CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 112, No. 3 September 2002

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 112, NO. 3, SEPTEMBER 2002

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 112, NO. 3, SEPTEMBER 2002

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.*

Reviewers of Manuscripts, 2000 and 2001

Each year the Journal endeavors to publish a list of all the persons who reviewed manuscripts during the preceding year. Such a list is a compendium of names supplied by the Associate Editors. Because our peer review system depends strongly on the continuing anonymity of the reviewers, the Journal publishes these names in alphabetical order without identification of the associate editors who provided the names and without identification of the papers they reviewed. The last such published list, for 1998 and 1999, appeared in November 2000. The list below is intended to include the names for both 2000 and 2001. The primary reason for the publication of the list is to express the Journal's gratitude to its reviewers. Reviewing a paper is often a very time consuming and demanding task, and the anonymity requirement yields no professional recognition to those who generously provide their time to help the Associate Editors decide which papers should be published and to give constructive criticisms to the authors. The Journal is justifiably proud of this list, which includes a goodly proportion of all the researchers and eminent authorities in acoustics and related fields. In a compendium of this length, omissions and errors are inevitable. If anyone notices such, please send the corrections and missing names via e-mail or regular mail to either Elaine Moran (asa@aip.org) or Allan Pierce (adp@bu.edu).

Abbas, Paul J.
Abdala, Carolina
Abe, Masanobu
Abom, Mats
Abraham, Doug
Acero, Alex
Achenbach, J. D.
Adler, Laszlo
Ainslie, Michael
Akagi, Masato
Akeroyd, Michael A.
Alku, Paavo
Allan, Prudence
Allard, Jean-François
Allen, John
Allen, Jont
Altman, Jacob A.
Andersen, Lars
Anderson, Aubrey
Anderson, Grant
Ando, Yoichi
Antes, H.
Apfel, Robert E.
Arehart, Kathryn H.
Arnason, Brian
Arnott, W. Patrick
Arons, Barry
Ashmead, Daniel H.
Askenfeldt, Anders
Astrup, Jens
Atalla, Noureddine
Atchley, Anthony A.
Attenborough, Keith
Aubauer, Roland
Augspurger, G. L.
Avan, Paul
Averkiou, Michalakis
Avis, Mark R.
Ayorinde, Emmanuel
Bachorowski, Jo-Anne
Backhaus, Scott
Badiey, Mohsen
Baer, Thomas
Bagshaw, Paul C.
Bahr, Ruth H.
Bai, Mingsian R.
Bailey, Michael R.
Bailliet, Helene
Bailly, Gerard
Baird, Gordon E.
Baird, John C.
Baker, Andrew
Baker, Richard
Balachandran, Bala
Balizer, E.
Ballato, A.
Banchet, Julien
Barbone, Paul E.
Barney, Anna
Barron, Michael
Barry, John
Barry, William J.
Bartram, J. F.
Bass, Henry E.
Bates, S. M.
Batliner, Anton
Baum, Shari R.
Beauvois, Michael
Bech, Soren
Bedard, Alfred J.
Beddor, Patrice S.
Behne, Dawn
Behrman, Alison
Belcher, Ed
Benjamin, K. C.
Bennet-Clark, Henry
Benoit-Bird, Kelly
Benthien, George
Beran, Mark
Beranek, Leo L.
Berg, Bruce G.
Berger, Elliott H.
Berke, Gerald S.
Berliner, Marilyn
Berman, David
Bernhard, Robert J.
Bernstein, Leslie
Berry, Alain
Berry, David A.
Berrymann, James G.
Berthelot, Y. H.
Berthommier, Frederick
Bickley, C.
Bilsen, Frans
Bishop, Dorothy
Bistafa, Sylvio
Blackstock, David T.
Blamey, Peter
Blauert, Jens
Blomgren, Michael
Bloothooft, Gerrit
Boden, Hans
Boe, Louis-Jean
Boersma, Paul
Boettcher, Flint A.
Bolia, Robert S.
Bolton, J. Stuart
Boluriaan, Said
Bondaryk, Joseph
Bonneau, Anne
Boone, Rinus
Booth, Newell O.
Bottega, Bill
Bouchard, Martin
Boudreau, Alex
Boulanger, Patrice
Boutillon, Xavier
Bowles, Ann
Bradbury, Jack
Bradley, Charles E.
Bradley, John S.
Braida, Louis
Brancazio, Lawrence
Branstetter, Brian
Brant, Larry J.
Breebaart, D. J.
Briggs, Kevin
Bright, C. B.
Brill, Randy
Broad, David J.
Bronkhorst, Adelbert W.
Brown, Carolyn
Brown, Charles
Brown, David K.
Brown, Judith C.
Brown, Michael G.
Bruce, Ian
Bruneau, Michel
Brungart, Douglas S.
Bucaro, J. A.
Buchta, Edmund
Buck, John
Buck, Karl
Bucker, Homer P.
Buckingham, Michael
Buell, Thomas N.
Bunnell, H. Timothy
Burdisso, Ricardo
Burkard, Robert
Burkhard, Mahlon D.
Burkhardt, John
Burnett, David
Burnett, I. S.
Burnett, Theresa
Burns, Edward
Burov, V. A.
Burroughs, C.
Busch-Vishniac, I. J.
Buss, Emily
Butler, J. L.
Buus, Soren
Cacace, Anthony
Cain, Charles A.
Caldersmith, Graham
Cantrell, John H.
Carello, Claudia
Carey, William M.
Carlile, Simon
Carlyon, Robert
Carre, Rene
Carstensen, Edwin L.
Caruthers, Jerald
Carvalho, Antonio
Castagnede, Bernard R.
Cathingnol, Dominic
Cato, Douglas
Cawley, Peter
Cazzolato, Ben S.
Cerda, Salvatore
Cerisara, Christophe
Chaigne, Antoine
Chambers, James M.
Chan, Roger W.
Chandler-Wilde, Simon N.
Chandrashekhara, K.
Chapman, David M. F.
Chapman, Ross
Chatterjee, Monita
Chazan, Dan
Chen, Weiqiu Q.
Cheng, Alexander H. D.
Cheng, C. H.
Cheng, Hsien K.
Chertoff, Mark
Cherukrui, Harish
Chevret, P.
Chew, Weng Cho
Chiba, Masakatsu
Chimenti, D. E.
Chiu, Ching-Sang
Cholewiak, Roger
Chondros, Thomas G.
Chotiros, Nicholas
Chou, Fu-chiang
Christensen, Laurel
Christopher, Ted

Chu, Dezhang
Chu, Wing T.
Church, Charles C.
Ciocca, Valter
Claribois, Jean-Pierre
Clark, Christopher
Clark, William W.
Clay, Clarence S.
Cleveland, Robin O.
Clifton, Rachel L.
Cobbold, Richard
Cohen, Lawrence T.
Colburn, H. Steven
Coleman, R.
Collins, John A.
Collins, Leslie M.
Collins, Michael D.
Colosi, John
Commander, Kerry
Connaughton, Martin
Cooke, Martin P.
Cooper, Nigel P.
Coulouvrat, Francois
Craik, R. J. M.
Crandell, Carl
Cranen, Bert
Crawford, David H.
Crawford, John
Culling, John F.
Cummings, Alan
Cummins, Fred
Cunefare, Kenneth A.
Cuomo, F. W.
Curra, Francesco
Cuschieri, Joseph
Dacol, Dalcio
Dahl, Peter
Dahleim, Marilyn
Dai, Huanping
Daigle, Gilles A.
Dain, Yefim
Dalecki, Diane
Dalenback, Bengt-Inge
d'Alessandro, Christophe
Dancer, Anton
Dancer, Armand
Danilov, Sergey
D'Antonio, Peter
Darvennes, C.
Darwin, Christopher J.
Dau, Torsten
Davies, William J.
Dawe, Lloyd A.
Deane, Grant
Dear, Steve
de Belleval, Jean-Francois
DeBoer, Egbert
Debus, Jean Claude
de Cheveigne, Alain
Decraemer, Willem F.
de Jong, Kenneth J.
de Krom, Guus
DelBalzo, Donald R.
Delgutte, Bertrand
de Lima, Washington
Demany, Laurent
Deng, Li
Denham, Sue
Dent, Micheal
Desmet, Wim
Deutsch, Diana
de Veth, Johan
de Vries, D.
Dewhurst, Richard J.
Deymier, P.
Di Giulio, Augusto
Di, Xiao
Diachok, Orest I.
DiCarlo, David
Dickey, Joseph W.
Diehl, Randy
Dillon, Harvey
Ding, Haojiang J.
Dirks, Donald
Divenyi, Pierre
Dohner, Jeffrey L.
Doinikov, Alexander
Dolan, David F.
Don, Manuel
Donaldson, Gail
Dong, Charles
Dong, S. B.
Donovan, R. E.
Donskoy, D. M.
Donskoy, Dimitri
Dooling, Robert J.
Dorman, Michael
Dosso, Stanley E.
Dowell, Earl H.
Dowling, Ann
Dowling, David
Downing, J. Micah
Dragonette, Louis
Dreschler, Wouter
Drullman, Rob
D'Spain, Gerald
Du, Gonghuan
Dubno, Judy R.
Duchateau, Jacques
Duda, Richard
Duda, Timothy
Duhamel, Denis
Dunn, Floyd
Dutoit, Thierry
Duzinski, Kathleen
Dye, Raymond
Dyer, Ira
Ebbini, Emad
Eckel, Hans E.
Eddins, Ann Clock
Eddins, David A.
Edwards, Brent
Eggermont, Jos J.
Elder, S. A.
Elishakoff, Isaac
Elko, G. W.
Elliott, Stephen J.
Ellis, Dale
Embleton, Tony F. W.
Embrechts, J. J.
Erbe, Christine
Erhmet, Helmut
Erskine, Fred
Etemoglu, Cagri
Evans, Gary W.
Evans, Richard B.
Everbach, E. Carr
Everstine, Gordon C.
Eweda, E.
Faber, Alice
Fagelson, Marc
Fahey, Paul F.
Fahnline, John B.
Fahy, Frank
Farabee, Theodore M.
Farina, Angelo
Fastl, Hugo
Faulkner, Andrew
Faure, Paul
Fawcett, John A.
Fay, Richard R.
Fegant, Oliver
Feit, David
Feke, Donald
Feleppa, Ernest J.
Ferber-Viart, Christine
Ferguson, Brian
Festen, Joost M.
Feth, Lawrence L.
Feuillade, Christopher
Field, Robert L.
Fields, James M.
Finch, R. D.
Finette, Steven I.
Fink, Mathias
Finneran, James
Fishman, Louis
Fitch, W. Tecumseh
Fitting, D. W.
Fitzgibbons, Peter J.
Flannery, Colin
Flatte, Stanley
Flege, James E.
Fletcher, Neville
Florentine, Mary
Foote, Kenneth
Formby, C. Craig
Fornberg, Bengt
Fougeron, Cecile
Fowler, Carol A.
Fowlkes, Jeffrey B.
Fox, Warren
Franco, Horacio
Frankel, Steven H.
Franzoni, Linda P.
Frazier, Linda M.
Fredberg, Jeffrey
Frederickson, Carl K.
Freed, Daniel
Freeman, Dennis
Freyman, Richard
Frijns, Johan
Frik, Gerhard
Frikha, Slaheddine
Frisk, George
Fristrup, Kurt
Fromm, David
Fry, Francis J.
Fu, Qian-Jie
Fujiwara, Kyoji
Fulford, James
Fuzessery, Nick
Gabrielson, T. B.
Gade, Anders C. S.
Galland, Marie-Annick
Gallasch, Eugen
Gallun, Erick
Garbe-LaPell, Collen
Gardonio, Paolo
Garga, A. J.
Garrelick, Joel
Garrett, Steven
Gauger, Dan
Gaumond, Charles
Gaunaurd, Guillermo C.
Gaunt, Abbot
Gaunt, Sandra
Gauthier, François
Gentner, Tim
Gentry, Cassandra
Gerratt, Bruce R.
Gerstoft, Peter
Gibbs, Gary P.
Gibian, Gary
Gifford, Rene
Gilkey, Robert
Gjestland, Truls
Glasberg, Brian R.
Glegg, Stewart
Gobl, Christer
Goldstein, Julius L.
Gonneau, E.
Goodman, Ralph R.
Goold, John
Gordon, Jonathan
Gordon, Peter
Gordon-Salant, Sandra
Gorga, Michael P.
Grabe, Esther
Gracewiski, Sheryl M.
Gragg, Roger
Graham, W. R.
Granstrom, Bjorn
Grant, Kenneth W.
Grantham, D. Wesley
Greaves, Matthew J.
Green, Phil
Green, W. A.
Greenberg, Julie
Greene, Charles
Greenlaw, Charles
Greenleaf, James F.
Grimault, Nicolas
Grinfeld, Michael
Grinnell, Alan
Gros, John
Grose, John H.
Grosh, Karl
Grove, Deborah
Gucunski, Nenad
Guerich, Mohamed
Guervich, Boris
Guettler, Knut
Guicking, Dieter
Guilhot, Jean-Pierre
Guillot, F.
Guinan, John J.
Gummer, Anthony W.
Gusev, Vitalyi
Guyader, Jean-Louis
Guyer, Robert
Habeger, Charles
Haberman, R. C.
Hachiya, Hiroyuki
Haddrow, James B.
Haering, Edward A., Jr.
Hafter, Ervin
Hall, Donald E.
Hall, Joseph L.
Hall, Joseph W.
Hall, Marshall
Hamernik, Roger
Han, Mieko S.
Handel, Stephen J.

