood function. That is to say, the more realistic the likelihood function, the more realistic the bound. It is shown here that by including the forward motion of a line array in the likelihood function for the bearing of a continuous broadband signal, the CRLB on the bearing estimate is lower than that found for the case of the stationary array. This is a consequence of the fact that there is bearing information contained in the Doppler that is not exploited by conventional beamformers. Further, it is shown that this improvement in performance requires that a nuisance parameter in the form of a source frequency be jointly estimated along with the bearing. An example is given which uses a recursive processor and a synthetic broadband signal from a nonmoving source. It is shown that the performance improvement relative to a stationary array of equal physical aperture is a function of the array speed, the integration time, the bearing angle and the direction of motion of the array, i.e., whether the motion incurs ''up'' or ''down'' Doppler.

2:35

4pSP3. Cramer–Rao lower bounds applied to shallow water geoacoustic parameter estimation using the distribution of the interference invariant. Peter M. Daly and Peter N. Mikhalevsky (Sci. Applications Intl. Corp., 1710 SAIC Dr., MS 1-11-15, McLean, VA 22102, peter.m.daly@saic.com)

One can exploit the broadband acoustic striation patterns produced by loud merchant ships in shallow water to obtain geoacoustic parameters. One measures these patterns as a function of range and frequency, calculates their two dimensional Fourier transform to produce intensity as a function of wavenumber and delay, then performs an inverse Radon transform to obtain the distribution of the interference invariant. With this observable, one can perform a global inversion for parameters of interest, such as sediment and bottom sound speed, density, and attenuation. If the parameter estimate is unbiased and the observable vector Gaussian has high signal-to-noise ratio, then one can obtain the theoretical minimum variance and covariance associated with the parameters by solving for the Cramer–Rao lower bound (CRLB). In this presentation, the parameters of interest are assumed to be deterministic and observed ''noise'' is a function not only of additive Gaussian noise at the hydrophone, but also due to uncertainty in other model parameters, such as hydrophone depth and range bias. This talk will present numerically calculated CRLB as a method for determining the performance of this inversion method.

2:50

4pSP4. Localization and tracking techniques for navigation and guidance of underwater autonomous sensing systems. Hua Lee (Univ. of California, Santa Barbara, CA 93106)

The key to the successful execution of autonomous underwater imaging and sensing is the localization and tracking of the data-acquisition units, which can be achieved by estimating the relative location with respect to a set of underwater base stations. In a three-dimensional underwater acoustic localization, the objective parameters are the range distance and a two-dimensional bearing angle. At the base stations, the transmitters send a sequence of guidance waveforms. One signal is a pair of polarity signals, with a 180° left–right phase offset, for the estimation of the bearing angle in the horizontal direction. Similarly, another one is with a top–bottom polarity for the estimation of the bearing angle in the vertical direction. The received signals represent spatial-temporal samples of the interferences produced by the polarities of the transmitted signals. The range distance can be estimated from the slope of the phase profiles of the spectra. And the bearing angle is directly related to the variation of the complex amplitude of the received polarity signals and can be estimated accordingly. This paper includes system design, signal analysis, development of processing algorithms, and results from full-scale experiments. [Research is supported by the UC MICRO program and Sonatech.]

3:05–3:20 Break

3:20

4pSP5. Source localization in shallow water using artificial neural network. Gee-Pinn James Too and Eshine Lin (Dept. of Syst., and Naval Mechatronic Eng., National Cheng Kung Univ., Tainan, Taiwan 701, z8008070@email.ncku.edu.tw)

This study presents a method of passive localization for shallow water acoustic sources based on ray acoustic theory. ray acoustic theory is used to establish an underwater acoustic channel and to analyze physical quantities (arrival time and arrival angle) received by sensors, and to determine mutual relationship of rays between sources and receivers by using an artificial neural network for source localization. This ray-acoustic-based method of passive localization features simplicity, less calculation, and robust performance to environment variations. Hence, the method is more practicable and more valuable for source localization. In this study, the ray-acoustic-based method of passive localization is not only used in shallow water, but it is also verified by using a nontrained source to prove its accuracy and reliability. [Work sponsored by National Science Council.]

3:35

4pSP6. Subsurface flaw localization based on the spatio-temporal gradient array processing. Kenbu Teramoto (Dept. of Mech. Eng., SAGA Univ., Saga-shi, 840852 Japan, tera@me.sga-u.ac.jp)

Lamb waves are guided ultrasonic waves capable of propagating relatively long distances in plates and laminated structures, such as airframe skins, storage tanks and pressure vessels. Their propagation properties in these media depend on the vibrational frequency as well as on the thickness and material properties of the structure. Structural flaws such as disbonds, corrosion and fatigue cracks represent changes in effective thickness and local material properties, and therefore measurement of variations in Lamb wave propagation can be employed to assess the integrity of these structures. In this study, the localization based on the spatio-temporal gradient array processing is derived theoretically. The proposed method has an ability to characterize the phase velocity of the Lamb-wave propagation through the linearity among the 4-dimensional vector which is composed by following components: (1) a vertical (z-directional) displacement, (2) its vertical particle velocity, (3) x-directional and (4) y-directional out-of-plane strains. This paper is organized as follows: the second section presents the process of the localization and the proposed algorithm, the third section evaluates the proposed velocimetry via numerical experiments.

3:50

4pSP7. Acoustic system for aircraft detection and tracking based on passive microphones arrays. Gaetano Caronna, Ivan Roselli, Pierluigi Testa (Dept. of Tech. Phys., Faculty of Eng., Univ. of Rome ''La Sapienza,'' Via Eudossiana 18 - 00184, Rome, Italy), and Andrea Barbagelata (D'Appolonia S.p.A., 16145, Genoa, Italy)

The present study is conducted within a project included in the Sixth Framework Program of the European Union. The project, whose acronym is SAFE-AIRPORT, involves the development of an acoustic system based on two passive phased array microphone antennas, which are to be located in open environments, and capable of detecting and tracking airplanes at a distance in air and on ground. Each antenna array is made up of a series of microphones arranged on a rotating planar support. The Array's spatial selectivity is discussed, aiming to optimize beamforming design according to spectral properties of common jet aircrafts. In particular, the influence of the number and the geometric distribution of microphones is shown in function of the target's acoustic frequency and signal-to-noise ratio sensitivity. The system's range is estimated at different working frequencies and for different positions of the aircraft on the horizon. The effect of wind gradients on the estimation of the aircraft position is also modeled, and computer simulations showing consequent deviation of acoustic waves are performed. Finally, possible developments in antenna design, such as the use of three fixed planar arrays in tetrahedral configuration instead of one rotating planar array, are evaluated showing expected advantages.

Session 4pUW

Underwater Acoustics: Long-Range Acoustic Propagation Part II

Peter F. Worcester, Chair

Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0225

Invited Paper

1:30

4pUW1. 6+ years of acoustic thermometry in the North Pacific Ocean: A cycle of the Pacific decadal oscillation. Brian D. Dushaw and the NPAL Group[a] (Appl. Phys. Lab., 1013 N.E. 40th St., Seattle, WA 98105-6698, dushaw@apl.washington.eduAcoust. measurements of basin-averaged temperatures have been made in the North Pacific over the last 8 years. The acoustic observations give integral measurements of ocean temperature with the averaging needed to detect weak, gyre-scale signals out of the dominant mesoscale variability. An acoustic source on Pioneer Seamount off California transmitted to receivers distributed throughout the North Pacific from 1996 to 1999 as part of the Acoust. Thermometry of Ocean Climate (ATOC) project. A second source, located near Kauai, began transmitting in 1997. The Kauai acoustic transmissions are part of the North Pacific Acoust. Lab. (NPAL) project. A 6-year time series has been obtained with sampling six times per day every 4 days. The cabled-to-shore source and receivers provide near real-time data. Long-period changes in temperature are apparent in the time series from the Kauai source. The paths from Kauai toward the California coast show steady cooling. A path to the northwest showed warming until early 2003, when the ''normal'' annual cycle returned. Comparisons of the acoustic data to observations by altimetry and Argo floats are not definitive because of the great differences between data types. [a] (J. Colosi, WHOI; B. Cornuelle, SIO; B. Dushaw, APL-UW; M. Dzieciuch, SIO; B. Howe, APL-UW; J. Mercer, APL-UW; R. Spindel, APL-UW; P. Worcester, SIO)

Contributed Papers

1:55

4pUW2. Raytubes in an ocean acoustic environment. Matthew Dzieciuch and Bruce Cornuelle (Scripps Inst. of Oceanogr./UCSD, 9500 Gilman Dr., La Jolla, CA 92093)

In ocean acoustic tomography, sound propagation is typically modeled by ray propagation, which allows an easy answer to the question, ''Where does the sound go?'' Sound obeys the wave equation and ray propagation is only valid in the high frequency limit. It is still possible to answer the question by generating a raytube using the method outlined by Bowlin [J. Acoust. Soc. Am. **89**, 2663–2669 (1991)] for environments where full wave modeling is appropriate. It is thus possible to ask how the width of the raytube changes with frequency, bandwidth, and range. Perhaps more importantly, from an acoustical oceanography point of view, how does the environment impact the width? We investigate the effects of the strength of the internal wave field, the depth of the mixed layer, and the presence of ocean spiciness on the raytube width.

2:10

4pUW3. Sensitivity kernels for finite-frequency ocean acoustic observables. Michael Vera (College of Sci. and Technol., Univ. Southern Mississippi, 730 E. Beach Blvd., Long Beach, MS 39560), Emmanuel K. Skarsoulis (Inst. of Appl. and Computational Mathematics, 711 10 Heraklion, Crete, Greece), and Bruce D. Cornuelle (UC San Diego, La Jolla, CA 92093-0230)

Wave theoretic modeling was applied to obtain sensitivity kernels representing the amount by which ocean acoustic observables, such as pressure, intensity, or arrival times, are affected by localized sound-speed variations anywhere in the medium. These kernels can be derived using the first Born approximation, yielding an integral representation of observable perturbations as functionals of sound-speed perturbations. The kernel in each integral represents the spatial sensitivity of the observable to sound-speed perturbations in the medium. The travel-time sensitivity kernel represents a full-wave generalization of the highly localized ray kernel from geometric optics. This work extended recent wave-theoretic results for the travel-time sensitivity kernel for short-range propagation in simple environments to long-range transmissions in more complicated ocean environments relevant to ongoing propagation and inversion experiments. It was shown that the geometry of the wave-theoretic sensitivity kernels is related to the geometry of Fresnel volumes surrounding eigenrays, provided that the effects of refraction are taken into account for the calculation of the latter. [Work supported by ONR.]

2:25

4pUW4. Spice, internal waves, and sound speed in the upper ocean. Daniel L. Rudnick (Scripps Inst. of Oceanogr., UCSD, Mail Code 0213, La Jolla, CA 92093-0213)

Sound-speed variations in the upper ocean may be caused by spice fluctuations or isopycnal tilt associated with internal waves. Given finely resolved fields of temperature and salinity, the relative effects of spice and tilt on sound-speed can be ascertained. For example, in a 1000-km section in the North Pacific thermocline, root-mean-square sound-speed variability in the horizontal caused by spice is roughly one-half that caused by tilt. Sound-speed variability is almost exclusively caused by spice in the nearly unstratified mixed layer. During the June 2004 North Pacific Acoustics Laboratory (NPAL) Spice04 mooring deployment cruise, fine-scale hydrographic observations were made with the newly developed Underway CTD. Observations on a horizontal scale of 10 km and a vertical scale of 5 m show strong spice-caused sound-speed variability in the remnant of the winter-mixed layer. The largest sound-speed changes were greater than 6 m/s in a horizontal distance of tens of kilometers. The effect of these changes on acoustic propagation is a topic of continuing study.

2:40

4pUW5. Abyssal monitoring of upper ocean processes. Walter H. Munk, Matthew A. Dzieciuch, and Peter F. Worcester (Scripps Insitution of Oceanogr., Univ. of California at San Diego, La Jolla, CA 92093-0225)

The transition from the early surface-reflected to near-surface-refracted waves is associated with a triplication pattern of arrivals which is very sensitive to upper ocean processes, including mixed layer formation, spicy fronts and internal waves. The pattern is well sampled at surface-conjugate depths (3 to 5 km). Deep and relatively short vertical arrays operating in the still, uniform abyssal environment are an attractive alternative to shallow *in situ* measurements.

2:55

4pUW6. Underwater sound source with tunable resonator for ocean acoustic tomography. Andrey K. Morozov and Douglas C. Webb (Webb Res. Corp., 82 Technol. Park Dr., East Falmouth, MA 02536, moro@webbresearch.com)

A high Q tunable resonator makes possible the construction of an underwater sound source with a very high efficiency. Tuning this narrowband resonator to match the instantaneous frequency of a reference signal enables control of sound projection over a wide spectrum width. At any given time the source has one fundamental resonant frequency equal to the instant frequency of the frequency modulating signal and the sound radiates with maximum efficiency. Such sound sources find application in ocean acoustic tomography and deep-penetration seismic profiling. The analysis of the simplest known Helmholtz resonator shows its ability to instantaneously switch its frequency or to quickly adjust to a changing reference frequency. From this simplest lumped resonant source the consideration then turns to a complicated distributed system, such as a tunable organ pipe. The analysis shows that the organ pipe can also be quickly swept and instantly switch the radiated signal frequency. The 200–300-Hz swept frequency sound source was designed for ocean acoustic tomography. Test results for this sound source are presented showing that it has a high efficiency, up to 50%, 3 dBi directivity gain, and depth independent parameters.

3:10–3:30 Break

3:30

4pUW7. A simple transformation of the environment eliminates standard PE phase errors. Irina I. Rypina, Ilya A. Udovydchenkov, Michael G. Brown, and Francisco J. Beron-Vera (RSMAS/AMP, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149, irypina@rsmas.miami.edu)

It is shown that, in range-independent sound channels with a single sound speed minimum, the environment can be transformed in such a way that solutions to a standard parabolic wave equation in the transformed environment are asymptotically equivalent to solutions to the Helmholtz equation in the original environment. This equivalence is demonstrated using both ray- and mode-based wavefield expansions. Reasons that equivalence between solutions to the Helmholtz equation and the parabolic wave equation can only be asymptotic will be briefly reviewed. [Work supported by ONR and NSF.]

3:45

4pUW8. A very simple transformation of the environment significantly reduces Thomson–Chapman PE phase errors. Ilya A. Udovydchenkov, Irina I. Rypina, Michael G. Brown, and Francisco J. Beron-Vera (RSMAS/AMP, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149, ilya@rsmas.miami.edu)

In a range-independent environment with a single sound speed minimum, a transformation of the environment is available which guarantees that solutions of the standard parabolic wave equation in the transformed environment are asymptotically equivalent to solutions of the Helmholtz equation. In this paper an approximate transformation, to be used in conjunction with the Thomson–Chapman parabolic approximation, is described. Advantages of the approximate transformation over the exact transformation are its simplicity and the fact that it can be used in environments with multiple sound speed minima. Travel time errors, relative to the corresponding Helmholtz equation rays, associated with this approximation in typical deep ocean environments grow approximately linearly in range at a rate of approximately 0.1 ms/Mm. [Work supported by ONR and NSF.]

4:00

4pUW9. LOAPEX: The long-range ocean acoustic propagation experiment. James A. Mercer (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

This paper describes the recent long-range ocean acoustic propagation experiment (LOAPEX) and presents initial results. LOAPEX was conducted between 10 September and 10 October 2004, and was closely coordinated with two other experiments, BASSEX and SPICEX. LOAPEX provided low-frequency broadband acoustic transmissions to vertical hydrophone arrays covering most of the deep ocean water column, and to horizontal hydrophone arrays surrounding the Eastern North Pacific Ocean. Two of the primary objectives are to explain the strong acoustic arrivals that are received in the deep ocean shadow zone, and to better understand the roles of scattering and diffraction in general. Transmission paths to the vertical hydrophone arrays varied from 50 to 3200 km. [Work supported by the Office of Naval Research, Code 321OA.]

4:15

4pUW10. Prescription frequency-modulated signal design for the long-range ocean acoustic propagation experiment (LOAPEX). Rex K. Andrew, James A. Mercer, and Bruce M. Howe (Appl. Phys. Lab., 1013 NE 40th St., Seattle, WA 98105)

A ''prescription frequency-modulated'' signal is a swept FM signal whose sweep rate and possibly envelope are adjusted according to some *a priori* prescription. An idea advanced by Birdsall and Metzger in 1994 for the ATOC long-range tomography experiment, the concept is being revisited for LOAPEX as an alternative to ''m-sequences.'' The prescription can be used to tailor the output spectrum to a preferred shape: the sweep rate is slower at frequencies where more spectral energy is needed, and faster where less is needed. The prescription can also compensate for the transducer response. In addition, the prescription can drive the envelope to predefined maximum levels at every frequency, producing a signal which is ''optimal'' with respect to drive level. This signal is in general far less stressful on transducers than standard m-sequences. Examples of prescription FMs used during the LOAPEX experiment will be presented. [Work supported by ONR.]

4:30

4pUW11. The long-range ocean acoustic propagation experiment (LOAPEX) observable: Modal content as a function of arrival time. Part A: Relevance. Frank S. Henyey and Michael A. Wolfson (Appl. Phys. Lab., Univ of Washington, 1013 NE 40th St., Seattle, WA 92105)

It is well known that for gradual enough changes in the sound speed structure, the mode number is nearly constant. Thus, changes in mode number are the most direct measure of scattering by smaller structures, such as internal waves. The mode number changes can be determined from data by separating the signal by arrival time. It is shown how the separation happens in an ideal case without scattering, allowing an optimum choice of the time window. The modal content in a small time window can also be estimated from the depth dependence of the intensity in the region of the turning depths of the modes. Each mode has a rapid falloff near its turning depth, and the extension beyond that depth measures the higher mode content. The deep vertical line array at LOAPEX measured this depth dependence. In addition, for certain times, all of the arrival is in the deep array, allowing unambiguous projections onto modes, especially if the source is above the sound axis. [Work supported by the Office of Naval Research.]

4:45

4pUW12. The long-range ocean acoustic propagation experiment (LOAPEX) observable: Modal content as a function of arrival time. Part B. Simulation results. Michael Wolfson and Frank Henyey (APL/UW, Seattle, WA 98105)

Following the approach described in Part A, we present the acoustic mode content for fixed frequency, range, travel time, and time windowsize by performing broadband full-wave simulations using a wide-angle, parabolic equation model. We use a realistic stratification and internal wave spectrum to predict the acoustic arrival structure. The modal content is extracted using the same technique as we intend to use on the LOAPEX vertical line arrays (especially the deep array). Comparing simulations with and without internal waves reveals the additional modal spreading due to scattering.

Session 5aAA

Architectural Acoustics, Speech Communication and Committee on Standards: Speech in Architectural Spaces — Both Intelligibility and Privacy II

Kenneth P. Roy, Cochair
Armstrong World Industries, Innovation Center, 2500 Columbia Avenue, Lancaster, Pennsylvania 17604

Peter A. Mapp, Cochair
Mapp Associates, 5 Worthington Way, Colchester C03 4JZ United Kingdom

Chair's Introduction—8:45

Invited Papers

8:50

5aAA1. Design and acoustics in classrooms—a Philadelphia story. Kenneth P. Roy and Kenneth W. Good, Jr. (Armstrong Innovation Ctr., 2500 Columbia Ave., Lancaster, PA 17604)

The Lamberton School in Philadelphia was built in 1949 and may have met all of the acoustical design requirements of the day . . . but! An acoustic and A/V evaluation of a typical classroom was made both before and after an architectural intervention to address the ANSI S12.60 Classroom Standard. See the numbers and hear the difference, both for yourself, and through the opinions of the teacher who has experienced the change. What was done and the effects thereof will be discussed.

9:10

5aAA2. Recent experience with the voice intelligibility recommendations for fire alarm systems in large athletic spaces. Matthew Moore (Cavanaugh Tocci Assoc., 327 F Boston Post Rd., Sudbury, MA 01776, mmoore@cavtocci.com)

Beginning with the NFPA 72 National Fire Alarm Code, 1999 edition, there has been a recommendation for voice intelligibility with fire alarm systems. The code appendix references a Common Intelligibility Score (CIS) of 0.70 or greater. This paper discusses how the design and construction process for some recent sports facility projects has been affected, and some of the difficulties with meeting this score. Example projects and measurements will be used to illustrate the challenges in these unique environments.

9:30

5aAA3. Computer model studies to predict qualitative and quantitative measures of speech intelligibility in class rooms. Hyeong-seok Kim, Gary Siebein (School of Architecture, Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611-5702), Brian Kreisman, and Carl Crandell (Univ. of Florida, Dauer Hall, Gainesville, FL 32611)

Computer model studies were used to predict qualitative and quantitative measures of speech intelligibility in classrooms under realistic conditions of background noise and reverberation. Fifteen different acoustical measurements related to speech intelligibility were made at multiple locations in three actual classrooms and in computer models of the classrooms. Speech intelligibility (MRT) tests were given to human subjects in each of the actual classrooms at five signal-to-noise ratios. Speech intelligibility tests were also prepared from aural simulations obtained by convolving anechoic speech tracts with impulse responses obtained in the computer models. Correlations ($R2$) between acoustical measures made in the full size classrooms and the computer models of the classrooms of 0.92 to 0.99 with standard errors of 0.033 to 7.311 were found. The scores on the speech intelligibility tests given in the actual rooms in the five noise conditions were closely duplicated in the equivalent tests conducted in a sound booth using the simulated speech signals obtained in the computer models. Both quantitative and qualitative measures of speech intelligibility in the actual rooms were accurately predicted in the computer models.

9:50

5aAA4. A verification of computer modeling predictions of speech intelligibility. Erica E. Bowden, Jonathan Rathsam, and Lily M. Wang (Architectural Eng. Program, Univ. of Nebraska Lincoln, 245 PKI, 1110 S. 67th St., Omaha, NE 68182-0681, ebowden@mail.unomaha.edu)

Computer modeling has become a popular tool in the field of architectural acoustics to predict the acoustical characteristics of a built environment. One characteristic commonly investigated in computer models is speech intelligibility. This project compares speech intelligibility measures of the speech transmission index (STI), the rapid speech transmission index (RaSTI), and the articulation loss of consonants (ALCons) predicted by computer models to measurements taken in existing spaces. Three spaces in the Omaha, NE, area of varying size and function are analyzed: a theater, church, and classroom. Research methodology and conclusions are presented. The results aid in understanding the reliability of 3D computer modeling in speech intelligibility predictions. [Work partially supported by the Univ. of Nebraska Center for Building Integration.]

10:10–10:25 Break

10:25

5aAA5. Some common error mechanisms in making STI intelligibility measurements. Peter A. Mapp (Peter Mapp Assoc., Copford, Colchester, Essex, UK)

STI and its derivatives (RaSTI and STIPa) have become the internationally accepted methods for acoustically measuring the potential intelligibility performance of both sound systems and rooms. Their use has unquestionably brought about a significant improvement in public address/voice alarm system quality and intelligibility. However, in practice, many of the measurements made on site may be unwittingly based on flawed techniques. The paper examines a number of common problems found to affect measurement accuracy. Significant differences between different measurement platforms and techniques are reported. Furthermore, different stimulus formats and even their recording and playback medium are also shown to potentially affect the final result. It is shown that neither STI nor STIPa, in their current formats, accurately predicts the intelligibility of sound systems with irregular frequency responses, particularly when these are operating in reverberant, high signal-to-noise environments, a common effect found in real-world voice alarm and emergency sound systems.