Hansen, Colin H.
Hansen, John H. L.
Harkin, Anthony
Harris, Andrew S.
Harris, Francis P.
Harris, G. R.
Harris, John G.
Harrison, Chris
Harrison, George H.
Hart, Lynette
Hartmann, Bruce
Hartmann, William M.
Hartung, Klaus
Hasegawa-Johnson, Mark
Hatfield, Julie
Hattis, Dale
Havelock, D. I.
Hawkins, Sarah
Hawks, John W.
Hawley, Monica
Hayden, M. S.
Haykin, Simon
Hazan, Valerie
He, Ning-Ji
He, Ping
Heaney, Kevin
Heard, Garry
Hedberg, Claes
Heffner, Henry
Heil, Peter
Heinz, Michael G.
Heller, Eric
Heller, Laurie
Hellman, Rhona P.
Hellman, William
Helweg, David
Henderson, Brenda
Henning, Bruce
Henyey, Frank S.
Herdic, Peter C.
Herman, Lloyd A.
Herstein, Louis
Hertrich, Ingo
Herzel, Hanspeter
Herzing, Denise
Heutschi, Kurt
Hickey, Craig
Hickling, Robert
Hicks, Michelle
Hidaka, Takayuki
Hildebrand, John
Hill, Nicholas
Hillenbrand, James M.
Hiller, Robert
Hines, Paul
Hirao, Masahiko
Hirose, Keikichi
Hirsch, Karl W.
Hirschberg, Avraham
Hirsekorn, Sigurn
Hixson, Elmer
Hodge, Megan
Hodgson, Murray R.
Hoff, Lars
Hoffman, Howard J.
Hoffman, Michael W.
Hogden, John
Holger, Jens
Holland, Christy K.
Holliday, D. Vance
Hollien, Harry
Holt, Lori L.
Hom, C. L.
Homm, Anton
Honarvar, Farhan
Hoole, Philip
Horne, John
Hornsby, Benjamin
Horonjeff, Richard D.
Horoshenkov, Krill
Horst, Weibe
Horwitz, Amy R.
Hossack, J. A.
Hothersall, David C.
Houde, John
Houston, Brian H.
Houtgast, Tammo
Houtsma, Adrianus
Hovem, Jens
Howarth, Thomas R.
Howe, Bruce
Howe, Michael S.
Hoy, Ronald
Hsu, David K.
Huang, Lixi
Hubbard, Allyn
Huebner, C.
Hukin, Robert
Hurst, Seward
Hygge, Staffan
Hynynen, Kullervo H.
Igusa, Takeru
Ih, Jeong-Guon
Ih, Young-Sang
Ihlenburg, Frank
Ilinskii, Yurii
Insana, Michael F.
Isabelle, Scott
Ise, Shiro
Ivakin, Anatoli
Ivansson, Sven
Jackson, Darrell
Jacobs, Laurence J.
Jacobsen, Finn
Jamieson, Janet
Janssen, Thomas
Jansson, Erik
Jech, J. Michael
Jen, Phillip
Jenison, Rick L.
Jesteadt, Walt
Jesus, Sergio
Johansson, Orjan
Johnson, C. S.
Johnson, Daniel L.
Johnson, David L.
Johnson, Marty E.
Jones, Chris
Jones, D. F.
Jones, Dylan M.
Jongman, Allard
Joris, Philip
Joseph, Phillip
Kaernbach, Christian
Kaimya, N.
Kalisstrotov, Maragarita A.
Kalko, Elizabeth
Kalveram, Karl Th.
Kang, Jian
Kaplunov, Julius
Karjalainen, Matti
Kargl, Steven G.
Karnell, Michael P.
Karpov, Sergey
Kashino, Makio
Kastak, David
Kates, James M.
Katz, William F.
Kawahara, Hideki
Kay, Lesley
Keefe, Douglas H.
Keenan, Ruth
Kewley-Port, Diane
Khismatullin, Damir
Khokhlova, Vera
Kidd, Gary R.
Kidd, Gerald D., Jr.
Kiddy, J. S.
Kidner, Mike
Kieffer, Richard S.
Kiefte, Michael
Kihlman, Tor
Kil, Hyun-Gwon
Kim, Hong Kook
Kim, Kwang Yul
Kistler, Doris J.
Klatte, Maria
Klauson, Aleksander
Kleiner, Mendel
Knio, Omar
Knobles, David
Ko, S. H.
Koch, Robert
Koehnke, Janet D.
Koessl, Manfred
Kohlrausch, Armin
Kollmeier, Birger
Konrad-Martin, Dawn L.
Koopmann, Gary H.
Kopec, John W.
Korman, Murray
Krane, Michael K.
Kreiman, Jody
Kristiansen, Ulf R.
Krokstad, Asjorn
Krumbholz, Katrin
Kubin, Gernot
Kuehn, David P.
Kumada, Masanobu
Kumar, Arun
Kummer, Peter
Kuo, Sen Maw
Kuperman, William A.
Kurz, Thomas
Kushwaha, Manvir
Kuttruff, Henrich
Kwon, S.
Laboissiore, Raphael
Ladd, Robert D.
Lafarge, Denis
Lafluer, L. Dwynn
Lagakos, N.
Lai, Y.-C.
Lakatos, Stephen
Lam, K. Y.
Lam, Y. W.
Lammers, Marc
Landsberger, Brian J.
Lane, Harlan
Langston, Charles
Laprie, Yves
Larsen, Ole
Larson, Charles
Larson, Gregg D.
Lauchle, Gerald
Laukkanen, Anna-Maria
Laura, Patricio
Lauriks, Walter
Lavenex, Pamela Banta
Lavery, Andone
Lebedeva, I. V.
LeCar, Howard
Lee, Jang Moo
Lee, Robert A.
Leek, Marjorie
Lees, Sidney
Legge, Katherine
Lehr, Andre
Leighton, Timothy G.
Leishman, Timothy
Leissa, A. W.
Leistner, P.
Lemonds, David
Lentz, Jennifer J.
LePage, Kevin
Lercher, Peter
Letcher, Stephen V.
Levitt, Harry
Lewers, T.
Lewin, Peter A.
Li, K. M.
Li, Yadong
Lieberman, Philip
Lightfoot, Jay
Liljencrants, Johan C. W. A.
Lilley, Geoffrey M.
Lim, Raymond
Lin, Jenshinn
Lin, Tai
Lindblad, Sven
Lindestad, P.
Lindevald, I. M.
Lipkens, Bart
Litovsky, Ruth
Liu, Qing-Huo
Lizzi, Frederic L.
Ljunggren, Sten
Llyrintsis, Anastasios S.
Lo, Kam
Lobel, Phil
Lock, James A.
Lockard, David
Loizou, Philipos
Long, Glenis R.
Lorenzi, Christian
Loughlin, Patrick J.
Love, Richard
Lu, Jian-Yu
Lu, Tian J.
Lubman, David
Lucero, Jorge C.
Ludlow, Christy
Ludvigsen, Carl
Lutfi, Robert A.
Lutman, Mark E.
Luz, George A.
Lynch, James
Lyon, Richard H.
Lyons, Tony
Maa, Dah-You
Macpherson, Ewan
Madden, John
Madsen, Ernest L.
Maeda, Shinji

Maeda, Tatsuo
Maestrelo, Lucio
Maglieri, Domenic J.
Makarov, Sergey N.
Makris, Nicholas
Mangiante, G.
Manley, Geoffrey
Mann, David
Mann, J. Adin, III
Marburg, Stefan
Marcus, Martin
Margolis, Robert H.
Marsh, Alan H.
Marshall, A. Harold
Marshall, Lynne
Marston, Philip L.
Martin, Glen K.
Martin, Nadine
Martin, Paul
Mason, Christine
Massaro, Dominic
Masson, Patrice
Mast, T. Douglas
Masters, Mitch
Matsuo, Ikuo
Matula, Thomas
May, Bradford
Mayer, Andreas
Mayer, Larry
Mayer, W. G.
Maynard, Julian D.
Maze, Gerard
McAlpine, David
McAnally, Ken
McCammon, Diana
McClements, Julian
McCowan, Brenda
McCoy, John J.
McDaniel, J. Gregory
McDonald, B. Edward
McDonald, Marl
McFadden, Dennis
McFarland, David
McGhee, Duncan
McGowan, Richard S.
McKay, Colette
McKisic, J. M.
McLaughlin, Stephen
McPherson, David
Mechel, F.
Meddis, Raymond
Medwin, Herman
Mehl, James B.
Melcher, Jennifer
Mellinger, David
Meloni, Tommaso
Mendel, Lisa Lucks
Mendoza, Lee
Menge, Christopher W.
Menotti, Francis R.
Menotti, Frank
Menounou, Penelope
Mercer, James
Mershon, Donald
Mettin, Robert
Meuhleisen, Ralph
Michaelis, Dirk
Micheau, Philippe
Michielson, Axel
Middlebrooks, John
Miedema, Henk M. E.
Migliori, A.
Mikhalevsky, Peter
Miller, Douglas L.
Miller, James G.
Miller, Judi Lapsley
Miller, Lee
Miller, Morton W.
Miller, Nicholas P.
Miller, Patrick
Mills, David M.
Mills, Harold
Mills, John H.
Minachi, Ali
Ming, Ruisen
Moffett, M. B.
Mohl, Bertel
Mom, Thierry
Mommertz, Eckard
Monaghan, Alex I. C.
Mongeau, Luc
Moody, David
Moon, Seung-Jae
Moore, Brian C. J.
Moore, Patrick
Moore, Susan
Morley, Robert
Moros, E. G.
Morris, Howard
Morris, Philip
Moss, Cynthia
Mould, John
Mountain, David C.
Mourad, Pierre D.
Mozurkewich, George
Muehleisen, Ralph
Mueller, Rolf
Munhall, Kevin G.
Munjal, M. L.
Murphy, Peter J.
Myrberg, Arthur A.
Naatanen, Risto
Nabili, Renato
Nachtigall, Paul
Nagem, Raymond
Nagy, Peter B.
Naidu, Ram
Nakamura, Satoshi
Namba, Seiichiro
Naugolnykh, Konstantin
Nearey, Terrance M.
Nederveen, Cornelis
Neeley, Stephen T.
Neff, Donna L.
Nelson, David
Nelson, Philip A.
Nemeth, Edwin
Nero, Redwood
Neuman, Arlene C.
Nicholas, Michael
Nielsen, Peter
Nightingale, Trevor N. T.
Noble, William
Nobles, David
Nolle, A. W.
Norris, Andrew
Norton, Guy V.
Novarini, Jorge C.
Nuttall, Alfred L.
Nyborg, Wesley L.
Nystuen, Jeffrey
O'Brien, William D.
O'Donnell, Mathew
Oda, Hitoshi
Oelze, Michael
Oguz, H.
Ohala, John J.
Ohde, Ralph N.
Ohl, Claus-Dieter
Ohlemiller, Kevin
Olson, Elizabeth
Ophir, Jonathan
Orduna-Bustamante, Felipe
Orr, Marshall H.
Orris, Gregory
Osetrov, A
Osler, John
Ostrovsky, Lev
Ostry, David J.
Otto, Norman C.
Oxenham, Andrew J.
Ozdamar, Ozcan
Ozer, Bulent
Page, John
Paillard, Bruno
Paliwal, Kuldip
Palmer, Alan
Palmer, David
Pan, Jie
Panton, Ronald
Park, Minkyu
Parker, Kevin J.
Parkins, John W.
Parthasarathy, Sarangarajan
Pastore, Richard E.
Patterson, Roy D.
Patuzzi, Robert
Pavic, Goran
Payton, Karen L.
Pearsons, Karl S.
Peat, Keith
Pelorson, Xavier
Peppin, Richard
Perkins, Mark E.
Perrier, Pascal
Perrott, David
Peters, Robert W.
Petersson, B.
Pfingst, Bryan
Pflug, Lisa A.
Phillips, Dennis P.
Photiadis, Douglas
Piacsek, Andrew
Picaut, Judicael
Piercy, J. E.
Piquette, Jean C.
Pireto, Thomas E.
Piterman, Michel
Pitts, Todd A.
Plack, Christopher J.
Plotkin, Kenneth J.
Polich, John
Pollard, Howard F.
Polling, Hugh
Pompoli, Roberto
Ponton, Curtis
Popovics, J. S.
Popper, Arthur
Port, Robert F.
Porter, Michael B.
Potter, John
Pouliquen, Eric
Povey, Malcolm J. W.
Prada, Claire
Prasad, M.
Preisig, James
Premat, Eric
Pressnitzer, Daniel
Preves, David
Prieve, Beth A.
Prince, Mary M.
Probtst, Rudolf
Prosperetti, Andrea
Puria, Sunil
Qiu, Xiaojun
Qu, Jianmin
Quatieri, Thomas
Raatgever, Johan
Rabinkin, Daniel
Rafaely, Boaz
Raggio, Marcia
Rajan, Subrahaniam
Rakerd, Brad
Ramsay, Gordon
Randall, Jan
Raphael, David T.
Rasmussen, Karsten Bo
Raspet, Richard
Ravicz, Michael E.
Rawlins, Tony
Ray, Laura
Reardon, Fredrick H.
Rebinsky, Douglas A.
Recio, Alberto
Redekop, David
Reichard, Karl M.
Reid, John M.
Reinhardt-Rutland, Anthony
Relkin, Evan
Remington, Paul J.
Renault, Stephane
Restorff, James B.
Reynolds, Steven
Rho, Jae-Young
Rhode, William S.
Richards, Virginia M.
Richardson, Michael
Richardson, Virginia
Ricker, Dennis
Rickert, Martin
Ricketts, Todd
Ridgway, Sam H.
Riedlinger, Rainer E.
Rigoll, Gerhard
Rindel, Jens Holger
Ring, M. D.
Robert, Daniel
Roberts, Brian
Robertson, William M.
Robinson, H. C.
Robinson, Leick D.
Rogers, James C.
Rogers, P. H.
Rokhlin, Stanislas
Rolt, Kenneth D.
Romano, Tony
Romero, Jose
Rose, Joseph
Rosen, Stuart
Rosowski, John J.
Rossing, Thomas D.
Rott, Nicholas
Roure, Alain
Rouseff, Daniel