10:45

5aAA6. Prediction of speech intelligibility in rooms—a comparison of five methods. Murray Hodgson (School of Occ. & Env. Hygiene, Univ. of BC, 3rd Fl., 2206 East Mall, Vancouver, BC V6T 1Z3, Canada, hodgson@mech.ubc.ca)

In this paper, five methods for calculating speech intelligibility at listener positions in rooms, from measured impulse responses and speech and noise levels, are presented and compared. All methods involve combining a measure of room reverberation with the received signal-to-noise level differences in the occupied room. The first method involves measured octave-band C50s and signal-to-noise level differences to determine octave-band U50s. The remaining four methods comprise detailed (involving octave-band values) and simplified (involving single- and/or combined-frequency values) versions of two approaches. The first approach (second and third methods) involves measured background-noise levels and values of the Transmission Index measured for infinite signal-to-noise level difference. The second approach (fourth and fifth methods) involves 1000-Hz octave-band early-decay times and A-weighted signal-to-noise level differences. The five methods are briefly presented. Then, predictions by the methods for a number of classrooms are presented and compared, and the differences, the merits, and the disadvantages of the various methods are discussed.

11:05

5aAA7. Comparison of three methods for determining speech transmission index (STI). Robert C. Coffeen (School of Architecture and Urban Design, Marvin Hall, The Univ. of Kansas, Lawrence, KS 66045)

Speech transmission index (STI) is a commonly used descriptor whose goal is to quantify speech intelligibility. STI can be calculated from computer modeling of a particular architectural space, calculated from an impulse response of an existing space, and measured by introducing modulated noise in an existing space. This paper will present calculated and measured STI ratings for a particular and existing architectural space so that these three methods for determining STI can be compared.

11:25

5aAA8. A comparison of computational models for predicting speech intelligibility and speech privacy. Ralph T. Muehleisen (Civil and Architectural Eng., Illinois Inst. of Technol., Chicago, IL, 60616, muehleisen@iit.edu) and C. Walter BeamerIV (Univ. of Colorado, Boulder, CO 80309)

Speech intelligibility (SI) and speech privacy (SP) are very important considerations in the design of classrooms, auditoria, conference rooms, and offices. Predictions of SI and SP are usually done with simple analytic models using Sabine/Eyring theory or CAD programs such as CATT, EASE or ODEON. In this paper we will compare the predictions of speech intelligibility index (SII), speech transmission index (STI), and privacy index (PI) using analytic theories of Sabine/Eyring and a simplified acoustic radiosity theory, a simple computational model using acoustic radiosity and CAD prediction using CATT acoustics. Predictions will be compared with the results of a similar study by Bistafa and Bradley [J. Acoust. Soc. Am. **109**(4), 1474–1482 (2001)].

Contributed Paper

11:45

5aAA9. Acoustical retrofit in residential buildings. Ballard W. George (Environtech Consultants, 1367 Bobolink Circle, Sunnyvale, CA 94087, kingstaco@yahoo.com)

This paper deals with situations where for some reasons problems arise, and one has to apply noise control measures after or near the end of construction. Three cases are discussed. In two of these, recommendations for window acoustical performance were not implemented. In one case, window improvement was needed to achieve consistency with documentation given to the city and to meet design goals. The owner elected to use the option of improved windows combined with sound absorption in terms of a high-performing carpet. In another case, again window recommendations were not followed, and a curtain wall with Lexan lights also contributed to excessive exterior-to-interior noise levels up to approximately 10 dB. A spaced-out gypsum board inner wall layer arrangement, with ventilation openings, was developed by the architect. The proposed mitigation arrangement involved trade-offs regarding acoustical separation and privacy between units. The third case involved a new hotel, where kitchen-serving equipment, not shown on plans provided, created rumble that resulted in complaints regarding one guest unit. A menu of corrective measures was provided, which could be implemented in sequential fashion.

Session 5aAB

Animal Bioacoustics: Animal Bioacoustics General Topics

Ann E. Bowles, Chair
Hubbs Sea World Research Institute, 2595 Ingraham Street, San Diego, California 92109

Chair's Introduction—8:25

Contributed Papers

8:30

5aAB1. An investigation of hearing threshold shift in bats due to blast wave pressure. Larry Pater (ERDC/CERL, 2902 Farber Dr., Champaign, IL 61821, larry.l.pater@erdc.usace.army.mil), Elizabeth Brittan-Powell, Cynthia Moss, Robert Dooling (Univ. of Maryland, College Park, MD 20742), Alexandra Loubeau, and Victor Sparrow (The Penn State Univ., University Park, PA 16802)

A project is described to determine damage to chiropteran (bat) hearing by shock waves such as those emitted by weapons and explosions. Experimental and analytical means will be used to accurately define the stimulus, particularly as regards wave form and spectra in the frequency range 10–100 kHz in which bat hearing is sensitive. Initial results regarding the stimulus as well as hearing thresholds determined via the auditory brainstem response are presented. [Work supported by US Army Engineer Research and Development Center CERL.]

8:45

5aAB2. Acoustic adaptation hypothesis in macro- and micro environments: An analysis of frog calls. Yuan Yao (Dept. of Organistic Biol., Ecology and Evolution, UCLA, Los Angeles, CA 90095) and Ying Lin (UCLA, Los Angeles, CA 90095)

The acoustic adaptation hypothesis states that acoustic signals are structured so as to maximize their performance under the constraints of the environmental acoustics that characterize their native habitats. However, less attention has been paid to the channels in which acoustic communication takes place and the specific constraints that different channels put on acoustic signals. For frog calls, the channel characteristics depend on two factors: the macro environment—such as the habitat; and the micro environment—such as the sites of the sender and receiver. The data in our analysis came from recordings of 81 species of frogs in French Guyana. These species span a variety of macro environments, such as forest, coastal marsh, and savanna, and a variety of micro environments, such as on vegetation, on ground, and under litter. Two methods—spectrogram correlation and hidden markov models—are used to derive a similarity matrix between these calls. The results indicate that calls show a higher similarity within the same macro environment than between different macro environments. On the other hand, measurements of spectral and temporal features suggest that the micro environment also plays a role, while its effect is more salient for frequency components than for time components.

9:00

5aAB3. Statistical automatic species identification of microchiroptera from echolocation calls: Lessons learned from human automatic speech recognition. Mark D. Skowronski and John G. Harris (Computational NeuroEng. Lab, Elec. and Computer Eng., Univ. Florida, Gainesville, FL, markskow@cnel.ufl.edu)

Current automatic species identification of microchiroptera through echolocation search calls utilizes holistic features, such as call duration, bandwidth, and frequency extrema, which stem from expert knowledge of sonograms of individual calls. This approach parallels the early acoustic-phonetic paradigm of human automatic speech recognition (ASR), which relied heavily on expert knowledge to account for variations in canonical linguistic units of interest. Acoustic-phonetic ASR gave way to the statistical paradigm of ASR primarily because of the superior ability of machine learning to account for variations in the acoustic signal (noise, speaker characteristics) [Juang and Furui, Proc. IEEE **88** (8), 1142–1165 (2000)]. In the current work, machine learning methods from human ASR, hidden Markov models (HMM) and dynamic time warping (DTW), are considered for the problem of automatic species identification of microchiroptera from echolocation calls. The extensive history of ASR provides valuable lessons, which are highlighted and applied to species recognition from field-recorded acoustic data. Frame-based features of the acoustic signal (fundamental frequency, log energy, temporal derivatives) are incorporated into HMM and DTW classifiers, and experimental results indicate the superior approach of statistical automatic species identification compared to current techniques.

9:15

5aAB4. Automatic song-type classification and individual identification of the ortolan bunting (*Emberiza hortulana L*) bird vocalizations. Kuntoro Adi (Speech and Signal Processing Lab, Marquette Univ., Milwaukee, WI 53233), Tomasz S. Osiejuk (Adam Mickiewicz Univ. PL-61-485 Poznan, Poland), and Michael T. Johnson (Marquette Univ., Milwaukee, WI 53233)

This paper presents a method for automatic song-type classification and individual identification of the ortolan bunting (*emberiza hortulana L*). This method is based on hidden markov models (HMMs) commonly used in the signal processing and automatic speech recognition research communities. The features used for classification and identification include both fundamental frequency and spectral characteristics. Fundamental frequency features consist of center frequency $f0$, relative $f0$ to moving-average baseline, peak strength, and peak bandwidth. Spectral features are derived from frequency-weighted cepstral coefficients. Using these features one HMM is trained for each type of vocalization both for each individual bird and across the entire population. Preliminary results indicate accuracies of above 90% for both song-type classification and individual identification tasks.

9:30

5aAB5. Comparison and evaluation of animal vocalization enhancement techniques. Jidong Tao and Michael T. Johnson (Dept. of Elec. and Comput. Eng., Marquette Univ., Haggerty Hall, Milwaukee, WI 53201)

Animal vocalization recordings are often corrupted by wideband background noise and interfering signals; however, signal processing methods for enhancement of these waveforms has not received as much attention in the literature as has human speech enhancement. In order to improve vocalization intelligibility and quality, a variety of enhancement methods taken from the field of speech processing are investigated here. These techniques range from traditional approaches such as spectral subtraction to more advanced ones such as Ephraim Malah log-spectral estimation and

wavelet denoising. Results, measured by improvement in signal-to-noise ratio (SNR) and subjective perceptual tests, are given for several noise environments. Signal enhancement is demonstrated across a variety of species, including African elephant, Beluga whale, and ortolan bunting vocalizations. [Work supported by NSF.]

9:45

5aAB6. Classification of natural landmark with biosonar. Maosen Wang and Andreas Zell (RA Dept., WSI, Univ. of Tuebingen, Sand 1, Tuebingen, Germany)

Echolocating bats can make nocturnal flights in acoustically cluttered environments with the use of echolocation. Their marvelous ability to evaluate natural targets in complete darkness provides us an opportunity to learn target detection, classification, and identification with similar biomimetic platforms. In this work, natural landmark classification with a binaural system, a sequential sensing strategy, and a frequency after reconstruction algorithm are adopted to provide sequential acoustic images for target classification. Experimental results suggest that considerable improvements in classification accuracy can be achieved by the use of this sequential classification method.

10:00

5aAB7. Vocalizations of *Equus caballus*: Frequency analysis of horse whinnies. David G. Browning (Dept. of Phys., Univ. of Rhode Island, Kingston, RI 02881, decibeldb@aol.com) and Peter M. Scheifele (Univ. of Connecticut, Storrs, CT 06269)

There are six recognized horse sounds: scream, squeal, nicker, whinny, snort, and blow. Of these whinnies are the most interesting with a high enough source level for communication, perhaps being the horse equivalent of a bark, and a characteristic change in frequency content during the vocalization. This frequency variation is not observed in many farm animal vocalizations, cows, goats, and sheep, for example. An analysis is made of the variation in typical horse whinnies in order to identify any general characteristics as well as any specific variations that can be identified with specific behavior.

10:15

5aAB8. Diel periodicity of fish sound production in Charlotte Harbor, Florida. James Locascio (Univ. of South Florida College of Marine Sci., 140 Seventh Ave. South, St. Petersburg, FL 33701) and David Mann (Univ. of South Florida College of Marine Sci., St. Petersburg, FL 33701)

Diel periodicity in sound production of spawning aggregations of fishes was documented in Charlotte Harbor, Florida from 7 May 2003–10 June 2003. The Long Term Acoustic Recording System (LARS) recorded 10 s of sound every 10 min within the frequency range of 0–1250 Hz. Field data collected in this study revealed diel patterns in fish sound production in great detail. Autocorrelation results demonstrated pronounced diel periodicity in fish sound production with significant lags occurring at 24-h periods. Chorusing events dominated by sand seatrout, *Cynoscion arenarius*, lasted for several hours each night and were not highly variable in duration, maximum recorded SPL, or start and end times. Mean daily (24-h period) maximum sound-pressure level was 119 decibels, mean daily chorus start time was 17:26 h Eastern Standard Time, mean daily chorus end time was 02:10 h Eastern Standard Time, and the average nightly chorus duration lasted 8 h and 43 min.

FRIDAY MORNING, 19 NOVEMBER 2004 ROYAL PALM SALONS 1 & 2, 9:00 TO 11:55 A.M.

Session 5aNS

Noise: Soundscapes and Sound Quality

Brigitte Schulte-Fortkamp, Cochair
Institute of Technical Acoustics, Technical University Berlin, Einsteinufer 25, 10587 Berlin, Germany

Patricia Davies, Cochair
School of Mechanical Engineering, Purdue University, Ray W. Herrick Laboratories, 140 South Intramural Drive, West Lafayette, Indiana 47907-2031

Chair's Introduction—9:00

Invited Papers

9:05

5aNS1. Perception of noise from diesel-engine powered vehicles. Aaron Hastings (Ray W. Herrick Lab., School of Mech. Eng., Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47907-2031), Patricia Davies, and Aimee M. Surprenant (Purdue Univ., West Lafayette, IN 47907)

A program was developed to simulate diesel engine sounds to have control over timing and amplitude variations in combustion events. The simulation includes this combustion-related noise, together with a noise floor and families of tones from auxilliary components. Additional shaping filters to control the spectral balance of the sounds are also included. Sound and vibration measurements from ten vehicles powered by diesel engines were taken and combustion event timing and amplitudes were extracted to determine what ranges and types of variations typically occurred. It was found that there were random, deterministic and individual cylinder variations present in one or more of the vehicles. From this analysis, the simulation program was used to generate 199 sounds for a listening test in which 40 subjects participated. Spectral balance, tonalness, and, fluctuations all affected the subjects' responses

and, above certain levels, as they increased so did subjects' annoyance ratings. Based on the results of this test, modifications to the psychoacoustic annoyance model [Zwicker and Fastl, *Psychoacoustics: Facts and Models*, 2nd ed. (Springer, New York 1998)] are proposed. These modifications include the incorporation of a term that quantifies how the tonal character of the sound affects annoyance. [Work supported by Isuzu Motor Company.]

9:25

5aNS2. Characterizing short-time transient events in the information technology soundscape. Wade R. Bray (HEAD acoustics, Inc., 6964 Kensington Rd., Brighton, MI 48116)

Information technology (IT) products form an increasing part of modern life in residential, recreational, and work contexts. Although average levels and sound power are generally low and have been improving for years, many IT products and components produce transient sounds which draw attention and elicit customer complaints. Standardized characterization by average sound power, with reference to tonal and impulsive characteristics, may not adequately capture subjectively important time-varying behavior. Time/frequency, time/magnitude, impulsiveness analysis methods, and an adaptive relative pattern-recognition technique will be applied to the short-time transient noises produced by a variety of information technology devices.

9:45

5aNS3. Sound quality in environment: "Psychoacoustic mapping." Klaus Genuit (HEAD acoustics GmbH, Ebertstrasse 30a, 52134 Herzogenrath, Germany, klaus.genuit@head-acoustics.de)

The annoyance due to noise in the environment is often predicted by analysis software which calculates the A-weighted sound pressure level distribution depending on the sound sources and on the sound propagation caused by them. The actual subjective perceived noise annoyance of the environment does not only depend on the A-weighted sound pressure level, but also on the so-called sound quality along with other parameters like the aspects of a soundscape. The question regarding noise annoyance cannot be predicted by the A-weighted sound pressure level alone. The sound quality perceived by the human hearing depends on, among other things, loudness, roughness and sharpness. The previously known methods for the prediction of the spatial A-weighted sound pressure level distribution based on the propagation are not suitable to predict psycho-acoustic parameters in an adequate way. Especially the roughness caused by modulation or the sharpness generated by the contribution of high frequency sound energy at various distances cannot readily be predicted. In the following it will be shown what the requirements are and especially what the challenges are for making such a psychoacoustic mapping.

10:05–10:20 Break

10:20

5aNS4. Community reaction to noise from a replaced bridge span and associated approach. Paul R. Donavan (Illingworth & Rodkin, Inc., 505 Petaluma Blvd. South, Petaluma, CA 94952, pdonavan@illingworthrodkin.com) and Bruce Rymer (California Dept. of Transportation, Sacramento, CA 95814)

A newly constructed replacement of the westbound span of the Carquinez Bridge on Interstate I-80 in the San Francisco Bay Area was recently opened for service. Although the traffic volume and mix remained the same as it was before the project, there was a very strong reaction to the noise produced by the new construction from nearby residences. For these residences, the predicted noise levels were only expected to increase by less than 2 dB due to the bridge and its approaches being somewhat closer to the community. The complaints had two aspects: generally higher traffic noise levels and impulsive slaps from the expansion joints in the viaduct connecting the bridge deck to the at-grade roadway. Although clearly audible in the community, the joint slaps did not contribute to any of the common community traffic noise metrics. For noise mitigation, the viaduct surface was ground to reduce overall tire/pavement noise, however, there was concern that by reducing the masking effect of the pavement texture generated noise, the slaps from joints would become even more noticeable and objectionable. The before and after noise measurements and community response are presented and discussed in this presentation.

10:40

5aNS5. Explorative sound evaluation. Stephan Paul, Brigitte Schulte-Fortkamp (ITA/TU-Berlin, Einsteinufer 25 D-10587, Berlin, Germany), and Klaus Genuit (HEAD-acoustics, Herzogenrath, Germany)

Exposure group description of sound based on AISP (exploration of associated imaginations on sound perception) is a powerful tool in sound quality evaluation. Used in a public research project it was stated that AISP needs further adaptation for automotive use. The present study was intended as a first approach refining the method for industrial application regarding the evaluation of an appropriate test design. The study was carried out using a driving simulator and a car, both with changeable acoustic properties. Verbal evaluations by the driving test individuals and the sound were recorded. Later on an open interview with the driver was carried out based upon given commentaries. The verbal data were analyzed through a qualitative approach. By means of this data the car with changeable acoustic properties driven on road was identified as most appropriate for sound evaluation of car interior sounds. The verbal data analysis and the physical analysis of the sounds gave valuable hints for troubleshooting and target sound definition. As the study showed the strong dependency of sound evaluation on the test design a representative evaluation should be done in an adequate design (car on road). Furthermore, initial usable tools and guidelines for industrial application were developed.

11:00

5aNS6. Ambient noise in Johns Hopkins Hospital. Ilene J. Busch-Vishniac, James E. West, Colin Barnhill, and Tyrone Hunter (Dept. of Mech. Eng. and Dept. of Elec. and Comput. Eng., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218)

Although noise in hospitals is widely recognized as a problem, there has been little work to characterize hospital noise reported in the literature. From the literature that does exist, it is clear that noise in hospitals poses problems for patient safety, increases staff stress, and is a leading source of complaints by patients and visitors. It is also clear that few hospitals currently meet the recommended noise limits of the World Health Organization or the US EPA. In this paper we present the first round of results of a noise study at Johns Hopkins Hospital in Baltimore, MD. Some interesting characteristics of the noise include its high level, its unusually flat spectrum through the speech range, and its relative invariability with location.

11:20

5aNS7. Natural and urban soundscapes: The need for multi-disciplined research. Robert Kull (Parsons, 5800 Lake Wright Dr., Ste. 101, Norfolk, VA 23502), Brigitte Schulte-Fortkamp, Sebastian Rossberg (Tech. Univ. Berlin, Berlin Germany) and Tim Lavallee (P.E. LPES, Inc., Smithfield, VA 23430)

Characterizing soundscapes represents a broad variety of approaches (ecological, sociological, phenomenological) in a continuum from completely natural to highly urban environments. There is an urgent need for the use of innovative designs which integrate the different levels of current analyses (qualitative and quantitative; individual and aggregate levels). The criteria (beyond sound level) of a good soundscape (What is a sensitive soundscape? What are soundscape requirements for a resort area or natural quiet?), that should be protected are central to the origin of the soundscape idea. Other classical acoustics questions, such as the use of audibility intrusiveness under critical conditions and under condition of mixed sources and time pattern, should be asked differently to deepen insight and understanding. A first attempt was made to monitor acoustical changes in an Italian mountain resort: Classical indicators (L95, L5, Leq) were used to locate geographical areas of increases in intrusiveness of the soundscape and linked to the activity pattern of the area. Several projects were conducted to define and localize quiet areas for a nature conservation program. Descriptive and analytical techniques, such as participatory sound and listening walks, cognitive maps, acoustical spectrographic maps, and soundscapegraphy are new techniques to consider for multi-disciplined approaches in different environments.

Contributed Paper

11:40

5aNS8. Queuing for quiet−The natural soundscape microstructure from a visitor perspective. Richard D. Horonjeff (Richard D. Horonjeff Consultant in Acoust. and Noise Control, 81 Liberty Square Rd. #20-B, Boxborough, MA 01719) and Grant S. Anderson (Harris Miller Miller & Hanson Inc., Bultington, MA 01803)

The passage of PL100-91 The National Parks Overflights Act of 1987 and PL106-181 The National Parks Air Tour Management Act chartered the National Park Service (NPS) and Federal Aviation Administration (FAA) to restore natural quiet to park settings, especially those with lengthy occurrences of nonindigenous sounds. Dose−response studies sponsored by NPS and FAA showed that among several dependent variables tested, nonindigenous time audible was the best predictor of visitor-reported annoyance and interference with natural quiet. Other than for Grand Canyon NP, PL100-91 lacked specificity for depicting nonindigenous times audible. Since 1991, several field investigations have produced over 400 h of source identification logs maintained by audiometrically screened observers. At over 25 sites in 7 parks, the data acquisition protocol used a forced-choice, hierarchical menu structure with audible source state changes timed to the nearest second. The identical protocol across all studies provided the basis for examining the amount of time a visitor would have to wait (T) to experience a contiguous block of indigenous-only sounds of duration (D), given a random arrival time. The paper reports means and distributions in T for various values of D, observed data clusterings across site groups, and the effects of mitigating variables.

Session 5aSC

Speech Communication: Boundaries, Rhythm and Timing in Spoken Language (Poster Session)

Amalia Arvaniti, Chair
Linguistics Department, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0108

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 9:00 a.m. to 10:30 a.m. and contributors of even-numbered papers will be at their posters from 10:30 a.m. to 12:00 noon.

5aSC1. Relating the performance and perception of phrasal boundaries. Jelena Krivokapic (Dept. of Linguist., USC, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693, krivokap@usc.edu)

This study examines the correspondence between the production of prosodic structure and perceptual judgments regarding prosodic boundary strength. In the production component of the study, three speakers read sentences in which one specific juncture is manipulated (varying in syntax, position in sentence, and phrase length) to elicit phrasal boundaries of differing strengths. Lengthening and pause duration at the boundary were measured. The same sentences, in written form, were presented to a second different group of three speakers who provided estimates of the strength of the target boundary on a scale of eight degrees. The results of the production and the estimation portions of the experiment demonstrate significant correlations between the production boundary strength as reflected in durational properties at the juncture and the boundary strength as estimated in the judging task. These correlations are roughly linear. Further, in both the production and perception domains, a range of boundary strength is exhibited rather than a small discrete set of boundary types. We also examine whether speaker's own boundary strength estimates agree with their productions to a greater extent than estimates of other speakers. [Work supported by NIH.]

5aSC2. On the temporal scope of boundary effects in articulation. Dani Byrd, Jelena Krivokapic (Dept. of Linguist., USC, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693, dbyrd@usc.edu), and Sungbok Lee (USC, Los Angeles, CA 90089)

Boundary-adjacent acoustic lengthening is well explored, and the articulatory bases for this lengthening are becoming better understood. However, the temporal scope of boundary effects has not been examined in the articulatory domain. The few acoustic studies examining the distribution of lengthening indicate that boundary effects extend from one to three syllables leftward from the boundary, and that effects diminish as distance from the boundary decreases (Cambier-Langeveld, 1997; Shattuck-Hufnagel and Turk, 1998; Turk, 1999; Berkovits 1993a,b, 1994). This diminishment is predicted by the pi-gesture model of prosodic influence (Byrd and Saltzman, 2003). We present an experiment testing the leftward and rightward scope of articulatory lengthening. One condition (CV#C1VC2VC3V) tested the scope of effects after the boundary, and another (C3VC2VC1V#CV) the scope preceding the boundary (where # indicates an intonational phrase boundary) as compared to a no-boundary control. Movement-tracking (EMA) data allowed the evaluation of constriction formation and release duration, acceleration, and spatial magnitude. Further, FDA (functional data analysis) was used to examine continuous temporal warping properties. Results of both analyses indicate an asymmetrical distribution of boundary effects, in that leftward effects are shorter in scope than rightward effects, though effects in both directions exist around a phrase boundary. [Work supported by NIH.]