Roux, Philippe
Roy, Ronald A.
Royster, Larry H.
Royston, Tom J.
Rozenfeld, Ilya
Rubinstein, Jay
Ruggero, Mario
Rumerman, Mel
Russell, Daniel A.
Russell, Ian
Rutledge, Janet
Sabatier, James M.
Saberi, Kourosh
Salomons, Erik M.
Salsburger, Hans-Jurgen
Salt, Alec N.
Salvi, Richard
Samar, Vincent
Samejima, Toshiya
Sammelman, Gary S.
Sanchez-Perez, J. V.
Sanchis, Lorenzo
Sanes, Dan
Sapienza, Christine
Sapozhnikov, Oleg
Sarkissian, Angie
Sarvazyan, Armen
Sas, Paul
Saunders, Timothy J.
Saunders, William R.
Sayich, Laela
Scaife, Ronan
Scherer, Ronald C.
Schlauch, Robert S.
Schlaugh, Gottfried
Schmiedt, Richard A.
Schneider, Bruce
Schneider, Hans
Schoen, Peter
Schoentgen, Jean
Schomburg, Annette
Schomer, Paul D.
Schroeder, Manfred
Schroeter, Juergen
Schroger, Erich
Schul, Johanes
Schulauch, Robert
Schulte-Fortkamp, Brigitte
Schumacher, Robert T.
Schusterman, Ronald J.
Selamet, Ahmet
Selbie, Scott
Sessler, G. M.
Shackleton, Trevor
Shadle, Christine H.
Shah, N. Jon
Shamma, Shihab
Shang, Er-Chang
Shankar, P.
Shannon, Robert V.
Sharma, Anu
Sharma, Sangit
Sharp, Ben H.
Shatz, Lisa
Shaw, Leonard
Sheft, Stanley E.
Shen, I. Y.
Sheperd, Rob
Shepherd, Kevin P.
Shera, Christopher A.
Sherman, C. H.
Shield, Bridget M.
Shimamura, Tetsuya
Shinn-Cunningham, Barbara G.
Shorter, Philip J.
Shriberg, Elizabeth
Shung, K. Kirk
Siderius, Martin
Siegel, Jonathan H.
Siegmann, William
Simmons, James
Sinclair, A. N.
Sinder, Daniel J.
Sinex, Donal
Singh, R. J.
Sinha, Bikash
Sininger, Yvonne S.
Sirrevaag, Erik
Sivadas, Sunil
Skinner, Margaret
Slifka, Janet
Smith, George B.
Smith, Kevin B.
Smith-Abouchacra, Kim
Smits, Roel
Smoorenburg, Guido F.
Smurzynski, Jacek
Snell, Karen B.
Snyder, Robert
Snyder, Russell
Snyder, Scott D.
Soegijanto B.
Solodov, Igor
Somers, Mitchell S.
Sommerfeldt, Scott D.
Sommers, Mitchell S.
Sondhi, Mohan M.
Song, Hee Chum
Soquet, Alain
Sorokin, Victor
Southall, Brandon
Souza, Pamela
Spangler, Henry
Sparrow, Victor W.
Spector, Alexander
Speisberger, John L.
Spelman, Sandy
Spicer, J.
Spiesburger, John
Spoor, Philip
Srstulovik, Sacha
Stachurski, Jacek
Stafford, Kathleen
Stanton, Timothy K.
Stapells David
Steele, Charles
Stein, Peter
Steinshneider, Mitch
Stellmack, Mark A.
Stepanishen, Peter R.
Stephen, Ralph
Stergiopoulos, Stergios
Stern, Richard M.
Stevens, Kenneth
Stewart, Catherine M.
Stewart, Noral D.
Stiller, Detlef
Stinson, M. R.
Stoll, Robert
Stone, Michael
Storey, Brian
Story, Brad H.
Strasberg, Murray
Strel'tsov, Vladimir
Strick, Helmer
Strickland, Elizabeth A.
Strong, William J.
Stroud, John
Strube, Hans W.
Strybel, Thomas
Studebaker, Gerald A.
Stuliff, Daniel
Stylainou, Yantis
Sugimoto, Nobumasa
Sujith, R. J.
Sullivan, Brenda M.
Sullivan, E. J.
Sum, K.
Summerfield, Quentin
Summers, Van
Sun, Jian-Qiao
Sundberg, Johan E. F.
Suni, Ian I.
Supin, Alexander Y.
Surlykke, Annamarie
Sussman, Harvey M.
Sussman, Joan E.
Sutherland, Louis C.
Sutin, Alexander M.
Svec, Jan
Svensson, U. Peter
Svirsky, Mario
Swanson, David C.
Swift, Gregory W.
Szabo, Thomas L.
Sze, K. Y.
Tabain, Marija
Taherzadeh, Shahram
Takagi, Kenshiro
Takahashi, D.
Takahashi, Jack
Takahashi, Terry
Talmadge, Carrick L.
Tan, Hong Z.
Tanaka, N.
Tang, D. J.
Tang, S. K.
Tappert, Frederick D.
Taroudakis, Michael
Tavakkoli, Jahangir
Temkin, S.
ter Haar, Gail R.
Terhardt, Ernst
Ternstrom, Sten
Terrill, Eric
Tewary, Vinod K.
Theriault, Jim
Thibodeau, Linda
Thode, Aaron
Thompson, S. C.
Thompson, Shirley J.
Thomson, David J.
Thorax, John
Thornton, Christine
Thorsos, Eric
Tiede, Mark
Tindle, Chris
Ting, Robert
Ting, Thomas
Titterton, Paul J.
Tocci, Gregory C.
Tollin, Daniel J.
Tolstoy, Alexandra
Toro, E. F.
Torprevaa, Hans
Tracey, Brian H.
Tracy, Maureen B.
Trahey, Gregg E.
Trancoso, Isabel
Traunmiller, Hartmut
Tremblay, Kelly
Tressler, J. F.
Trevorrow, Mark
Trine, Timothy
Tubis, Arnold
Turgut, Altan
Turner, Joseph
Tyack, Peter
Uchanski, Rosalie
Umnova, Olga
Unger, Evan C.
Vakakis, Alexander F.
Valenzuela, Miriam
Valle, Christine
Van Buren, A. L.
Vance, Timothy J.
van de Par, Steven
van den Honert, Chris
van der Heijden, Marcel
Van Dijk, Pim
Van Moorhem, William K.
van Opstal, John
van Santen, Jan
van Son, Robert J.
Vaudrey, Michael
Veldhuis, Raymond
Vera, Michael
Verdier, Claude
Verschuure, J.
Vieira, Maurilio N.
Viemeister, Neal F.
Vieregge, Wilhelm H.
Vipperman, Jeffrey S.
Viswanthan, Vishu
Vlahopoulos, Nickolas
Vollmann, Jacqueline
Voran, Stephen
Vorlaender, Michael
Voronovich, Alexander
Vos, Henk
Vos, Joos
Voss, Susan
Waag, Robert C.
Wagner, Hermann
Wakefield, Gregory H.
Wales, Stephen C.
Walsh, Edward J.
Wang, B. T.
Wang, C. Y.
Wapenaar, C. P. A.
Warren, Richard
Watson, Charles S.
Watts, Greg
Watts, Lloyd
Wauer, Joerg
Waxler, Roger M.
Wear, Keith A.
Weaver, Richard L.
Weber, Dan
Weber, Reinhard
Weisenberger, Janet M.
Werner, Lynne
Westwood, Evan

Wettergren, T.
Whalberg, Magnus
White, Michael J.
Wiegrebe, Lutz
Wightman, Frederic L.
Wilen, Larry A.
Williams, David
Williams, Kevin
Wilson, D. Keith
Wilson, O. B., Jr.
Winkler, Istvan
Wit, Hero P.
Withnell, Robert H.
Wittenberg, Thomas
Wodicka, George R.
Wojtczak, Magdalena
Wolf, Steven
Wolfe, J.
Wolfson, Michael
Wong, G. S. K.
Woodhouse, James
Wooten, Janine
Worcester, Peter
Wouters, Jan
Wright, Beverly A.
Wu, Junru
Wu, Sean
Wurmser, Daniel
Xiang, Ning
Xu, Yi
Yan, Yonghong
Yang, Tsih C.
Yano, Takeru
Yantorno, Robert
Yates, Graeme K.
Ye, Zhen
Yehia, Hani
Yin, Thomas
Yooon, Suk Wong
Yost, William A.
Yu, Yung H.
Zabolotskaya, Evgenia
Zagzebski, James A.
Zahorik, Pavel
Zakarauskas, Pierre
Zatorre, Robert
Zeng, Fan Gang
Zhong, Pei
Zhou, Jin-Xun
Zhou, X.
Zimmer, Karin
Zingarelli, Robert A.
Ziomek, Lawrence J.
Ziskin, Marvin C.
Zoltan, Nick
Zuckerwar, Allan F.
Zurek, Patrick

Reports of Technical Committees 2001–2002

(See October issue for additional technical committee reports)

Acoustical Oceanography

As the first order of business in this report, the Technical Committee on Acoustical Oceanography (AO) would like to thank its outgoing Chair, James F. Lynch, for three years of outstanding service. He initiated the AO ''Mini-Course in Oceanography,'' in which two oceanographers, generally from outside the ASA attend the technical sessions and give one hour mini-tutorials about various areas of oceanography (which are generally related topically to our sessions). He was also responsible for organizing and overseeing the selection process for the first recipient of the Medwin Prize in Acoustical Oceanography. Under his leadership AO has been an active and exciting Technical Committee.

The fall 2001 ASA meeting in Ft. Lauderdale was a busy one for Acoustical Oceanography. The first Medwin Prize in Acoustical Oceanography was awarded to Timothy G. Leighton of the Institute of Sound and Vibration Research, Southampton University, ''for the effective use of sound in the discovery and understanding of physical processes and parameters in the sea, particularly ocean bubbles.'' He gave the Medwin Prize Lecture, entitled ''Surf Zone Bubble Spectroscopy: The Role of the Acoustic Cross-Section.'' Robert Spindel presented a special lecture, entitled ''History of Acoustical Oceanography,'' memorializing the significant achievements and milestones in acoustical oceanography during the first three quarters of the Society's first century. The lecture is part of the History Lecture Series organized by the ASA Committee on Archives and History. Alex Hay organized an AO special session on ''Turbulence and Finestructure Studies,'' which included AO Mini-Course presentations by Ann Gargett on water column turbulence and John Trowbridge on boundary layer turbulence. AO also cosponsored a special session on ''Acoustic Image Reconstruction Using Tomographic Techniques.'' The issues surrounding marine mammals and undersea sound were highlighted at the meeting. AO cosponsored two special sessions related to these issues, ''Acoustics and Marine Mammals'' and ''Ocean Research and Marine Mammal Regulatory Issues.'' In addition, Peter Worcester represented AO in a panel discussion on ''ASA's Role in Marine Mammal Bioacoustics.'' Kelly J. Benoit-Bird was selected for the Best Student Paper Award for her paper entitled ''Energy: Converting from acoustic to biological resource units.'' The AO representatives to the Technical Program Organizing Meeting (TPOM) had a challenging task given the large number of special sessions and papers in both AO and UW. We are indebted to Michael Brown and David Palmer for their efforts.

The spring 2002 ASA meeting in Pittsburgh was also busy and successful. The 2002 Medwin Prize in Acoustical Oceanography was awarded to Bruce D. Cornuelle of the Scripps Institution of Oceanography, University of California at San Diego, ''for the effective use of sound in the discovery and understanding of physical processes in the sea, using inverse methods to interpret long-range acoustic propagation data.'' He gave the Medwin Prize Lecture, entitled ''Ocean Tomography, Inverse Methods, and Broadband Ocean Acoustics.'' (Subsequent Medwin Prizes will continue to be awarded at the spring ASA meetings.) Mohsen Badiey organized an AO special session entitled ''Ed Hamilton Memorial Session on Sediment Properties,'' and R. Lee Culver organized one entitled ''Acoustical Studies of Bubbles in the Ocean.'' The Best Student Paper Awards for Pittsburgh have not yet been announced. We are indebted to David Palmer (again!) and Kyle Becker for representing AO at the TPOM.