5aSC3. Interacting effects of phrasal and syllable position on consonant production. Sungbok Lee (Dept. of Linguist., USC, and USC Viterbi School of Eng., 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693, sungbokl@usc.edu), Dani Byrd (USC, Los Angeles, CA 90089), Jason Adams (USC, Los Angeles, CA 90089), and Daylen Riggs (USC, Los Angeles, CA 90089)

The complexities of how prosodic structure, both at the phrasal and syllable levels, shapes speech production have begun to be illuminated through studies of articulatory behavior. Here, we pursue the goal of understanding prosodic signatures on articulation by examining the effects of phrasal and syllable position on the constriction formation and release of consonants. Articulatory kinematic data were collected for five subjects using electromagnetic articulography (EMA) to record target consonants (labial, labiodental, and tongue tip), located in (1) either syllable final or initial position and (2) either at a phrase edge or phrase-medially: #C, ##C, C#, C##. The duration, displacement, and time-to-peak velocity of constriction formation and release were determined for the target consonants based on kinematic landmarks in the articulator velocity profiles (zero crossings and extrema). ANOVA determines that syllable and phrasal position consistently affect the movement duration; however, effects on displacement were more variable. For the majority of subjects, the boundary-adjacent portion of the movement (release for a preboundary coda and constriction formation for a postboundary onset) is not differentially affected in terms of phrasal lengthening; both lengthen equivalently. However, two subjects show an interaction such that the codas are lengthened more than the onsets. [Work supported by NIH.]

5aSC4. Towards standard measures of articulatory timing. Leonardo Oliveira, Mariko Yanagawa, Louis Goldstein (Yale Univ. & Haskins Labs., 270 Crown St., New Haven, CT 06511), and Ioana Chitoran (Dartmouth College, Hanover, NH 03755)

Many studies have investigated the relative timing (or overlap or latency) between articulatory events to make inferences about coordination in speech production or about phonetic structure. A major problem is that these studies employ different measures of relative timing. One criterion for choosing a standard measure could be the variability exhibited across repetitions and talkers for a given set of gestures in a given phonological context. If a single measure appears to be the least variable across a

variety of experiments, then it is a good index of the stable, phonetically relevant properties of coordination. In this study, 33 measures were compared using coefficient of variation. In four corpora of articulatory data on consonant clusters (EMMA and x-ray microbeam), three articulatory events were defined: gesture onset, target, and release. Based on these events, coefficients of variation for the 33 measures were calculated across repetitions and speakers under a number of different linguistic conditions, i.e., stress, syllable, rate, cluster type and language (English and Georgian). Despite the very different properties of the corpora, the least-variable measures were always very similar and involve normalizing the latency between the two gestures by the constriction formation time (target to onset). [Work supported by NIH.]

5aSC5. Phrase-final frication lengthening in a variety of English spoken in the Iron Range of Northern Minnesota. Matt Bauer (Dept. of Linguist., Georgetown Univ., 37th and O St., Washington, DC 20057)

This study examines acoustic and perceptual correlates of final-devoicing in a variety of English spoken throughout the Iron Range of Northern Minnesota. A pilot study of one speaker indicates devoicing in the dialect is present phrase-finally but not phrase-medially $x2(1) = 0.89$, $p < 0.05$. Measurements of voicing and frication in words like tens and tense reveal devoicing is not due to shortening of the duration of vocal fold vibration (which is what a devoicing process predicts). Instead, frication of phrase-final fricatives are lengthened compared to frication of phrase-medial fricatives $F(2,27) = 17.36$, $p < 0.05$, but vocal fold vibration phrase-finally is unaffected. Thus, the percept of devoicing results from a lengthening process that does not alter voicing duration. The present study will determine the extent to which final-lengthening is present in the speech of other speakers. The project is interesting because (a) Iron Range English was previously thought to exhibit devoicing, not final-lengthening [G. Underwood. Pub. Am. Dialect Society, 67 (1982)] and (b) results address how representations of sound are transferred into speech unfolding in time. Intuitions offered by articulatory phonology indicate timing of phrase-final frication lengthening in Iron Range English is not a reflex of a phonological rule, but instead of differentiated timing of independent gestures.

5aSC6. Final lowering: Fact, artifact or dialectal variation? Amalia Arvaniti (Dept. of Linguist., UCSD, 9500 Gilman Dr. #0108, La Jolla, CA 92093 amalia@ling.ucsd.edu)

The existence of final lowering has been disputed by Grabe (1998), who suggested that Liberman and Pierrehumbert's 1984 finding was due to declination, as there was an extra syllable, ''and,'' between the last two accents in their materials (e.g. ''RASPberries, MULberries and BRAMBLEberries''). Arvaniti and Godjevac (2003), however, replicated the results of L&P with and without the extra syllable (c.f. ''LIma beans, NAvy beans and SOY beans'' vs. ''LIma beans, GREEN beans and SOY beans''). Here, the materials used in A&G were elicited from speakers of Standard British English (SBE). The results confirmed that the disagreement between L&P and Grabe is due to dialect: the SBE speakers did not exhibit final lowering under either condition. They showed less steep $F0$ slopes, so all accents after the first had similar scaling, while the last accent showed a rise (rather than a fall, as in American English), enhancing the lack of final lowering. These differences show that final lowering is better seen as a phonological device that a linguistic variety may employ in a given melody. Also, the cross-varietal comparison supports the view that $F0$ downtrends consist of distinct components (declination, final lowering), each of which may be utilized independently of the others.

5aSC7. Optical phonetics and visual perception of lexical and phrasal boundaries in English. Edward T. Auer, Jr. (House Ear Inst., 2100 West Third St., Los Angeles, CA 90057), Sahyang Kim, Patricia Keating, Rebecca A. Scarborough, Abeer Alwaan (UCLA, Los Angeles, CA 90095-1543), and Lynne E. Bernstein (House Ear Inst., Los Angeles, CA 90057)

Detection of lexical and phrasal boundaries in the speech signal is crucial to successful comprehension. Several suprasegmental acoustic cues have been identified for boundary detection in speech (e.g., stress pattern, duration, and pitch). However, the corresponding optical cues have not been described in detail, and it is not known to which optical cues to boundaries speechreaders are attuned. In a production study, three male American English talkers spoke two repetitions each of eight pairs of sentences in two boundary conditions (one-word versus two-word sequences, one-phrase versus two-phrase sequences). Sentence pairs were constructed such that they differed minimally in the presence of a boundary. Audio, video, and 3D-movements of the face were recorded. Sentence pairs in both boundary conditions differed on both optical and acoustic duration measurements. A subset of the sentence pairs was also presented in a visual perception study. Pairs were chosen *a priori* because they differed or did not differ on the production measures. Fourteen perceivers detected boundary presence from video-alone in a forced-choice paradigm. Both lexical and phrase boundaries were reliably detectable by perceivers. The relation of visual intelligibility to the optical characteristics of production will be discussed. [Research supported by NSF IIS 9996088 and NIH DC04856.]

5aSC8. Pitch/amplitude mismatches in Mandarin. Deborah S. Davison (Res. Compliance, Stanford Univ., 1215A Welch Rd., MC 5401, Stanford, CA 94305)

Pitch and amplitude contours covary in Beijing Mandarin [Sagart (1986)]. Tianjin Mandarin amplitude range of stressed low (L) phonological tones is equal to that of stressed high (H) tones for $>50\%$ of L tokens. $F0$ range for 75–85 dB lexical tonemes T1L and T3LH versus T4HL was 100–158 Hz versus 135–180 Hz, respectively. Dimensions of $f0$/amplitude range for T1L 75–150 Hz/70–85 dB versus T4's H 110–180 Hz/70–85 dB were virtually identical: 75 Hz/15 dB versus 70 Hz/15 dB, respectively. Stressed T1L and T3LH low $f0$'s relatively higher amplitude may compensate for reduced perceptual saliency of low $f0$. T2''H'''s [Yip (2002)] pattern resembles T4H assuming T2's single H tonal target aligns to right rather than left syllable edge (cf. Beijing lexical T2 ''mid-to-high rising tone'') when fully stressed, hence bimoraic. Tonally unspecified left edge T2 $f0$ does not correlate with amplitude. Both monomoraic ''toneless-stressless'' neutral tone T0 syllables preceding stressed T1L and T3LH and lexical T1L and T3LH ''sandhi-changed'' to LH and H preceding stressed T1L and T3LH, respectively, have high $f0$ but lower amplitude than the following stressed low tone. Stress-conditioned $F0$/dB mismatches support the conclusion that high $f0$/dB attracts pitch accent [Davison (2003)], contrary to the assumption that lexical tone languages such as Chinese do not permit phrasal pitch accents internal to intonational phrases [Ladd (1996)].

5aSC9. Form and function of intonational phrase heads. Cynthia Girand (Dept. of Linguist., UCB 295, Univ. of Colorado, Boulder, CO 80309)

This study examines the phonetic and phonological features essential to the investigation of the internal structure of the intonational phrase. This is crucial for identifying and explaining the forms and functions of basic intonational units. There are two primary intonational contour modeling theories [Ladd (1983) and references therein]. One suggests that phrasal contours are the basic units of intonation. Within this theory, contour shapes are associated with particular functions or meanings. In contrast, a more recent theory claims that individual tones (i.e., abstract phonological units) are the basic units of intonation, and intonational contours result from the concatenation of adjacent tones in a phrase. Using 756 utterances

from the Switchboard and Buckeye corpora, the present study takes a closer look at the basic units that compose the intonational contour. While the nucleus has long been identified as a functionally important part of an intonational phrase, the head of an intonational phrase has not been considered, in kind. This work examines global and local phonetic and phonological features (e.g., intensity, pitch range, pitch height, downstepped tones) of intonational phrase heads in an attempt to better understand their forms and functions.

5aSC10. Italian Raddoppiamento: Prosodic effects on length. Rebeka Campos-Astorkiza (Univ. of Southern California, University Park GFS 301, Los Angeles, CA 90089, rebekaca@usc.edu)

Traditionally, Italian Raddoppiamento refers to a lengthening process that targets word-initial consonants after a final stressed vowel. Unfortunately, most theoretical accounts of Raddoppiamento lack a solid phonetic foundation. We report an acoustic study, based on the data obtained from four Tuscan Italian native speakers, investigating both word-initial consonants and vowels in the Raddoppiamento environment. Further, we consider two different prosodic positions, C Raddoppiamento phrase-internally versus across an intonational phrase boundary, which, according to previous analyses (Nespor and Vogel, 1986), should prevent the lengthening from occurring. Finally, stressed and unstressed environments are tested. Thus, the quality and magnitude of Italian Raddoppiamento lengthening as a function of segmental identity, prosodic context, and stress is reported. The results show that consonantal lengthening takes place as expected in the traditional Raddoppiamento environment. On the other hand, word-initial vowels do not lengthen. Stress has an effect on word-final vowel length that is incompatible with previous theoretical accounts of Raddoppiamento. Finally, the presence of an intonational phrase boundary does not, as previously predicted, categorically block the process. Overall, this empirical evidence challenges previous accounts of Raddoppiamento and provides a systematic phonetic documentation of the phenomenon. [Work supported by NIH.]