The Acoustical Oceanography Technical Committee now has a new web site, for which we are indebted to Gopu Potty and James Miller. It can be reached through the ASA web page by clicking on ''Committees.'' Please check it out and e-mail any material, additional links, or comments that you might have to potty@oce.uri.edu or miller@oce.uri.edu.

The greatest concern that I have encountered during my first year as Chair of the AO Technical Committee is the extent to which AO and UW sessions sometimes overlap at ASA meetings. AO and UW have formed a small joint committee, chaired by James Lynch and Henrik Schmidt, to make recommendations on possible approaches to minimizing this problem.

I would like to close by thanking all of the volunteers who make the AO Technical Committee a success!

PETER F. WORCESTER
Chair

Animal Bioacoustics

The Animal Bioacoustics Technical Committee and its activities continued to expand during the past year. Members and friends of the committee met at the Ft. Lauderdale and Pittsburgh ASA meetings with approximately 30 attending in Ft. Lauderdale and 20 in Pittsburgh.

Animal Bioacoustics sponsored three special sessions at the fall meeting in Ft. Lauderdale: Fish Audition and Sound Production organized by David Mann and Mardi Hastings, Acoustics and Marine Mammals organized by Joseph Blue and Edmund Gerstein, and a Session Honoring the Contributions of William Watkins organized by Peter Tyack. In addition Mardi Hastings gave a Hot Topics presentation that included contributions from several members. The program on Acoustics and Marine Mammals, co-sponsored with the Technical Committees on Underwater Acoustics, Acoustical Oceanography and Noise, consisted of five very well attended sessions that culminated with a lively panel discussion on ASA's Role on Marine Mammal Bioacoustics. Charles Schmid organized and moderated the panel discussion. Members of the panel were Whitlow Au (Hawaii Institute of Marine Biology, University of Hawaii), David Bradley (Applied Research Laboratory, Pennsylvania State University), Roger Gentry (National Marine Fisheries Service, National Oceanographic on Atmospheric Administration), Bob Gisiner (Marine Mammal Science and Technology program, Office of Naval Research), Darlene Ketten (Woods Hole Oceanographic Institute and Harvard Medical School), Joel Reynolds (Natural Resource Defense Council), Ben Stein (American Institute of Physics), and Peter Worcester (Scripps Institution of Oceanography). The panel and audience discussed ideas to improve standards, research, education and communication pertaining to current issues in marine mammal bioacoustics.

At the spring meeting in Pittsburgh Jim Simmons gave a very exciting and well-attended tutorial lecture, ''Keep You Ear on the Ball: Display of Targets in the Bat's Sonar Receiver.'' Ann Bowles organized special sessions on the Acoustics of Terrestrial Mammals and Birds and Jeanette Thomas organized sessions Honoring the Contributions of Bill Evans. Animal Bioacoustics joined the Technical Committees on Noise and Speech Communication as co-sponsor of a session on Public Policy Issues. This session highlighted several issues to be examined by the newly formed ad hoc Panel on Public Policy.

Four special sessions are planned for the upcoming meeting in Cancun. In addition Arthur Popper and Robert Dooling will give an hour presentation on the History of Animal Bioacoustics in the Society. A written

version of this lecture will appear in *The Acoustical Society of America - The First 75 Years*, a book being prepared by the Archives and History Committee for distribution at the 75th anniversary meeting in New York in Spring 2004.

Congratulations go to Jim Finneran who received the R. Bruce Lindsay Award in Pittsburgh for his contributions to hearing in teleost fishes and dolphins. Also during the past year Darlene Ketten, John Richardson, Andrea Simmons and William Watkins were honored by being elected Fellows of the Society for their contributions to the field of animal bioacoustics.

Whitlow Au and Charles Schmid are planning and organizing the 1st Symposium on Acoustic Communication by Animals, using special technical initiative funds from ASA. This symposium will be held July 27–30, 2003, at the University of Maryland-College Park. Several colleagues from Animal Bioacoustics are serving on the organizing and technical committees for this workshop.

Animal Bioacoustics continues to be served by many excellent volunteers. Charles Greene is replacing Whitlow Au as the Animal Bioacoustics representative to the Medals and Awards Committee as Whit is currently a member of the Executive Council. Darlene Ketten will replace Charles on the Membership Committee. Ann Bowles will continue as the Animal Bioacoustics representative to ASACOS and Micheal Dent as representative to the ASA Student Council. David Mann attended the Technical Program Organizing Meetings for Ft. Lauderdale and will also go to the one for Cancun. David Bradley from Underwater Acoustics graciously filled in for Animal Bioacoustics at the organizing meeting for Pittsburgh. A special thanks goes to all the session organizers and speakers who contributed to the success of the Animal Bioacoustics program during the past year.

Mardi Hastings
Chair

Architectural Acoustics

TCAA had a very busy and productive year, with numerous special sessions in Chicago and Ft. Lauderdale, and numerous activities supported by technical initiatives from the Technical Council.

Highlights included the ASA meetings, continued work on classroom acoustics, adopting management of the Newman Award, the development of an ASA ''Registry of Acoustical Places,'' development of a book on halls for music performance, and even support of a video production for classroom acoustics.

The local meeting committee for Chicago was comprised primarily of TCAA members, including Scott Pfeiffer (the general chair), along with Paul Calamia, John Kopec, Martha Larson, and Ian Hoffman.

Special sessions at the Chicago meeting included concert hall measurements, first 80 milliseconds in auditoria, a poster session of halls for music performance, a joint poster session (with Engineering Acoustics and Noise) on acoustical test facilities, student design competition in architectural acoustics, architectural acoustic modeling and imaging, a walking tour of three Chicago theaters, measurements and predictions of building materials and systems (joint with Engineering Acoustics and Noise), privacy and articulation in classrooms and other occupied spaces (joint with Noise), and the sound of baseball (joint with Noise, Physical Acoustics, and Engineering Acoustics). Session chairs included Christopher Jaffe, Jerald Hyde, Christopher Storch, Timothy Foulkes, Ian Hoffman, Charles Moritz, Brigitte Schulte-Fortkamp, Daniel Raichel, Robert Coffeen, Lily Wang, Robin Glosemeyer, Neil Shaw, Brandon Tinianov, David Lubman, and Joseph Pope. Paper sorters were Brandon Tinianov and Robin Glosemeyer. Robert Coffeen, Lily Wang, and Robin Glosemeyer directed the student design competition. K. Anthony Hoover was introduced as the new TCAA chair. Dorie Najolia reported that the student council was developing a new web page. David Egan was elected as chair of the subcommittee managing the Robert Newman Award. William Cavanaugh reported on the development of a ''Registry of Acoustical Places.'' TCAA members Angelo Campanella and Soren Bech were recognized as new Fellows of ASA. TCAA, Animal Bioacoustics, Music, Noise, Structural Acoustics, and the National Council of Acoustical Consultants jointly sponsored a student reception. Scott Pfeiffer, Paul Calamia, and Martha Larson arranged a lecture presentation at nearby Orchestra Hall, and also worked with Richard Talaske in arranging a theater crawl of Auditorium Theater, the Palace Theater, and Goodman Theater.

Sessions at the Ft. Lauderdale meeting included room acoustics potpourri, speech privacy design in office spaces, the TCAA Vern O. Knudsen Distinguished Lecture, integration of synthesis techniques with ''acoustic'' music (joint with Musical Acoustics), cruise ship acoustics (high seas acoustical adventures) (joint with Noise), speech intelligibility and metrics (joint with Speech Communications), and measurements and materials. Session chairs included David Marsh, John Erdreich, Kenneth Roy, David Lubman, K. Anthony Hoover, Richard Campbell, Russell Cooper, and Peggy Nelson. Paper sorters were Lily Wang and David Marsh. David Bradley reported that the student council is preparing a section on the ASA website. Jerry Lilly was recognized as a new Fellow of the Acoustical Society of America. TCAA, Animal Bioacoustics, Music, Noise, and the National Council of Acoustical Consultants cosponsored a student luncheon. Members were encouraged to attend the Acoustical Society Foundation fund raising dinner featuring Sebastian Junger.

Technical initiatives continue to support TCAA's website maintenance, student receptions, and the printing of the classroom acoustics brochure. Technical initiatives also supported AIA (American Institute of Architects) certification for courses taught by members of TCAA who offered courses to architects.

TCAA members continued to assist in the development of the guidelines for proper acoustics in classrooms, with a goal of completing an ANSI standard.

Roughly 70 members attended the TCAA meeting in Chicago and about 50 attended in Ft. Lauderdale.

Members continue to actively participate in all aspects of TCAA, both at ASA meetings and beyond.

Active participation remains strong in TCAA, and our membership, including student members, continues to grow.

K. ANTHONY HOOVER
Chair

Biomedical Ultrasound/Bioresponse to Vibration

Twenty-one members attended the B&B technical committee (TC) meeting in December 2001, in Ft. Lauderdale, FL. Fall meetings are our ''off'' meetings, at which we usually try to sponsor innovative programs rather than a host of special sessions. Hence Pierre Mourad organized and chaired the all-day Topical Meeting on the ''Physics of Ultrasound in Relation to the Biology of its Therapeutic Effects: HIFU, LOFU, and Imaging HIFU Lesions.'' Michael Bailey and Ibrahim Hallaj were contributed-paper session chairs and Subha Maruvada was the TPOM representative. At the Ft. Lauderdale meeting, B&B had 32 papers in four sessions (two of which were the Topical Meeting), and all were well-attended (peaks of more than 40 attendees at every session). B&B also cosponsored two sessions: ''Theory and Applications of Acoustic Time Reversal'' and ''Acoustic Image Reconstruction Using Tomographic Techniques.'' Mattias Fink and Shira Broschat were made New Fellows at this meeting, and Tim Leighton was awarded the first Medwin Prize.

Robin Cleveland, Charlie Church, and Mark Schafer were nominated for B&B chair to succeed Carr Everbach; Robin won and begins his 3-year stint as chair at the Cancun meeting, 2–6 December 2002. Tyrone Porter agreed to be the B&B student council representative from the Pittsburgh through the Nashville meetings.

The Pittsburgh TC meeting in June 2002 was attended by 37 people. Special sessions were organized by Robin Cleveland, Constantin C. Coussios and Carr Everbach, Christy Holland and Mark Schafer, Brian Fowlkes, and Tom Matula. Mark Schafer was the TPOM representative. At the Pittsburgh meeting, B&B had 38 papers in five sessions; peaks of more than 41 attendees occurred at every session except for the last one (late Friday), which crested at 33.

In Pittsburgh, Robert E. Apfel received the Gold Medal of the Society and gave a moving acceptance speech. James Finneran and Tom Royston were each awarded R. Bruce Lindsay Awards.

The student paper contest in Pittsburgh was successful, with 11 entries and 13 judges. First prize went to Stanley Samuel (Univ. of Michigan Med. Ctr.) for his paper ''5pBB7: A robust roughness quantification technique using a standard imaging array transducer,'' and second place went to Yufeng Zhou (Duke Univ. ME, Mater. Sci, and Urologic Surgery) for ''5aBBa1: Reduction of tissue injury without compromising stone comminution in shock wave lithotripsy.'' Many thanks to all the students and judges for your efforts. B&B is slated to have student paper contests at both the Cancun and Nashville meetings.

Since the Fall 2002 ASA meeting is a special one, B&B is sponsoring special sessions in Cancun. These include ''Acoustic Microscopy,'' ''Lithotripsy,'' ''Ultrasound Mediated Drug Delivery and Gene Transfection,'' and a lecture on the history of Biomedical Ultrasound (sponsored by the Archives and History committee which rotates among the TCs). At the Nashville meeting (28 April–2 May 2003), B&B will sponsor special sessions on ''Acoustic Cavitation,'' ''Biomedical Applications of Image-Guided Ultrasonic Therapy,'' ''HIFU,'' and ''Audible-frequency Medical Diagnostic Methods.'' We hope to see you at these or other upcoming meetings.

E. CARR EVERBACH
Chair 1999–2002

Engineering Acoustics

The Engineering Acoustics Technical Committee (EATC) met at each of the two meetings of the Society, in Fort Lauderdale and in Pittsburgh.

In Newport Beach, the EATC sponsored two special sessions and two sessions of contributed papers. The special sessions and their organizers were

- Small Scale Acoustics and Acoustics Devices—Daniel Warren
- Ferroelectric Single Crystals: From Manufacture to Devices—Harold Robinson.

In addition, EATC was a cosponsor of the session ''Celebration of Miguel C. Junger's Contributions to Acoustics.'' The Committee offers many thanks to Elizabeth McLaughlin, who served as the representative to the Fort Lauderdale TPOM from the Committee.

At the Fort Lauderdale meeting, Ilene Busch-Vishniac was awarded the Silver Medal in Engineering Acoustics. Hearty congratulations to Ilene, who has been a longstanding member of EATC.