5aSC11. Children's use of prosody to identify ambiguous sets of compound nouns. Michiko Yoshida and William F. Katz (Univ. Texas—Dallas, Callier Ctr. for Commun. Disord., 1966 Inwood Rd., Dallas, TX, 75235, michiko@utdallas.edu)

Research has shown that young children can process word duration and fundamental frequency ($F0$) information in an adultlike manner to disambiguate simple conjoined English phrases [Beach *et al.*, J. Acoust. Soc. Am. **99**, 1148–1160 (1996)]. However, the processing of pause duration has not been investigated in this manner. The present study investigates how children use pause duration and $F0$ to perceive word boundaries, and whether the manner of information processing differs between children and adults. Adults and children (5 and 7 years old) completed a picture-pointing task based on auditory stimuli. Stimuli were computed-edited tokens of the words "sun," "flower," and "pot." From a recording by a male talker, vocoded tokens were created with five steps of pause duration and three steps of fundamental frequency. Both continua ranged between patterns suggesting the interpretations "sun, flowerpot" and "sunflower, pot." Preliminary results indicate a cue-trading relation between duration and $F0$ for all listeners, with pause duration playing the dominant role. These data suggest adultlike processing in children as young as 5 years old. Experiments with younger children and a formal evaluation of the pattern recognition processes used by listeners are currently underway in our laboratory.

5aSC12. Perception of stress in Thai. Rattima Nitisaroj (Dept. of Linguist., Georgetown Univ., 37th and O St., NW, Washington, DC 20057)

This study assesses the relative contribution of four parameters—duration, amplitude, fundamental frequency and vowel quality—to the perception of stress in Thai. Previous acoustic studies [e.g., S. Potisuk *et al.*, Phonetica **53**, 200–220 (1996)] have identified these four parameters as varying with stress, but perceptual tests in which these parameters are independently controlled have been needed. For this experiment, stimuli were created by digitally manipulating the four parameters in minimal pairs (compounds versus phrases) that differ only in stress pattern. Subjects were asked to choose between two contexts signaling whether the stimuli they heard contained a compound or a phrase token. Duration was found to be the strongest perceptual cue to stress. Sometimes, stress was signaled by the combination of duration and vowel quality. Results also showed an effect of subjects bias to judge noun–verb tokens as phrases and noun–noun tokens as compounds. If the bias conflicted with decision signaled by duration and vowel quality, effects caused by amplitude variation were observed. Finally, fundamental frequency did not appear to play any role in signifying stress. The results thus support the following hierarchy of perceptual correlates to stress in Thai: duration, vowel quality/bias, amplitude, and fundamental frequency.

5aSC13. Comparing rhythm and melody in speech and music: The case of English and French. Aniruddh D. Patel, John R. Iversen, and Jason C. Rosenberg (The Neurosci. Inst., 10640 John Jay Hopkins Dr., San Diego, CA 92121)

Does the prosody of a nation's language leave an imprint on its music? We address this question by comparing British English and French, a stress-timed vs syllable-timed language with salient intonational differences. We previously showed that an empirical difference between speech rhythm in the two cultures is reflected in instrumental music. In this study we expand on these rhythmic measurements and provide new data on melody. We compare English and French intonation using a measure which can also be applied to musical melodies. In a database of read speech we converted the intonation contour of each sentence into a sequence of vowel pitches, using the mean fundamental frequency of each vowel to represent its pitch. We found that the size of pitch intervals between successive vowels varied more in British English than in French speech. We then examined classical instrumental music and found that pitch intervals between successive notes varied more in English than in French music. We also examined differences in the way rhythm and melody are aligned in both speech and music. Overall, we find that the prosody of a culture's language is reflected in the structure of its instrumental music. [Supported by Neurosciences Research Foundation.]

5aSC14. Acoustic modifications in choral reading. Meredith A. Poore and Sarah H. Ferguson (Dept. of Speech-Lang.-Hearing, Univ. of Kansas, 3001 Dole Ctr., 1000 Sunnyside Ave., Lawrence, KS 66045, mpoore@ku.edu)

Choral reading is the condition in which one or more talkers read an assigned text aloud and in synchrony. Choral reading has several practical and theoretical applications, including use with beginning readers to enhance oral literacy, with people who stutter to evoke fluency, and as a tool for investigating speech rhythm. However, little is known about the speech acoustic modifications that occur when talkers perform the task of choral reading, or how different methods of eliciting choral reading may affect these modifications. In stuttering therapy, choral reading is usually elicited in pairs, either with two live talkers (choral reading) or one talker live and one prerecorded (track reading). While previous studies have assumed that these two reading conditions are equal, a recent study on speech timing suggests otherwise [F. Cummins, ARLO **3**, 7–11 (2002)]. Measurements of pitch, amplitude and vowel duration will be compared across solo, track and choral reading of texts which vary in their perceived rhythmic organization. Acoustic data will be interpreted in reference to various theories regarding the effectiveness of choral reading for fluency enhancement. [Work supported by a University of Kansas Honors Program Undergraduate Research Award.]

Session 5aUW

Underwater Acoustics: Localization, Classification and Processing

Harry A. DeFerrari, Chair
RSMAS Applied Marine Physics, University of Miami, 4600 Rickenbacker Causeway, Miama, Florida 33149

Contributed Papers

8:00

5aUW1. Multi-array passive search in a littoral environment. Donald R. DelBalzo, Erik R. Rike, and David N. McNeal (Neptune Sci., Inc., 40201 Hwy. 190 E, Slidell, LA 70461 delbalzo@neptunesci.com)

Acoustic barrier tactics were developed during the Cold War for deep, uniform underwater environments. Oceanographic and acoustic conditions in littoral environments are extremely complex and dynamic. The spatial and temporal variability of low-frequency signal and noise fields in these complex environments destroys the basic homogeneous assumption associated with standard tactical search concepts. Genetic algorithms (GAs) have been applied to both the signal and noise parts of this problem to produce near-optimal, nonstandard search tracks that maximize probability of detection in such inhomogeneous noise fields. In the present work, a GA was used to optimize tactics of several passive searchers by constructing optimal barriers in complex, littoral environments. The dynamic ambient noise model (DANM) was used to produce realistic, low-frequency, directional noise fields, based on discrete ship tracks. The cost function to be minimized was the probability that a target crossed the barrier undetected. Both standard and GA-derived tactics were evaluated and compared. The results show the importance of nonstandard tactics and careful consideration of environmental complexity when designing optimal passive search tactics. [Work sponsored by NAVSEA under the LCS project.]

8:15

5aUW2. Beamformer waveguide invariant source localization on the New Jersey shelf (USA) during winter acoustic propagation conditions—RAGS03. Altan Turgut, Bruce H. Pasewark, Marshall H. Orr (Naval Res. Lab., Acoust. Div., Washington, DC 20375), and Daniel Rouseff (Univ. of Washington, Seattle, WA 98105)

Horizontal and vertical beamformer (waveguide) invariants are used to robustly localize broadband noise sources (ships) using signals in the 50–200-Hz band. The localization is realized by using signatures recorded on one horizontal and three vertical arrays. In brief, waveguide invariant theory applied to beamforming by two vertical arrays provides a range ratio of the source to the receivers. Beamforming by a horizontal array provides time-evolving spectrum for a particular look direction (LOFARgram). As a result the trajectories of the striations observed in LOFARgrams can be used to estimate the speed, range and direction of a broadband source. In December 2003 the Naval Research Laboratory moored three vertical arrays and one horizontal array at ranges of 10, 20 and 30 km from two fixed acoustic sources on the New Jersey (USA) shelf. Acoustic signatures from passing merchant and research vessels were recorded on the arrays and used to demonstrate source localization. [Work supported by ONR.]

8:30

5aUW3. A new invariant method for instantaneous source range estimation in an ocean waveguide from passive beam-time intensity data. Sunwoong Lee and Nicholas C. Makris (MIT, 77 Massachusetts Ave., Cambridge, MA 02139, makris@mit.edu)

A new method is derived for instantaneous source range estimation in an ocean waveguide from passive beam-time intensity data received on a horizontal or vertical line array. While the method is as simple and robust as the ''waveguide invariant'' method described by Brekhovskikh and Lysonov (2003), it does differ significantly from the latter. Some advantages are that (1) it only requires source signal measurement at a single rather than many ranges so that range can be instantaneously estimated, and (2) it is invariant over all horizontally stratified ocean waveguides. The paramter β of the waveguide invariant method, on the other hand, varies significantly with vertical sound-speed structure. Since the invariant parameter of this new method varies only with the geometry of the array, the term ''array invariant'' method is here introduced. The method does not rely on near-field processing techniques but is also applicable at arbitrary far-field ranges. The method is applied to data from the Main Acoustic Clutter Experiment 2003 for source ranges between 2 to 10 km, where it is shown that simple, accurate, and efficient source range estimates can be made using this new method.

8:45

5aUW4. Source localization in environments with deterministic and stochastic uncertainties. Ralph N. Baer and Michael D. Collins (Naval Res. Lab., Washington, DC 20375)

Focalization is an approach for localizing an acoustic source when there are uncertainties in the sound speed or other environmental parameters [J. Acoust. Soc. Am. **90**, 1410–1422 (1991)]. The original work on focalization was based on deterministic parameters, but the concept was recently applied to problems involving uncertainties associated with internal waves [J. Acoust. Soc. Am. **115**, 2550 (2004)]. Due to a parameter hierarchy in which source position outranks environmental parameters, it is often possible to localize a source without determining the true environmental parameters. This is a very fortunate situation, especially when dealing with stochastic parameters that would be essentially impossible to determine. Focalization is currently being extended to problems involving uncertainties in both deterministic and stochastic parameters. [Work sponsored by ONR.]

9:00

5aUW5. Optimal sensor placement in highly variable acoustic fields to increase detection. Erik R. Rike and Donald R. DelBalzo (Neptune Sci., Inc., 40201 Hwy. 190 E, Slidell, LA 70461)

Optimal placement of sensors for detection depends on spatial and temporal properties of ambient noise (AN) and transmission loss (TL). These acoustic quantities are highly variable in littoral environments and their predictions are often overly smoothed when estimating detection probabilities. This variability was characterized using the dynamic ambient noise model (DANM) and the parabolic equation (PE) model. DANM was used to produce space and time estimates of directional noise based on discrete ship tracks and the PE was used to characterize spatial variability in TL. Then, Brown's algorithm for search planning was modified to include realistic ordered search against diffuse target distributions and used to determine the relative importance of non-homogeneous noise statistics and TL in optimizing search tactics. The new capability is COBRA (coherent optimization for Baysean resource allocation), which is efficient and accurate. The results in one littoral area show that spatial variations in

TL are twice as important as those in AN for optimal search planning. This work has immediate application to design of optimal sonobuoy field patterns and future application to mobile sonar search planning. [This work was sponsored by ONR under the LADC project.]

9:15

5aUW6. Horizontal array gain variability measured on the New Jersey shelf during late fall and early winter propagation conditions—RAGS03. Bruce H. Pasewark, Altan Turgut, Marshall H. Orr, Jeffery A. Schindall, Michael McCord, and Earl Carey (Naval Res. Lab., Code 7120, Washington, DC 20375-5350)

A conventional beamformer has been applied to 300 Hz CW and 300 Hz center frequency LFM acoustic signals received on a 96-channel 465 m bottomed horizontal array. The temporal history of the array gain for five subapertures (30, 60, 120, 240, 465 m) will be presented. The gain of each subaperture varied over time scales ranging from less than 1 min to 13 days. The time dependence is being correlated to sound speed variability induced by shelf slope front movement, the internal tide, nonlinear internal waves and atmospheric forcing. The longer apertures did not achieve ideal array gain. Temporal variability of the horizontal spatial coherence lengths will be discussed. The acoustic sources were moored 18 m above the bottom in 64 m of water. The propagation path was cross shelf. The receiving array was located near the New Jersey (USA) shelf break 20 km from the source. Water depth at the array location was 89 m. [Work supported by ONR.]

9:30

5aUW7. Hybrid joint probability density functions for active sonar. James M. Gelb and Brian R. La Cour (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78758)

Methods are presented for estimating joint probability density functions (PDFs) of statistically dependent features with a focus on sparse data with planned application for computing likelihood functions for feature-based target classification in active sonar. The estimators involve a new class of hybrid models, i.e., adjusted combinations of low-dimensional, nonparametric PDFs and high-dimensional, multivariate, parametric models. One goal is to preserve physically meaningful feature identity. The efficacy of the methods to model PDFs and to classify data sets using these PDFs will be presented for simulated and actual data. Two general forms for the PDFs are presented: (1) hybrid models, $\Pi_i p_i(f_i) \times [M_N(\mathbf{f})/\Pi_i \mathbf{M}_i(\mathbf{f}_i)]$, where $\mathbf{f}$ represents a set of N features; $p_i(f_i)$ are the marginal probabilities, and $M_N(\mathbf{f})$ is a model for the N-dimensional multivariate PDF with model marginals $M_i(f_i)$. (2) Expansion models: the multivariate PDF is expanded in terms of its correlations, $\Pi_i p_i(f_i)[1 + \Sigma_{i<j}\xi_{ij} + \Sigma_{i<j<k}\zeta_{ijk} + \cdots]$, with, for example, the two-point correlation functions defined as $\xi_{ij} = p_{ij}(f_i, f_j)/p_i(f_i)p_j(f_j) - 1$. In both cases successes and failures of the models and necessary modifications to them are presented to improve their general utility. [This work is funded by the ONR.]

9:45

5aUW8. Object classification using low frequency scattering characteristics obtained from structural acoustic target models. Alessandra Tesei and Mario Zampolli (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19138 La Spezia, Italy)

The capability to distinguish between different objects using low frequency scattered signals is a topic of fundamental interest to the sonar research community. To address this issue, a 3-D frequency domain finite element structural acoustics tool (FESTA) is used to synthesize multistatic scattered field time series from frequency sweeps. The time-frequency plots for scattering from a variety of manmade objects, like elastic spherical shells and cylindrical shells of various types, are analyzed, and the characteristic elastic effects are exploited to obtain low frequency classification clues. Comparisons between FESTA and other models are shown, and where available, experimental data are used to validate the results. Guidelines for the optimal level of modeling accuracy required to capture the relevant scattering details using the finite element tool are established, and the tradeoffs between accuracy, computation time and memory use are addressed.

10:00–10:15 Break

10:15

5aUW9. M-sequences and bi-static active sonar. Harry DeFerrari and Hien Hguyen (RSMAS, Univ. of Miami, FL)

The application of m-sequences to bi-static active sonar is evaluated with data from acoustic propagation experiments in the Florida Straits. Over the past 20 years, several numerical methods for processing M-sequences have been developed by Birdsall, Metzger and others. These methods come together for a possible sonar application by the following approach: (1) continuous transmission and reception of long m-sequences, (2) synchronous sampling to form a CON (complete ortho-normal) data set, (3) direct blast removal by HCCO (hyperspace cancellation by coordinate zeroing), and (4) a full range waveform Doppler search. Ultra-fast Hadamard transforms speed up the direct waveform pulse m-sequence pulse compression and the inverse pulse waveform transform and thereby allow timely execution of the intensive computational burden. The result is a numerically demonstrable approach that produces a gain of 36 dB over a simple pulse and 16 dB over other active signals. In addition, the direct arrivals and their Doppler leakage are eliminated, thus shifting the detection problem from a reverberation limited to a noise limited Doppler process. Here, m-sequence data from propagation experiments are reprocessed to test the fundamental concepts. Data results are shown to be in close agreement with ideal numerical simulations.

10:30

5aUW10. Multisensor registration for distributed active sonar systems. Brian R. La Cour, Kevin Johnson, and Son Quach (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

A technique is described whereby discrepancies between and biases within active sonar receivers may be corrected using acoustic signals. The method utilizes information contained in the direct blast recorded on each receiver over multiple active pings to estimate the ranges and bearings of the one or more sources relative to each receiver. Each sensor is assumed to have limited beamforming capability. In addition, the use of known acoustic scatters to resolve sensor misalignment will be discussed. Finally, results of an application of these techniques to active sonar data from the Deployable Experimental Undersea Multistatic Undersea Surveillance system (DEMUS) will be discussed and used to assess the algorithm. [This work was funded under ONR Contract No. N00014-00-G-0450-21.]

10:45

5aUW11. Target detection using a beam space time-reversal operator. Geoffrey F. Edelmann, Joseph F. Lingevitch, David M. Fromm, Charles F. Gaumond (Naval Res. Lab., 4555 Overlook Ave., Washington DC 20375, geoff@ccs.nrl.navy.mil), and David C. Calvo (Naval Res. Lab., Washington DC 20375)

Time-reversal operator methods such has DORT have recently garnered interest as promising techniques to ensonify ocean targets. However, measuring the time-reversal operator in the ocean by independently firing each element of a time-reversal mirror is difficult due to low signal-to-noise levels. This talk will discuss a portion of the shallow-water TREX-04 experiment in which energy was focused on a target using a vertical array of 64 source/receive transducers. The time-reversal operator was measured by transmitting energy from the full aperture of the time-reversal mirror instead of from just a single element. The time-reversal operator, made in beam space, was then used to successfully focus energy on a target 1 km away. [Work supported by the Office of Naval Research.]

11:00

5aUW12. An underwater acoustic detection experiment based on forward scattering in a time reversal mirror. J. Mark Stevenson, Alessandra Tesei, Piero Guerrini, Pierangiolo Boni (NATO Undersea Res. Ctr., La Spezia, Italy), Philippe Roux, Heechun Song, William S. Hodgkiss, William A. Kuperman, and Tuncay Akal (Scripps Inst. of Oceanogr., La Jolla, CA)

Following earlier work by Song *et al.* [IEEE, J. Ocean. Eng. **28**, (2003)] we conducted an underwater acoustic barrier experiment based upon forward scattering in a time reversal mirror. The experiment was conducted at a carrier frequency of 15 kHz in a shallow water waveguide. The waveguide length was about 16 water depths with one interface being an undulating, soft seafloor. After establishing a focus at one end of the waveguide, stationary and moving objects were placed in the water column, including an autonomous undersea vehicle (AUV). Before and after comparison of the received acoustic signal, including the temporal and spatial abberation of the focus, provided object detection. Several anomaly detection techniques were applied to the acoustic data in post-processing. Target speed was observed to be a discriminator in anomaly detection performance, suggesting that a combination of different anomaly detection techniques is probably desirable in addressing the moving target problem. [Work supported by NATO and ONR.]

11:15

5aUW13. Examination of bit error characteristics due to a drifting source in passive phase conjugation underwater acoustic communications. Jong R. Yoon, Kyu-Chil Park (Div. of Electron., Comput. and Information Eng., Pukyong Natl. Univ., Korea 608-737, jryoon@pknu.ac.kr), and Daniel Rouseff (Univ. of Washington, Seattle, WA 98105)

Recent experimental work in acoustic communications using passive phase conjugation has shown that the demodulation error depends on the relative drift rate between the transmitter and receiver [Rouseff *et al.*, IEEE J. Ocean Eng. **26**, 821–831 (2001)]. The observed effect involves the mismatch between the initial impulse response and the subsequent response after the source or receiver has changed locations. The result can also be interpreted in terms of the spot size of the retro-focused field. In the present work, the effect of relative drift between source and receiver is studied by numerical simulations. The communications bit error rate is quantified as a function of drift rate, carrier frequency, array geometry and source-receiver range. [This work was supported by Pukyong National University Research Abroad Fund in 2003.]

11:30

5aUW14. Focal depth shifting of a time reversal mirror in a range-independent waveguide. Shane C. Walker, Philippe Roux, and William A. Kuperman (Marine Phys. Lab., S.I.O., UCSD, 9500 Gilman Dr., Mail Code 0238, La Jolla, CA 92039-0238, shane@physics.ucsd.edu)

A time reversal mirror refocuses back at the original probe source position. A goal has been to refocus at different positions without model based calculations. A method to refocus at different ranges has been developed earlier [Song *et al.*, J. Acoust. Soc. Am. **103**, 3224–3240 (1998)] using frequency shifting. Here we present a technique to refocus at different depths than the original probe source in a shallow ocean range-independent waveguide. The requirement is to collect data from various ranges at a single depth, as from a moving broadband radiator, over a distance sufficient to construct the relevant frequency-wavenumber structure of the waveguide. With this information, it is then possible to focus at arbitrary depth at any of the ranges that the probe source data were taken. Theory and laboratory measurements are presented.

FRIDAY AFTERNOON, 19 NOVEMBER 2004 PACIFIC SALON 3, 1:00 TO 4:45 P.M.

Session 5pUW

Underwater Acoustics: Noise, Sound Sources and Sonar Systems

David L. Bradley, Chair

Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

Contributed Papers

1:00

5pUW1. Underwater sound pressures from marine pile driving. Richard B. Rodkin and James A. Reyff (Illingworth & Rodkin, Inc., 505 Petaluma Blvd. South, Petaluma, CA 94952, rrodkin@illingworthrodkin.com)

Marine pile driving of steel shell piles has resulted in high underwater sound pressures that have been lethal to fish. High sound pressures have resulted in harassment of pinnepeds under the marine mammal protection act. Most waterways in the nation include fish and marine mammals that are protected by State and/or Federal agencies. Impacts from pile driving have contributed to costly construction delays for some major bridge projects. Recent construction activities in the marine environments of Northern California have provided the opportunity to characterize these sound pressures and evaluate control measures to protect fish and marine mammals. Sound control measures evaluated include different pile-driving methods, cofferdams (with and without water), confined air bubble curtain systems, and unconfined bubble curtain systems. Some control measures have achieved over 30 dB of noise reduction. Each situation, however, presents difficulties in achieving targeted reduction goals.

1:15

5pUW2. Sounds and vibrations recorded on and beneath landfast sea ice during construction of an artificial gravel island for oil production in the Alaska Beaufort Sea. Charles R. Greene, Jr. (Greeneridge Sci. Inc., 1411 Firestone Rd., Goleta, CA 93117)

Underwater and airborne sounds and ice-borne vibrations were recorded from sea-ice near an artificial gravel island during its initial construction. Recordings were made in winter 2000 when BP began constructing Northstar Island about 5 km offshore. Underwater and airborne sounds were recorded over frequencies from 10 to 10 000 Hz; particle motion in the ice was recorded from 10 to 500 Hz. Ice thickness was about 1.6 m.

Water depth at the island was 12 m, and depths ranged between 4 and 13 m at recording stations. Recording distances were from about 100 m to over 5 km. Activities recorded included ice augering, pumping sea water to flood the ice and build an ice road, a dozer plowing snow, cutting ice with a Ditch Witch, trucks hauling gravel over an ice road to the island site, and a backhoe trenching the sea bottom for a pipeline. Dominant frequencies were generally below 500 Hz. The Ditch Witch cutting ice was one of the stronger sounds and was audible underwater and as ice vibration to distance 3 km. Other sounds diminished to background levels by a distance of 2 km. Airborne sound detection depended on wind speed and direction. [Work supported by BP.]

1:30

5pUW3. Issues for improved characterization of the acoustic field due to radiated noise of ships in shallow water environments. Christopher Barber and Kyle M. Becker (Appl. Res. Lab., Penn. State Univ., P.O. Box 30, State College, PA 16804-0030)

Estimation of the sound pressure levels radiated from ships to underwater sensors in littoral environments is a problem of increasing importance. Current areas of interest range from underwater surveillance and homeland security of ports, to the threat posed to naval vessels by influence mines. Recent literature addressing shallow water propagation has focused primarily on object detection using active sonar systems or on environmental characterization through inverse methods. While conclusions from these studies regarding the impact of environmental variability on acoustic propagation may be generally applicable, there are several additional areas that need to be investigated in order to understand the acoustic field at short ranges in shallow water. This presentation seeks to identify areas where further investigation may be warranted. In particular, existing propagation codes are examined for their applicability and/or limitations for the modeling of near-field radiation from nonsimple distributed sources, evaluation of complex source signals containing both broadband noise and narrow-band tones over a broad frequency range, and the impact of small-scale local variability. This presentation seeks to motivate a discussion toward new approaches and methodologies for addressing gaps between modeling requirements and capabilities. [Work supported by PEO-LMW, AIMS Program.]