The EATC met in Fort Lauderdale on the evening of Tuesday, 4 December, with 27 members and friends participating.

In Pittsburgh, the EATC sponsored two special sessions and two sessions of contributed papers. The special sessions and their organizers were

- Electroacoustic Systems for 3-D Audio—Gary Elko
- Mine Hunting Sonar—Thomas Howarth.

The first of these special sessions was cosponsored with the Technical Committee on Architectural Acoustics. The Committee offers many thanks to William Thompson, who served as the representative to the Pittsburgh TPOM from the Committee.

At the Pittsburgh meeting, Tony Embleton was awarded the Gold Medal of the Society. Sincere congratulations to Tony.

The EATC met in Pittsburgh on the evening of Tuesday, 4 June, with 22 members and friends participating.

It was announced at the meeting that Stanley Ehrlich will be stepping down after serving for eighteen years as an Associate Editor of the Journal in technical areas relevant to the Committee. Sincere appreciation was expressed at the meeting for Stanley's service to the *Journal* and its readers. Stan is a former Chairman of this Committee and continues to serve the Committee and the Society in many ways. We also salute Allan Zuckerwar who will take over the duties of Associate Editor in many of the technical areas relevant to the Committee.

The Chair of EATC would like to thank all of the people who have worked to make the activities of EATC successful for this year, and to invite all interested parties to the EATC meetings to be held at future meetings of the Society.

STEPHEN C. THOMPSON
Chair

Musical Acoustics

2001–2002 was an active year for the Technical Committee on Musical Acoustics (TCMU). A number of successful special sessions were organized for the Ft. Lauderdale and Pittsburgh meetings, and a new translation project was launched using technical initiative funds.

At the December meeting in Ft. Lauderdale there were four special sessions in musical acoustics. Two were cosponsored by the Education Committee: a session on musical technology organized by Tom Rossing, and ''Musical Acoustics in Teaching Science, Engineering and Technology'' organized by Robert Collier and Carr Everbach. Paul Wheeler and Tom Rossing organized ''Caribbean Musical Instruments and Traditions.'' Tony Hoover and Richard Campbell were the organizers of a joint Architectural Acoustics–Musical Acoustics session on the integration of synthesis techniques with ''acoustic'' music.

The Pittsburgh meeting featured four special sessions in musical acoustics. These included a session on naturalness in synthesized speech and singing organized by Sten Ternström (cosponsored by Speech), a session on interactive computer music systems organized by Roger Dannenberg, and a session on music recognition techniques organized by Jim Beauchamp (cosponsored by P&P). The session ''Free Reed Instruments: Historical Perspectives and Recent Research,'' organized by Jim Cottingham, was followed by a concert by classical accordionist Henry Doktorski, who played for an enthusiastic full house audience of over 200. This concert was to have also included harmonica virtuoso Howard Levy, whose flight did not arrive in Pittsburgh in time for the event due to extensive weather delays. Many from the ASA meeting did get a chance to hear Levy perform later that evening at a local club.

Efforts to promote student involvement in musical acoustics continue through the student paper competition and joint sponsorship of student socials. Rachel Romond, who continues to serve as the Musical Acoustics representative on the ASA Student Council, was elected Chair of the Student Council beginning in 2003. There were five entrants in the student paper competition in Ft. Lauderdale and eight entrants in Pittsburgh. The winner of the Best Student Paper Award in Musical Acoustics at Ft. Lauderdale was Pantelis Vassilakis of UCLA, who presented ''Auditory roughness estimation of complex spectra—roughness degrees and dissonance ratings of harmonic intervals revisited.'' At Pittsburgh two student paper awards were given. The first place paper was ''Regimes of oscillation and reed vibrations in lingual organ pipes,'' presented by Eric Cox of Northern Illinois University. The second place award went to Michael Busha of Grinnell College, who presented ''Laboratory measurements on free reeds from the reed organ, accordion, and khaen.''

A technical initiative begun in 2000 and continuing into 2002 is a project conducted by Peter Hoekje involving preparation of a set of demonstrations in musical acoustics to be available on CD-ROM and on the web. A new technical initiative begun in 2002 is a project in which selected papers of current interest in musical acoustics are being translated into English. The editorial board for this is headed by Tom Rossing and includes Jim Beauchamp, Ian Lindevald, and Jim Cottingham.

The representatives to the Technical Program Organizing Meetings in 2001–2002 were Jim Cottingham for the Ft. Lauderdale meeting and Peter Hoekje for the Pittsburgh meeting. Representation from the TCMU on ASA committees includes Jim Beauchamp on the Medals and Awards Committee, Uwe Hansen on Membership Committee, and Ian Lindevald on the Committee on Standards. Neville Fletcher was appointed in December 2001 as a *JASA* Associate Editor for Musical Acoustics, replacing Dean Ayers. The other *JASA* Associate Editor for Musical Acoustics is Stephen McAdams, appointed at the Chicago meeting in June 2001.

At the Ft. Lauderdale meeting Colin Gough and Thomas D. Rossing were each presented the ASA Science Writing Award for Professionals. Newly elected Fellows of the Acoustical Society this year included René Caussé and Andrzej Rakowski. James P. Cottingham has been reelected Technical Committee Chair, to serve until 2005. Members appointed or reappointed this year to serve on the technical committee for terms expiring in 2005 include George A. Bissinger, Annabel J. Cohen, Diana Deutsch, Uwe J. Hansen, Roger J. Hanson, Peter L. Hoekje, James H. Irwin, Ian M. Lindevald, Stephen E. McAdams, Gary P. Scavone, and Gabriel Weinreich.

The web site for the TCMU continues to maintain current information on the technical committee and on musical acoustics in general at http://www.public.coe.edu/~jcotting/tcmu.

JAMES P. COTTINGHAM
Chair

Noise

The Technical Committee on Noise (TCN) has had a successful year! Although noise is still defined as ''unwanted sound,'' the interest of those in the Acoustical Society, and of the general public at large, in noise issues is strong and is growing. There seems to be a sense of purpose and urgency concerning noise and its control that is unprecedented in recent years. Per-

haps the awareness of noise has reached the critical mass needed to achieve real results. As always, your technical committee has been a vital exponent for noise awareness and noise control, in fulfillment of the ASA purpose ''to increase and diffuse the knowledge of acoustics and promote its practical application.'' More on that later.

The Technical Committee on Noise met twice in the past year: in the fall of 2001 in Ft. Lauderdale, FL, and in the spring of 2002 in Pittsburgh, PA. Both meetings had a great turnout.

The ASA meeting in Ft. Lauderdale was well attended, perhaps due to its attractive location. This was a successful meeting, with seven technical sessions primarily sponsored by TCN, and five additional sessions cosponsored with other technical committees. Nearly 80 papers were presented. Topics included classroom acoustics, noise measurement and control case studies, industrial noise prediction, cruise ship acoustics, community noise, and noise effects on marine animals. A retrospective session on Miguel Junger's work was held. In addition, Tony Embleton regaled us with his fascinating tutorial on noise propagation and prediction outdoors!

There are many people to thank for the success of the Ft. Lauderdale meeting, including our Technical Program Organizer, Mahlon Burkhard; TCN session chairs, John Erdreich, Joe Pope, Frank Brittain, Bennett Brooks, Nancy Timmerman, George Luz, and Angie Campanella; cosponsoring chairs Ira Dyer and David Feit (EA), Russ Cooper and Neil Shaw (AA), Darlene Ketten, Ed Gerstein, and Joe Blue (AB), and, of course, all of the paper presenters.

The ASA meeting in Pittsburgh offered a number of fine Noise sponsored sessions on a variety of topics including impulse and continuous noise effects, mining noise, glazing acoustics, noise policy, and noise control devices. TCN also cosponsored several sessions with Architectural, Structural, Speech and Animal Bioacoustics. Additionally, the Engineering and Structural committees sponsored several sessions that focused on flow-induced noise and active noise control. Thanks go to the Program Organizers, Dick Godfrey and John Seiler; TCN session chairs, John Seiler, Greg Tocci, Larry Finegold and Dick Godfrey; cosponsoring chairs Lily Wang (AA), Ahmet Selamet (EA), Tan-Fan Hwang (SA), and Scott Sommerfeldt (SA); and the nearly 70 paper presenters.

A special treat at the Pittsburgh meeting was the Plenary Session, at which our own Tony Embleton received the Gold Medal of the Society. This award was well deserved, and it was a pleasure to see Tony accept it. Congratulations Tony!

At the Pittsburgh meeting, TCN approved holding a joint meeting with the Institute of Noise Control Engineering (INCE) in Vancouver in 2005. This should be a great event, given the success of our last joint meeting in Newport Beach. In addition, our dedicated members have developed noteworthy special sessions and tutorials for the upcoming meetings in Cancun, Nashville, and Austin. These meetings promise to be very interesting.

As of this writing, the Standard on Classroom Acoustics was just approved for publication by ANSI. Everyone associated with this effort deserves hearty congratulations. This standard should help to contribute to quieter classrooms, and a more educationally friendly environment in our schools.

Over this past year TCN has had a fine group of candidates for the Outstanding Papers by Young Presenters award. I am pleased to announce that Wendy Sanders of the University of Michigan received the Outstanding Paper award for her paper ''Turbulent boundary layer pressure fluctuations at large scales,'' which was presented at the Pittsburgh meeting. Nice work, Wendy!

Technical Initiative funding for TCN activities continued in the areas of the younger presenter awards, web page maintenance, the student reception, the development of the online library (see www.nonoise.org), funds to distribute the classroom acoustics booklet in PDF format, and support for the National Hearing Conservation Association (NHCA).

The TCN Student Representative to the ASA Student Council is Jennifer Cooper. The Student Council has been very active, encouraging participation by our younger members, who are the future of this organization.

As mentioned above, this year has seen an upsurge in general interest in noise. One effort that is gaining momentum is the assessment and revision of noise policy. This effort is being undertaken by an INCE study group, with the goal of reaching a national forum later this year. Also, there seems to be increased attention to noise on the part of the media. Your humble TCN Chair has been called upon, recently, to discuss noise issues for a wide variety of media outlets, including National Public Radio, the New York Times, and abcNEWS.com, to name a few. A goal of the TCN Task Force on noise control and the ASA was to gain greater awareness of noise issues by the public, and perhaps now it is really happening.

On a sad note, this year we mark the untimely passing of Dan Johnson. Dan made many important and lasting contributions to our Society. Among his accomplishments was his extensive work in the standards arena, including his service as Standards Director, and he was a past Chair of this committee. Dan was a mentor to many in the Society, including myself, and he will be greatly missed.

The TCN is fortunate to enjoy the enthusiastic support of its members. This fine group never fails to deliver the excellent technical work and personal commitment that we have come to rely upon. Because of you, we look forward to a bright future.

BENNETT M. BROOKS
Chair

Physical Acoustics

It's been another fine year for the Physical Acoustics Technical Committee (PATC), thanks in large part to the people who volunteer to do the jobs that need to be done. For the Ft. Lauderdale meeting, Julian Maynard and Joseph Gladden organized the special session ''Applications of Acoustic Resonance,'' and Anthony Atchley and Victor Sparrow put together ''High Amplitude Effects in Resonators.'' For the Pittsburgh meeting, Keith Wilson gave us ''Recent Developments in Wave Propagation in Random and Complex Media,'' and Matthew Poese organized a herd of students and some geezers to present ''Demonstration Devices in Physical Acoustics.'' We'll have four special sessions in Cancun thanks to Henry Bass, Glynn Holt, Todd Murray, Joseph Turner, and Richard Weaver. Glynn Holt, Richard Raspet, and James Sabatier who will bring us three more special sessions in Nashville. These sessions plus all the bread-and-butter contributed sessions are put together into a coherent meeting by our TPOM representatives, who were James Chambers and Kerry Commander for Ft. Lauderdale, and Bart Lipkens and James Sabatier for Pittsburgh. Sameer Madanshetty, Steven Garrett, and James Sabatier continue to serve, respectively, as our representatives to the Standards, Membership, and Medals and Awards Committees. Preston Wilson has been our representative to the Student Council. I suspect that the people with the most work of all are the Associate Editors. Some of those associated with physical acoustics, past and present, are Henry Bass, Yves Berthelot, Mack Breazeale, Dale Chimenti, Mark Hamilton, John Harris, Michael Howe, Steven Kargl, Philip Marston, Martin Ochmann, Ronald Roy, Victor Sparrow, Louis Sutherland, and Lonny Thompson. There must be others who have contributed significantly to our Technical Committee (TC) that I have overlooked who get my apologies as well as my thanks.

Some external events have touched our committee. Soon after the World Trade Center (WTC) collapse, Thomas Gabrielson, Matthew Poese, Anthony Atchley, and Thomas Donnellan rushed in to look for survivors using acoustical techniques. James Sabatier was flown in by Army helicopter to use his landmine detection equipment to monitor the structural health of the heavily damaged WTC Four building. Anthony Atchley described these efforts at a Hot Topics presentation at Ft. Lauderdale. Also, a report in *Science* of evidence for nuclear fusion in collapsing sonoluminescencent bubbles generated quite a stir with several members of our committee. Larry Crum gave a lively talk on the subject at our Pittsburgh Technical Committee meeting, and a special session is planned for Nashville.