1:45

5pUW4. Modeling of angular and range acoustic energy distribution from seismic exploration array in the Northern Gulf of Mexico. Arslan Tashmukhambetov, Natalia A. Sidorovskaia (Phys. Dept., Univ. of Louisiana at Lafayette, UL BOX 44210, Lafayette, LA 70504), George E. Ioup, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

Broad-band experimental recordings (up to 25 kHz) of the acoustic energy arrivals from surface seismic exploration array by a bottom-moored hydrophone were gathered by the Littoral Acoustic Demonstration Center (LADC) in the Northern Gulf of Mexico during the summer of 2003. Experimental data are available for a wide range of horizontal distances (up to 7 km) and arrival angles. The numerical reconstruction of the acoustic radiation pattern and the acoustic energy decay law in range subject to the actual propagation conditions in the Northern Gulf of Mexico based on the calibrated recorded data by the identification of the zero-to-peak value in the direct pulse arrival is presented. The comparison of the measured energy losses with the expected ones due to spherical and cylindrical spreading is given. The identification of different types of arrivals (direct pulse, surface ghost, bottom reflected, etc.) and evaluation of their frequency content depending on the azimuth and emission angles and the distance to the array are conducted. The complexity of propagation and source models required to predict the identified arrivals is discussed.

2:00

5pUW5. Noise characterization at the international station of Crozet Islands (H04). Antoine Roueff, Pierre-Franck Piserchia, Jean-Louis Plantet, Yves Cansi, and Gerard Ruzie (CEA/DASE/LDG, BP12, 91680 Bruyeres-le-Chatel, France)

In the context of the Comprehensive Nuclear-Test-Ban Treaty (CTBT), the hydroacoustic network gives the ability to state parties to check the natural and artificial activity in the oceans. To achieve this purpose, 11 hydroacoustic stations are going to be established. One of these stations has recently been launched by the CEA (France) near the Crozet Islands in the Indian Ocean. In this paper, we characterize the ambient noise of this new station. This study will enable one to estimate the detectability of the station. In this analysis, we associate the tide height with the noise level on the recorded signals. The lunar calendar perfectly matches the hydroacoustic noise variation. In addition, we will show the influence of this noise evolution on the rate of detection. In order to have a good statistic, we process a hundred days of data (recorded at 250 Hz). The tide heights supplied by ''le Laboratoire de gophysique et Ocanographique de Toulouse,'' are sampled with a 1-h step. Furthermore, the spectral characteristic of the ambient noise is also given. Finally, a comparison with other international stations, Diego Garcia (H08) and Cape Leeuwin (H01), is given.

2:15

5pUW6. A swept-frequency implosive source. LeRoy M. Dorman and Allan W. Sauter (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0220)

Implosive underwater sound sources, such as the 20-liter single-shot device we have constructed, generally emit a rarefaction followed by a compression when the implosion is complete. Some purposes, however, are better served by a swept-frequency, rather than an impulsive, signature, since the lower peak pressures of the swept source are less likely to cause environmental damage. Our current plan for a repeating implosive source operates by venting a bubble (initially at ambient pressure) into a low-pressure receiver through a control valve. Performing this process in steps with abrupt pauses induces oscillations characteristic of the several sizes of the evolving bubble. From modeling the bubble oscillations using the Gilmore bubble equation, it appears feasible to produce a range of a factor of 2 in frequency, with a peak pressure of a megapascal. The design of the control valve, however, is challenging, as it was for the single-shot version. The repeater, however, operates on air, whose density and viscosity are lower.

2:30

5pUW7. An airgun array source signature model for environmental impact assessments. Alexander O. MacGillivray, N. Ross Chapman (School of Earth and Ocean Sci., Univ. of Victoria, P.O. Box 3055 STN CSC, Victoria, BC V8W 3P6, Canada, alexm@uvic.ca), and David E. Hannay (JASCO Res. Ltd., Victoria, BC V8Z 7X8, Canada)

Environmental impact assessments for seismic surveys often include estimates of radially propagating sound levels, which are used to determine marine mammal impact zones. Sound levels may be estimated using computer-based acoustic modelling techniques but these require an accurate description of the survey source signature—arrays of airguns, in particular, have complex, highly directional source functions that depend on the array layout. To address this requirement, an airgun array source signature model has been developed for the purpose of underwater noise level prediction. The source model is based on published descriptions of the physics of airgun bubble oscillations and radiation [A. Ziolkowski, Geophys. J. R. Astron. Soc. **21**, 137–161 (1970)] and includes the effects of port throttling, motion damping and bubble interactions. The output of the model is a collection of notional signatures which may be used to compute the source function of the array in any direction. Free parameters in the model have been fit to a large collection of existing airgun signature data, for airguns ranging from 5 to 185 in^3. The output of the model is suitable for estimating sound levels resulting from airgun survey activity. [Work supported by NSERC.]

2:45

5pUW8. Measured and predicted array response to the vertical directivity of ambient noise. Andrew Holden (DSTL Winfrith, Dorchester, DT2 8WX, UK, apholden@dstl.gov.uk)

A great deal has been published on ambient noise. Most of this has covered (a) omni directional levels, and (b) the vertical and horizontal directivity of shipping noise at low frequencies. There is some published material on the vertical directivity of wind generated noise at lower frequencies, but very little at higher frequencies. In order to study wind generated ambient noise at higher frequencies, a small planar array from QinetiQ Bincleaves has been used to make ambient noise measurements. As well as measurements, a model called CANARY has been written to predict ambient noise vertical directivity and array responses to this noise. This paper contains some comparisons between CANARY predictions and (a) previously measured array response to ambient noise vertical directivity at 4.5 kHz, and (b) analysis of the planar array measurements at 25 kHz. The paper shows the nature of the ambient noise vertical structure and that the CANARY predictions are in good agreement with the measurements.

3:00–3:15 Break

3:15

5pUW9. Noise audit model for acoustic vector sensor arrays on an ocean glider. Michael Traweek (Office of Naval Res., 800 N. Quincy St., Arlington, VA 22217-5600), John Polcari (AETC, Inc., Arlington, VA 22202), and David Trivett (Georgia Inst. of Technol., Atlanta, GA 30332-0405)

The design of any new sensing array is a complex endeavor, demanding clear understanding of both signal exploitation options and the underlying noise environment (including ambient, structural, flow and electronic components), including cross-channel characteristics. The goal of this effort is to outline an approach for assessing potential advantages of vector sensors to provide ocean gliders with the acoustic sensory input necessary to execute ISR, characterize the environment, and support other mission-enhancing behaviors. At the heart of the approach is development of a noise audit model (NAM) which (coupled with appropriate signal characterizations) enables realistic and comparable evaluation of optimal processing performance for different classes of sensors (e.g., scalar, vector, or tensor). The NAM is a design tool that permits balanced design of the various array components so that the array output is ambient noise limited in the quietest operating environment. Noise components are broken into different source-transmission path chains; correct transfer functions are applied along each path; and sensor/array outputs are then obtained by incoherent sum. Properly structured, rapid evaluation of performance limits can also be supported. The approach will be discussed, and examples of both NAM components and associated performance evaluations will be presented. [Work supported by ONR.]

3:30

5pUW10. Simulation of synthetic aperture sonar performance under experimentally measured environmental fluctuations. Timothy H. Ruppel and Steve Stanic (Naval Res. Lab., Stennis Space Ctr., MS 39529)

In June 2003, NRL conducted a series of coherence measurements in very shallow water (8 m depth) off the coast of Panama City, FL under a wide range of ocean conditions. Low-frequency (1–10 kHz) and high-frequency (10–200 kHz) measurements were taken. The results show appreciable variability even during fairly calm conditions. This paper will report on the effects of this experimentally measured variability on the performance of a simple computer-simulated synthetic aperture sonar (SAS). SAS works by rapidly and repeatedly pinging a target from a moving acoustic array platform. If the underwater acoustic environment changes appreciably between pings, random phase errors are introduced on the individual SAS receiving elements, with the result that the SAS image might be significantly degraded. The use of a computer-modeled SAS allows for the investigation of these effects in the absence of such phenomena as random receiver motion (motion compensation) or specific auto-focusing algorithms. [Work supported by the ONR Program Element No. 62435N.]

3:45

5pUW11. Multibeam sonar observations of hydrodynamic forcing functions and bubble persistence in a ship wake. Thomas C. Weber, Anthony P. Lyons, and David L. Bradley (Appl. Res. Lab. and the Grad. Prog. in Acoust., P.O. Box 30, State College, PA 16804, tcw141@psu.edu)

High frequency (240 kHz) upward-looking multibeam echosounder data were collected from the wake of a twin-screw 50-m research vessel. Volumetric backscatter in each of the 101 1.5×1.5-deg beams was recorded at a repetition rate of approximately 10 s as the ship passed over the moored sonar system. The multibeam sonar's wide field of view and the low wake drift rate made it possible to image the wake over its entire lifetime of 30 min. The wake quickly separates into distinct port and starboard components. It is suggested that this is caused by vortices shed from the hull that propagate orthogonally to the ship's direction, entraining bubbles and creating a convergence zone on the outboard side of the vortices as they oppose the buoyant rise of the bubbles. The speed at which the vortices separate is calculated from the data and used in a simple model describing the evolution of the bubbles in the wake. Results show that the hull shedded vortex model is plausible, and that the gas transfer rate from the bubbles must be approximately 25 times less than it would be for surfactant-free bubbles in order to explain the bubble persistence observed in the data.

4:00

5pUW12. Synthetic aperture sonar applied to a multibeam volume imaging sonar. Michael Hamilton (SAIC, 10260 Campus Point Dr., MS-C4, San Diego, CA 92121)

The possibility of using synthetic aperture sonar (SAS) processing techniques on a multibeam volume imaging sonar (the Buried Object Search Sonar, or BOSS) is explored. A system algorithm is developed which involves first forming several parallel synthetic apertures from a translated array oriented in the cross-track direction, then beamforming the synthetic apertures to form a fully 3D volumetric representation. The validity of the technique is verified by taking data from a subaperture of an experimental cable-guided multibeam system and applying the SAS processing string. Test targets are examined, including spheres and cylinders. The process proved quite robust, particularly considering that the system was not originally designed for SAS processing, and no sophisticated motion compensation schemes were employed. Implications for multibeam SAS processing of future AUV-based systems are discussed. [Work supported by ONR's National Research Enterprise Internship Program, and performed at Naval Surface Warfare Center, Panama City, FL.]

4:15

5pUW13. High-frequency passive sonobuoy designs with PVDF wires and their predicted performance. Juan Arvelo, Jr., Patrick Ferat, and Ron Mitnick (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099)

A novel piezoelectric material technology is used to design high-frequency (>10 kHz) air-deployed sonobuoys that exploit ambient noise anisotropy to enhance their passive performance. High frequencies have the advantages of smaller arrays and reduced clutter from distant shipping noise. PVDF wires are used to design vector sensors in azimuth with vertical directionality. Such hydrophones are used to design vertical and volumetric array configurations. Performance predictions of developed array designs are presented for selected environmental conditions. Results clearly indicate that vertical aperture is necessary to increase array gain against wind-driven noise by exploiting its anisotropy while azimuthal discrimination is required to enhance gain against nearby shipping interference. [This work is sponsored by the Office of Naval Research (ONR).]

4:30

5pUW14. Hair cell transducer design based on optical fiber Bragg gratings. Francois M. Guillot, D. H. Trivett, and Peter H. Rogers (George W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405, francois.guillot@me.gatech.edu)

The development of a biomimetic transducer that reproduces the sensing mechanism of a single hair cell is presented. When light from a broadband source is input into an optical fiber terminated by a Bragg grating, the undisturbed grating reflects a fraction of the incident light at the Bragg wavelength, and the corresponding light intensity can be measured by a photodiode. The end of the fiber onto which the grating is written is adhered to two spacers separated by a 1-mm gap, and this assembly is adhered to two plates connected by a hinge, which is located below the gap. One plate is held rigidly and the tip of the other (free) plate experiences transverse vibrations, when ensonified. These vibrations induce longitudinal strains in the length of the fiber situated over the gap. Each unstrained portion of the grating reflects a small amount of light at the Bragg wavelength, resulting in two interfering light signals. The strain modulates the interference signal and produces an ac voltage at the photodiode output. The characteristics and performance of the transducer are presented, and its applications as the building block for underwater vector and tensor sensors are discussed. [Work supported by ONR.]