Congratulations go to Jin Liu who won the first prize Best Student Paper Award at Pittsburgh for her presentation ''Absence of excess attenuation in Celcor ceramic at low frequencies,'' and Joseph Gladden who won second prize with ''Thin film characterization using resonant ultrasound spectroscopy.''

We welcome everyone, especially students and new members, to our Technical Committee meetings held on the Tuesday evening of each ASA meeting. At least 64 people attended the TC meeting in Ft. Lauderdale and at least 77 were there at Pittsburgh. Also at Pittsburgh we had a very nice reception for students in physical acoustics at the Church Brew Works a few miles from the meeting site, attended by 22 students and 12 nonstudents. Technical Initiative funds were used this past year to support the social, the student paper awards, and travel for a few invited speakers.

The retirement of Logan Hargrove, the long time patron of physical acoustics at the Office of Naval Research (ONR), prompts much concern in our technical committee. At this point, it does not appear that ONR will be

immediately refilling his position. Many of us will sorely miss the good work that Logan has done over many years to support physical acoustics—and not just financially. Thanks, Logan.

Well, my time as the PATC chair is over. I was surprised by the amount of fun I had doing it. I hope Tom Matula, our new chair, has at least as good a time with it, and he is sure to do a great job.

ROBERT M. KEOLIAN
Chair 1999–2002

Psychological and Physiological Acoustics

In keeping with the policy of P&P to deemphasize the fall meetings, the Ft. Lauderdale meeting was sparsely attended by P&P members: There was only one P&P session. The TC Chair did not attend but was more than ably represented by Donna Neff, former TC Chair and current member of the Executive Council.

The spring meeting in Pittsburgh was well attended and there was considerable P&P activity: Six full podium or poster sessions, three special sessions, and one ''Hot Topic'' session presented by Bill Yost. In addition, there were three sessions that were cosponsored by P&P. Thanks to Brent Edwards and Dianne Van Tasell for organizing the session on hearing aid design and to Barbara Shinn-Cunningham for coorganizing (with AA) the session on virtual rooms. Both of these special sessions were interesting, well attended, and, by all reports, successful. Thanks also to Jan Weisenberger and Judith Lauter for organizing the special session honoring Ira Hirsh. Larry Feth organized the sessions and is thanked for accomplishing this difficult task with few conflicts.

The P&P open meeting was, in contrast to some previous meetings, well controlled and accomplished its modest, but important, goals. After acknowledging the efforts of the local organizing committee, especially those of the other Rich Stern, things became subdued when Bob Frisina and Erv Hafter eloquently helped us with the deaths of two promising P&P youngsters, Evan Relkin and Brian Shelton.

There was brief discussion of special sessions for the Nashville meeting (spring 2003), and five sessions, including three ''named'' sessions, were presented to the Technical Council on Friday, 7 June, and approved. In keeping with P&P policy, there are no special sessions planned for the fall meetings in Cancun (2002) or in Austin (2003). The P&P Associate Editors gave their reports and there appear to be no major changes in editorial activity. The TC Chair spared the audience of approximately 100 from an extended spiel about the importance of the editorial process and the efforts of the Associate Editors in maintaining the quality and rigor of JASA and that, in the opinion of the TC Chair, JASA is the single most important activity of ASA and that we have to maintain our well-earned status as the premier venue for scientific communication in psychological acoustics. It was an unfortunate omission, especially because it's all true and because the Associate Editors deserve far more recognition than their humble title suggests. Prior to the Pittsburgh meeting, Ginny Richards was elected by the TC to become the chair of P&P. The announcement of her replacing the current TC Chair was greeted with applause. It was not clear what was being applauded, however. Six new members of the TC were elected during the meeting: Søren Buus, Michelle Hicks, Lynne Marshall, Chris Plack, Don Sinex, and Magda Wojtczak. Their terms begin after the meeting in Spring 2003 (Nashville). Thanks to the people who agreed to be on the ballot and to the outgoing members, who prepared the ballot: Ann Clock Eddins (Election Chair), Huanping Dai, Dick Fay, Peggy Nelson, Bert Schlauch, and Stan Sheft. During the plenary session Bill Hartmann (mostly P&P) was acknowledged for his fine service as President of ASA. Congratulations to the most recent P&P Fellows: Rachel Clifton, Birger Kollmeier, David Marsh, Andrzej Rakowski, Christoper Shera, Andrea Simmons, and Lynne Werner. Neal Viemeister received the Silver Medal in Psychological and Physiological Acoustics and joins an awesome set of five predecessors. His acceptance speech, although fairly stylish, merits only a grade of B, primary because he did not adequately acknowledge the importance of the scientific environment provided by ASA and, even more importantly, that of his precious friends and colleagues in ASA.

P&P continues to be well represented in ASA. Jan Weisenberger's term as vice-president has ended; Bill Yost replaces her. Donna Neff continues her service on the Executive Council. Our thanks to all three. Finally, P&P continues with its standard Technical Initiatives, which include travel support for invited speakers, student receptions, and homepage maintenance. Suggestions for innovative uses for these funds, such as workshops, satellite meetings, etc., are welcome (richards@cattell.psych.upenn.edu).

NEAL VIEMEISTER
Chair 1999–2002

Structural Acoustics and Vibration

Most of the activity of the Structural Acoustics and Vibration technical committee took place at the regular meetings of the ASA. At the Ft. Lauderdale meeting (Fall 2001), the Structural Acoustics and Vibration TC sponsored three special sessions. ''Methods of passive dissipation in structural acoustics'' was organized by Courtney Burroughs of Penn State University, and focused on analysis and measurement of damping in structures, with some focus on composite structures. ''Numerical techniques in structural acoustics'' was organized by Joe Cuschieri of Florida Atlantic University, and reviewed some of the recent developments in numerical methods. ''A celebration of the career of Miguel Junger'' was organized by Ira Dyer and David Feit. This session was sponsored by five technical committees, which gives an indication of how prolific the career of Miguel Junger has been. The session was very informative from a historical perspective, as well as providing recent results that have succeeded the work of Miguel Junger. The TC also sponsored a Student Paper Award competition. The first place paper at Ft. Lauderdale was given by Joseph Gregory from North Carolina State University. The title of his paper was ''Experimental statistical energy analysis in the time domain.'' The second place paper was given by Andrei Zagrai from the University of South Carolina. The title of his paper was ''The electro-mechanical impedance method for damage identification in circular plates.'' There were again three special sessions sponsored by the TC at the Pittsburgh meeting (Spring 2002). ''Pyroshock I & II'' were organized by Harry Himelblau from Rocketdyne, and featured many of the leading researchers in the area of pyroshock. The session provided a nice review of recent developments in this research field. ''Flow-induced vibration and noise'' was organized by Yan-Fan Hwang from Penn State University and was a session of interest to a number of attendees. ''Active control of ducted fan noise and other active control issues'' was organized by Scott Sommerfeldt from Brigham Young University and featured a number of papers covering various aspects of active control. In addition to these sessions, there was a special session sponsored by Archives and History and SAV entitled ''History of structural acoustics and vibration.'' This special lecture was given by David Feit, Murray Strasberg (both form Naval Surface Warfare Center), and Eric Ungar (from Acentech, Inc.) This lecture provided a very informative and well-presented overview of the important developments in our field. The Student Paper Award competition was again sponsored, and we would like to thank Greg McDaniel of Boston University for his efforts with this competition. The first place paper at Pittsburgh was given by Manmohan Moondra from Wayne State University. The title of his paper was ''HELS based near-field acoustic holography for a highly nonspherical structure.'' The second place paper was given by Kent L. Gee from Brigham Young University. The title of his paper was ''Multi-channel active control of axial cooling fan noise.'' We would like to thank all of the students who have participated in the past two meetings for the high-quality presentations, and for their contributions to our meetings.

We note with sadness the recent passing of Dan Johnson. Dan was active in the standards efforts within ASA, and his contributions will be deeply missed. We continue to have four associate editors for structural acoustics: Jerry Ginsberg, Andrew Norris, Richard Weaver, and Earl Williams. In addition, Paul Remington is serving as an associate editor for ARLO in structural acoustics. Sabih Hayek has served as our representative for structural acoustics on ASA's Medals and Awards Committee, but was recently replaced by Mauro Pierruci. Courtney Burroughs continues to serve as our representative on the Membership Committee, and his efforts are also appreciated. Courtney also served as the general chair for the Pittsburgh meeting, and his efforts in contributing to a very successful meeting in Pittsburgh are appreciated.

SCOTT D. SOMMERFELDT
Chair

Underwater Acoustics

The UWTC representatives to the Ft. Lauderdale Technical Program Organizing Meeting (TPOM) were Ellen Livingston and John Perkins. Because of a number of special circumstances this meeting held an unusually high number of Special Sessions that were of interest to the underwater acoustics community. First, working with the Archives and History Committee, Ralph Goodman organized a Special Session on Insights into the History of Underwater Acoustics. For this Session we had a number of distinguished speakers who have been involved with significant achievements in underwater acoustics. A second Special Session, on Ocean Research and Marine Mammal Regulatory Issues, was organized by David Bradley. This Session was cosponsored by Acoustical Oceanography and Animal Bioacoustics. In conjunction with this Session, the UWTC was also a co-sponsor on a Special Session on Acoustics and Marine Mammals. This Session concluded with a lively panel discussion on ASA's Role in Marine Animal Bioacoustics. Kevin LePage organized a Special Session on Reverberation in Shallow Water. This Session had so many contributed papers that it lasted an entire day. As if this were not enough, the UWTC was also cosponsor of three other Special Sessions: (1) A Celebration of Miguel C. Junger's Contributions to Acoustics (led by the Engineering Acoustics Committee), (2) Theory and Applications of Acoustic Time Reversal (led by the Signal Processing Committee), and (3) Acoustic Image Reconstruction (led by the Signal Processing Committee).

The Underwater Acoustics Technical Committee and the Acoustical Oceanography Technical Committee have agreed to form a joint Subcommittee to explore ways to minimize the problems we had at Ft. Lauderdale with an abundance of overlapping sessions that are of interest to both groups. At the Ft. Lauderdale meeting, the UWTC joined with the other Technical Committees to sponsor a student luncheon. This did not seem to work out as well as the evening receptions we have held in the past—mainly because the food ran out very quickly! The UWTC also held a student paper competition. The winners were Ben Dzikowicz, Karim Sabra, and Purnima Ratilal. The chairman of the judges committee was Mohsen Badiey.

David Bradley and John Perkins represented the UWTC at the Pittsburgh TPOM. This meeting featured a Special Session in honor of Morris Schulkin. This Session was organized by Robert Spindel. The UWTC was also a cosponsor of a Special Session in honor of Ed Hamilton. The Acoustical Oceanography Committee was the lead for this Session. The UWTC was also cosponsor of three other Special Sessions: (1) Acoustical Studies of Bubbles in the Ocean (led by the Acoustical Oceanography Committee), (2) Recent Developments in Wave Propagation in Random and Complex Media (led by the Physical Acoustics Committee), and (3) Mine Hunting Sonar (led by the Engineering Acoustics Committee).

As for student activities at Pittsburgh, we cosponsored an evening reception with a group of Technical Committees. The food might not have been as good as at some other meetings, but it was a large and lively group. The UWTC also held a student paper competition in Pittsburgh. The winners were Luc Lenain and Yi-san Lai.

Looking forward to the Cancun meeting, we have four Special Sessions planned: (1) Concurrent Mapping and Navigation organized by John Leonard, (2) GeoClutter and Boundary Characterization organized by Nicholas Makris and Charles Holland, (3) Littoral Environmental Variability organized by Finn Jensen, Peter Dahl, and Jim Lynch, and (4) Underwater Acoustic Measurement Laboratories organized by Carlos Ranz-Guerra. In addition to these sessions, the UWTC is also cosponsoring several sessions being organized by other Technical Committees.

JOHN PERKINS
Chair

Regional Chapter News

Cincinnati Chapter: The Chapter presented three students with Dr. Daniel W. Martin Acoustic Awards at the Ohio State Science Day in May 2002. The cash awards totaled $500. The winners also received ribbons, certificates, and sets of tuning forks.

First Prize went to Chad Koenig, 11th grade, for his comprehensive comparison of insulating material and equipment modification to reduce machinery noise. His excellent presentation was titled ''Ah... That's Better.''

Second Prize was awarded to 10th grader Mathew Anker for ''The Acoustic Properties of Electrostatic Loudspeakers.'' Mr. Anker also won a Martin Acoustics Award in 2001.

Third Prize winner was Courtney Sniderman, 8th grade, for ''The Sound of Music.''

Judges were Sharon T. Hepfner and Sarah Seiger.

SHARON HEPFNER

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.

2002

16–20 Sept.	7th International Conference on Spoken Language Processing (ICSLP 2002) Interspeech 2002 [http://www.icslp2002.org/home.html].
2–6 Dec.	First Pan-American/Iberian Meeting on Acoustics (Joint Meeting: 144th Meeting of the Acoustical Society of America, 3rd Iberoamerican Congress of Acoustics, and 9th Mexican Congress on Acoustics), Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org/cancun/cancun.html].

2003

28 April–2 May	145th Meeting of the Acoustical Society of America, Nashville, TN [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
5–8 May	SAE Noise & Vibration Conference & Exhibition, Traverse City, MI [P. Kreh, SAE International, 755 W. Big Beaver Rd., Suite 1600, Troy, MI 48084; Fax: 724-776-1830; WWW: http://www.sae.org].
10–14 Nov.	146th Meeting of the Acoustical Society of America, Austin, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

2004

24–28 May	75th Anniversary Meeting (147th Meeting) of the Acoustical Society of America, New York, NY [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
22–26 Nov.	148th Meeting of the Acoustical Society of America, San Diego, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

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

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

Some indexes are out of print as noted below.

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

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

Volumes 21–30, 1949–1958: JASA, Contemporary Literature, and Patents. Classified by subject and indexed by author and inventor. Pp. 952. Price: ASA members $20; Nonmembers $75.
Volumes 31–35, 1959–1963: JASA, Contemporary Literature, and Patents. Classified by subject and indexed by author and inventor. Pp. 1140. Price: ASA members $20; Nonmembers $90.
Volumes 36–44, 1964–1968: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 485. Out of Print.
Volumes 36–44, 1964–1968: Contemporary Literature. Classified by subject and indexed by author. Pp. 1060. Out of Print.
Volumes 45–54, 1969–1973: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 540. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).
Volumes 55–64, 1974–1978: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 816. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).
Volumes 65–74, 1979–1983: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 624. Price: ASA members $25 (paperbound); Nonmembers $75 (clothbound).
Volumes 75–84, 1984–1988: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 625. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).
Volumes 85–94, 1989–1993: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).
Volumes 95–104, 1994–1998: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 632. Price: ASA members $40 (paperbound); Nonmembers $90 (clothbound).

Members of Technical and Administrative Committees of the Acoustical Society of America

The Technical and Administrative Committees listed below have been appointed by the Executive Council. These appointments, with such changes as may be made by the President from time to time, will be in effect until the Spring meeting of the Society in 2003.

Technical Committees 2002–2003

Acoustical Oceanography

	Term to
Peter F. Worcester, *Chair* to 2004	2004
Daniela Di Iorio	2005
Matthew A. Dzieciuch	2005
Gary J. Heald	2005
Daniel L. Hutt	2005
Timothy G. Leighton	2005
James H. Miller	2005
Carlos J. Robinson	2005
Daniel Rouseff	2005
Emmanuel K. Skarsoulis	2005
Jerome A. Smith	2005
Eric J. Terrill	2005
Gang Yuan	2005
Ching-Sang Chiu	2004
Bruce D. Cornuelle	2004
Orest I. Diachok	2004
Kenneth G. Foote	2004
Andrew M. Forbes	2004
Boris G. Katsnelson	2004
James F. Lynch	2004
Jungyul Na	2004
Gopu R. Potty	2004
Stephen A. Reynolds	2004
Mark V. Trevorrow	2004
Mohsen Badiey	2003
N. Ross Chapman	2003
Timothy F. Duda	2003
Oleg A. Godin	2003
Bruce M. Howe	2003
Iwao Nakano	2003
Marshall H. Orr	2003
David R. Palmer	2003
John R. Potter	2003
Hanne Sagen	2003
Alexandra I. Tolstoy	2003
Peter Traykovski	2003
Svein Vagle	

Ex officio:
John C. Burgess, Associate Editor of JASA
William M. Carey, Associate Editor of JASA
William L. Siegmann, Associate Editor of JASA
Ralph A. Stephen, Associate Editor of JASA
Michael J. Buckingham, member of Medals and Awards Committee
Christopher Feuillade, member of Membership Committee
James H. Miller, member of ASACOS

Animal Bioacoustics

	Term to
Mardi C. Hastings, *Chair* to 2003	2003
William C. Burgess	2005
James J. Finneran	2005
Adam S. Frankel	2005
Darlene R. Ketten	2005
Thomas Norris	2005
Larry L. Pater	2005
Peter L. Tyack	2005
Whitlow W. L. Au	2004
Robert J. Dooling	2004
Robert Hickling	2004
David A. Helweg	2004
Peter M. Narins	2004
Andrea M. Simmons	2004
James A. Simmons	2004
Ann E. Bowles	2003
John R. Buck	2003
Christopher W. Clark	2003
William C. Cummings	2003
Charles R. Greene	2003
Lynette A. Hart	2003
D. Vance Holliday	2003
Richard W. Mankin	2003
Sam H. Ridgway	2003

Ex officio:
Whitlow W. L. Au, Associate Editor of JASA and member of Medals and Awards Committee
William M. Carey, Associate Editor of JASA
Stanley A. Chin-Bing, Associate Editor of JASA
William L. Siegmann, Associate Editor of JASA
Ralph A. Stephen, Associate Editor of JASA
Darlene R. Ketten, member of Membership Committee
Ann E. Bowles, member of ASACOS

Architectural Acoustics

	Term to
K. Anthony Hoover, *Chair* to 2004	2004
Christopher N. Blair	2005
John S. Bradley	2005

Christopher N. Brooks	2005
Angelo J. Campanella	2005
Quinsan Ciao	2005
Robert C. Coffeen	2005
Peter D'Antonio	2005
Felicia Doggett	2005
Timothy J. Foulkes	2005
Richard D. Godfrey	2005
Clare H. Hurtgen	2005
J. Christopher Jaffe	2005
Mendel Kleiner	2005
David Lubman	2005
Ralph T. Muehleisen	2005
Michael T. Nixon	2005
Jack E. Randorff	2005
H. Stanley Roller	2005
Brigitte Schulte-Fortkamp	2005
Noral D. Stewart	2005
Ning Xiang	2005
Bennett M. Brooks	2004
Steven M. Brown	2004
Richard H. Campbell	2004
M. David Egan	2004
Jesse J. Ehnert	2004
Donna A. Ellis	2004
Ronald R. Freiheit	2004
Richard M. Guernsey	2004
Mark A. Holden	2004
Dana S. Hougland	2004
Jerald R. Hyde	2004
David W. Kahn	2004
Martha M. Larson	2004
Gary S. Madaras	2004
Charles T. Moritz	2004
Paul B. Ostergaard	2004
Dennis A. Paoletti	2004
Neil A. Shaw	2004
Richard H. Talaske	2004
Ewart A. Wetherill	2004
George E. Winzer	2004
Michael R. Yantis	2004
Leslie D. Blomberg	2003
David Braslau	2003
Paul T. Calamia	2003
William J. Cavanaugh	2003
Daniel Clayton	2003
Murray R. Hodgson	2003
Bertram Y. Kinzey, Jr.	2003
Richard J. Peppin	2003
Scott D. Pfeiffer	2003
Roy L. Richards	2003
Kenneth P. Roy	2003
Ludwig W. Sepmeyer	2003
Gary W. Siebein	2003
Louis C. Sutherland	2003
Gregory C. Tocci	2003
Lily M. Wang	2003

Ex officio:
Jerry H. Ginsberg, Associate Editor of JASA
Mendel Kleiner, Associate Editor of JASA
Andrew N. Norris, Associate Editor of JASA
Richard L. Weaver, Associate Editor of JASA
Earl R. Williams, Associate Editor of JASA
Ewart A. Wetherill, member of Medals and Awards Committee
Gregory C. Tocci, member of Membership Committee
George E. Winzer, member of ASACOS

Biomedical Ultrasound/Bioresponse to Vibration

	Term to
Robin O. Cleveland, *Chair* to 2005	2005
John S. Allen	2005
Michael R. Bailey	2005
Charles C. Church	2005
Gregory Clement	2005
Floyd Dunn	2005
E. Carr Everbach	2005
Christy K. Holland	2005
R. Glynn Holt	2005
James A. McAteer	2005
Wesley L. Nyborg	2005
Mark E. Schafer	2005
Diane Dalecki	2004
J. Brian Fowlkes	2004
Seyed H. R. Hosseini	2004
Kullervo Hynynen	2004
James F. Greenleaf	2004
T. Douglas Mast	2004
Thomas J. Matula	2004
Douglas L. Miller	2004
Sandra A. Poliachik	2004
Ronald A. Roy	2004
Thomas J. Royston	2004
Pei Zhong	2004
Shira L. Broschat	2003
Ibrahim M. Hallaj	2003
Peng Jiang	2003
Joie P. Jones	2003
Inder R. Makin	2003
Janet M. Weisenberger	2003
Junru Wu	2003
Suk Wang Yoon	2003

Ex officio:
Laurel H. Carney, Associate Editor of JASA
Floyd Dunn, Associate Editor of JASA
Neal F. Viemeister, Associate Editor of JASA
Wesley L. Nyborg, member of the Medals and Awards Committee
Anthony J. Brammer, member of the Membership Committee
E. Carr Everbach, member of ASACOS

Engineering Acoustics

	Term to
Stephen C. Thompson, *Chair* to 2003	2003
Steven R. Baker	2005
David A. Brown	2005
Stephen C. Butler	2005
Robert D. Corsaro	2005
Stephen E. Forsythe	2005
Brian H. Houston	2005
W. Jack Hughes	2005
Robert M. Koch	2005
L. Dwight Luker	2005
Arnie L. Van Buren	2005
Daniel M. Warren	2005
Joseph F. Zalesak	2005
Mahlon D. Burkhard	2004

James Christoff	2004
Fernando Garcia-Osuna	2004
Charles S. Hayden	2004
Dennis F. Jones	2004
Jan F. Lindberg	2004
Yushieh Ma	2004
Elizabeth A. McLaughlin	2004
Alan Powell	2004
Roger T. Richards	2004
Harold C. Robinson	2004
Kenneth D. Rolt	2004
Neil A. Shaw	2004
James F. Tressler	2004
Kim C. Benjamin	2003
Stanley L. Ehrlich	2003
Gary W. Elko	2003
Robert D. Finch	2003
Guillermo C. Gaunaurd	2003
Dehua Huang	2003
Sung Hwan Ko	2003
Theodore J. Mapes	2003
Victor Nedzelnitsky	2003
Yongrae Roh	2003
Ahmet Selamet	2003
James E. West	2003
George S. K. Wong	2003

Ex officio:
Andrew J. Szeri, Associate Editor of JASA
Allan J. Zuckerwar, Associate Editor of JASA
Mahlon D. Burkhard, member of Medals and Awards Committee and member of ASACOS
Thomas R. Howarth, member of Membership Committee

Musical Acoustics

	Term to
James P. Cottingham, *Chair* to 2005	2005
George A. Bissinger	2005
Annabel J. Cohen	2005
Diana Deutsch	2005
Uwe J. Hansen	2005
Roger J. Hanson	2005
Peter L. Hoekje	2005
James H. Irwin	2005
Ian M. Lindevald	2005
Steven E. McAdams	2005
Gary P. Scavone	2005
Gabriel Weinreich	2005
R. Dean Ayers	2004
Judith C. Brown	2004
Courtney B. Burroughs	2004
Robert D. Collier	2004
George F. Emerson	2004
Bozena Kostek	2004
Barry Larkin	2004
Daniel O. Ludwigsen	2004
Thomas D. Rossing	2004
Julius O. Smith	2004
William J. Strong	2004
Anders G. Askenfelt	2003
James W. Beauchamp	2003
Xavier Boutillon	2003
Murray D. Campbell	2003
Rene E. Causse	2003
Nicholas J. Giordano	2003
J. M. Harrison	2003
William M. Hartmann	2003
Norman C. Pickering	2003
James M. Pyne	2003
Daniel A. Russell	2003
Punita G. Singh	2003
Paul A. Wheeler	2003
Shigeru Yoshikawa	2003

Ex officio:
John C. Burgess, Associate Editor of JASA
Neville H. Fletcher, Associate Editor of JASA
Stephen E. McAdams, Associate Editor of JASA
Allan J. Zuckerwar, Associate Editor of JASA
James W. Beauchamp, member of Medals and Awards Committee
Uwe J. Hansen, member of Membership Committee
Ian M. Lindevald, member of ASACOS

Noise

	Term to
Bennett M. Brooks, *Chair* to 2003	2003
Elliott H. Berger	2005
Ann E. Bowles	2005
Frank H. Brittain	2005
Steven M. Brown	2005
Mahlon D. Burkhard	2005
Sanford Fidell	2005
Lawrence S. Finegold	2005
Samir N. Y. Gerges	2005
Richard D. Godfrey	2005
Lee D. Hager	2005
Murray R. Hodgson	2005
Jerry G. Lilly	2005
David Lubman	2005
George A. Luz	2005
Luc Mongeau	2005
Michael T. Nixon	2005
Matthew A. Nobile	2005
Richard J. Peppin	2005
Robert A. Putnam	2005
Jack E. Randorff	2005
Stephen I. Roth	2005
Kenneth P. Roy	2005
Paul D. Schomer	2005
Brandon D. Tinianov	2005
Nancy S. Timmerman	2005
Gregory C. Tocci	2005
Martin Alexander	2004
John P. Barry	2004
Leo L. Beranek	2004
Arno S. Bommer	2004
Anthony J. Brammer	2004
James O. Buntin	2004
Robert J. Cook	2004
Jim R. Cummins	2004
Kenneth A. Cunefare	2004
Joseph M. Cuschieri	2004
Paul R. Donavan	2004
William D. Gallagher	2004
Gerald C. Lauchle	2004
Miomir Mijic	2004
Thomas Norris	2004
John P. Seiler	2004
Noral D. Stewart	2004
Michael R. Stinson	2004

Alice H. Suter	2004
Louis C. Sutherland	2004
Jiri Tichy	2004
Dennis C. Walton	2004
Keith Attenborough	2003
Martin J. Beam	2003
Sergio Beristain	2003
Tarun K. Bhatt	2003
Leslie D. Blomberg	2003
Robert D. Bruce	2003
John C. Burgess	2003
Angelo J. Campanella	2003
William J. Cavanaugh	2003
Brian V. Chapnik	2003
Kejian Chen	2003
Kuo-Tsai Chen	2003
Li Cheng	2003
Raymond Cheng	2003
Robert J. Comparin	2003
James T. Cowling	2003
Malcolm J. Crocker	2003
Jaime Delannoy	2003
T. James DuBois	2003
John J. Earshen	2003
Jesse J. Ehnert	2003
Douglas K. Eilar	2003
Tony F. W. Embleton	2003
John Erdreich	2003
William J. Gastmeier	2003
Robert D. Hellweg	2003
David K. Holger	2003
Robert M. Hoover	2003
Patrick M. Hurdle	2003
Kazuo Ikegaya	2003
Jian Kang	2003
Seong-Woo Kang	2003
Tim Kelsall	2003
Tor S. Kihlman	2003
Sonoko Kuwano	2003
Mark A. Lang	2003
Chantal Laroche	2003
Yun-Hui Liu	2003
Robert Lotz	2003
Bjorn E. L. Lundquist	2003
Alan H. Marsh	2003
Henk M. E. Miedema	2003
Ruisen Ming	2003
Ikuharu Morioka	2003
Ralph Muehleisen	2003
Georgy L. Ossipov	2003
Kenneth J. Plotkin	2003
Joseph Pope	2003
Manuel Recuero	2003
Brigitte Schulte-Fortkamp	2003
Ben H. Sharp	2003
Lawrence Shotland	2003
Benjamin Soenarko	2003
Scott D. Sommerfeldt	2003
Kerrie G. Standlee	2003
Jing Tian	2003
Jean Tourret	2003
Ana M. Verzini	2003
Paul T. Weirich	2003
Kwangsee Allen Woo	2003
Barry R. Wyerman	2003
Serdar H. Yonak	2003
Robert W. Young	2003
James C. Yu	2003

Ex officio:
D. Keith Wilson, Associate Editor of JASA
Allan J. Zuckerwar, Associate Editor of JASA
Michael R. Stinson, member of Medals and Awards Committee
John Erdreich, member of Membership Committee
Richard J. Peppin, member of ASACOS

Physical Acoustics

	Term to
Thomas J. Matula, *Chair* to 2005	2005
Robert T. Beyer	2005
Robin O. Cleveland	2005
Lawrence A. Crum	2005
Kenneth E. Gilbert	2005
Robert A. Hiller	2005
R. Glynn Holt	2005
Bart Lipkens	2005
Ralph T. Muehleisen	2005
Harry Simpson	2005
John S. Stroud	2005
Preston S. Wilson	2005
David T. Blackstock	2004
David A. Brown	2004
John A. Burkhardt	2004
Kerry W. Commander	2004
Bruce C. Denardo	2004
Logan E. Hargrove	2004
D. Kent Lewis	2004
Julian D. Maynard	2004
George Mozurkewich	2004
Lev A. Ostrovsky	2004
Andrea Prosperetti	2004
Neil A. Shaw	2004
Victor W. Sparrow	2004
Richard Stern	2004
Roger M. Waxler	2004
Henry E. Bass	2003
Yves H. Berthelot	2003
James P. Chambers	2003
David I. Havelock	2003
Murray S. Korman	2003
Philip L. Marston	2003
Philip S. Spoor	2003
Larry A. Wilen	2003
D. Keith Wilson	2003
Wayne M. Wright	2003
Evgenia A. Zabolotskaya	2003

Ex officio:
Yves H. Berthelot, Associate Editor of JASA
Dale E. Chimenti, Associate Editor of JASA
Floyd Dunn, Associate Editor of JASA
Mark F. Hamilton, Associate Editor of JASA
John G. Harris, Associate Editor of JASA
Michael S. Howe, Associate Editor of JASA
Martin Ochmann, Associate Editor of JASA
Richard Raspet, Associate Editor of JASA
Andrew J. Szeri, Associate Editor of JASA
Louis C. Sutherland, Associate Editor of JASA
Lonny L. Thompson, Associate Editor of JASA
James M. Sabatier, member of Medals and Awards Committee
Steven L. Garrett, member of Membership Committee
Sameer I. Madanshetty, member of ASACOS

Psychological and Physiological Acoustics

	Term to
Virigina M. Richards, *Chair* to 2005	2005
John F. Culling	2005
Torsten Dau	2005
Armin Kohlrausch	2005
Robert A. Lutfi	2005
Christine R. Mason	2005
Brian C. J. Moore	2005
Bruce G. Berg	2004
Joan M. Besing	2004
Monita Chatterjee	2004
Robert D. Frisina	2004
Andrew J. Oxenham	2004
Pierre L. Divenyi	2003
Richard L. Freyman	2003
Wesley L. Grantham	2003
Barbara G. Shinn-Cunningham	2003
Patrick M. Zurek	2003

Ex officio:
Leslie R. Bernstein, Associate Editor of JASA
Laurel H. Carney, Associate Editor of JASA
Marjorie R. Leek, Associate Editor of JASA and member of Medals and Awards Committee
Brenda L. Lonsbury-Martin, Associate Editor of JASA
Neal F. Viemeister, Associate Editor of JASA
Ervin R. Hafter, member of Membership Committee
Peggy B. Nelson, member of ASACOS

Signal Processing in Acoustics

	Term to
James V. Candy, *Chair* to 2003	2003
David H. Chambers	2005
David J. Evans	2005
Howard A. Gaberson	2005
Charles F. Gaumond	2005
David I. Havelock	2005
Jean-Pierre Hermand	2005
George E. Ioup	2005
Juliette Ioup	2005
Sean K. Lehman	2005
Zoi-Heleni Michalopoulou	2005
Joseph Pope	2005
Leon H. Sibul	2005
Randall W. Smith	2005
James E. West	2005
Gary R. Wilson	2005
Ning Xiang	2005
Leon Cohen	2004
Jose A. Diaz	2004
Deborah M. Grove	2004
David I. Havelock	2004
Alan W. Meyer	2004
Richard J. Ruhala	2004
Roger W. Schwenke	2004
David C. Swanson	2004
Preston S. Wilson	2004
Geoffrey S. Edelson	2003
Stanley L. Ehrlich	2003
Brian Ferguson	2003
William M. Hartmann	2003
John Impagliazzo	2003
Hua Lee	2003
Patrick J. Loughlin	2003
James C. Preisig	2003
Edmund J. Sullivan	2003

Ex officio:
John C. Burgess, Associate Editor of JASA
Stanley L. Ehrlich, member of Medals and Awards Committee
David I. Havelock, member of Membership Committee
David J. Evans, member of ASACOS

Speech Communication

	Term to
Diane J. Kewley-Port, *Chair* to 2004	2004
Fredericka Bell-Berti	2005
Ann R. Bradlow	2005
Dani M. Byrd	2005
James M. Hillenbrand	2005
Lori L. Holt	2005
Hideki Kawahara	2005
Qiguang Lin	2005
Mario A. Svirsky	2005
Gary G. Weismer	2005
Catherine T. Best	2004
Rene Carre	2004
Robert D. Frisina	2004
Megan M. Hodge	2004
Jody E. Kreiman	2004
Charissa R. Lansing	2004
Shrikanth S. Narayanan	2004
John J. Ohala	2004
Dwayne Paschall	2004
Astrid Schmidt-Nielsen	2004
Samuel A. Seddoh	2004
Winifred Strange	2004
Elaine T. Stathopoulos	2004
Emily A. Tobey	2004
Mariko Aoki	2003
Shari R. Baum	2003
Suzanne E. Boyce	2003
Christopher S. Campbell	2003
Randy L. Diehl	2003
David Dorado	2003
Bruce Gerratt	2003
John H. L. Hansen	2003
Jean Paul Haton	2003
Sarah Hawkins	2003
Mark S. Hedrick	2003
Sun-Ah Jun	2003
Klaus J. Kohler	2003
Nancy S. McGarr	2003
Ian E. Rogers	2003
Robert Ruiz	2003
Caroline L. Smith	2003
Greg S. Turner	2003
Nagalapura Viswanath	2003

Ex officio:
Peter F. Assmann, Associate Editor of JASA
Randy L. Diehl, Associate Editor of JASA
Kenneth W. Grant, Associate Editor of JASA
Marjorie R. Leek, Associate Editor of JASA
Anders Lofqvist, Associate Editor of JASA
Douglas D. O'Shaughnessy, Associate Editor of JASA

Gary G. Weismer, member of Medals and Awards Committee
Maureen L. Stone, member of Membership Committee
Shrikanth Narayanan, member of ASACOS

Structural Acoustics and Vibration

	Term to
Scott D. Sommerfeldt, *Chair* to 2003	2003
John A. Burkhardt	2005
Kenneth D. Frampton	2005
Sunil Mehta	2005
Martin L. Pollack	2005
Paul J. Remington	2005
Jeffrey S. Vipperman	2005
Richard L. Weaver	2005
Kuangcheng Wu	2005
Paul E. Barbone	2004
Jeffrey E. Boisvert	2004
Robert L. Clark	2004
Joseph M. Cuschieri	2004
John A. Fahnline	2004
Jerry H. Ginsberg	2004
Stephen A. Hambric	2004
Peter C. Herdic	2004
Harry Himelblau	2004
Timothy W. Leishman	2004
Jerome E. Manning	2004
Philip L. Marston	2004
James G. McDaniel	2004
Karl M. Reichard	2004
Angie Sarkissian	2004
Victor W. Sparrow	2004
Eric E. Ungar	2004
Nickolas Vlahopoulos	2004
Benjamin Bard	2003
Alain C. Berry	2003
Joseph W. Dickey	2003
David Feit	2003
Allison B. Flatau	2003
Guillermo C. Gaunaurd	2003
Karl Grosh	2003
Sabih I. Hayek	2003
Francis Kirschner	2003
Jean R. Nicolas	2003
Allan D. Pierce	2003
Mauro Pierucci	2003
Vasundara V. Varadan	2003
Earl G. Williams	2003
Sean F. Wu	2003

Ex officio:
Jerry H. Ginsberg, Associate Editor of JASA
Andrew N. Norris, Associate Editor of JASA
Richard L. Weaver, Associate Editor of JASA
Earl G. Williams, Associate Editor of JASA
Mauro Pierucci, member of Medals and Awards Committee
Courtney B. Burroughs, member of Membership Committee
Louis A. Herstein, member of ASACOS

Underwater Acoustics

	Term to
John S. Perkins, *Chair* to 2003	2003
Paul B. Baxley	2005
Shira L. Broschat	2005
Douglas H. Cato	2005
Michael D. Collins	2005
Peter H. Dahl	2005
Grant B. Deane	2005
Gerald L. D'Spain	2005
Peter Gerstoft	2005
Charles W. Holland	2005
Kevin B. Smith	2005
Hee Chun Song	2005
Dajun Tang	2005
Christopher T. Tindle	2005
Alexandra I. Tolstoy	2005
Stephen N. Wolf	2005
Ahmad T. Abawi	2004
Michael G. Brown	2004
Dennis B. Creamer	2004
Christian P. de Moustier	2004
Stanley E. Dosso	2004
George V. Frisk	2004
Stewart A. L. Glegg	2004
Joseph F. Lingevitch	2004
Nicholas C. Makris	2004
Zoi-Heleni Michalopoulou	2004
Marshall H. Orr	2004
Gregory J. Orris	2004
James C. Preisig	2004
Daniel Rouseff	2004
William L. Siegmann	2004
Ralph N. Baer	2003
Kyle M. Becker	2003
John Buck	2003
Nicholas P. Chotiros	2003
Dezhang Chu	2003
David R. Dowling	2003
Steven Finette	2003
Paul C. Hines	2003
Anatoliy N. Ivakin	2003
Finn B. Jensen	2003
Sunny Khosla	2003
John J. McCoy	2003
B. Edward McDonald	2003
Kazuhiko Ohta	2003
John R. Preston	2003
Ralph A. Stephen	2003
Alexander G. Voronovich	2003
Lisa M. Zurk	2003

Ex officio:
John C. Burgess, Associate Editor of JASA
William M. Carey, Associate Editor of JASA
Stanley A. Chin-Bing, Associate Editor of JASA
William L. Siegmann, Associate Editor of JASA
Ralph A. Stephen, Associate Editor of JASA
David L. Bradley, member of Medals and Awards Committee
Peter H. Rogers, member of Membership Committee
Arnie L. Van Buren, member of ASACOS

Administrative Committees 2002–2003

Archives and History

	Term to
Henry E. Bass, *Chair* to 2004	2004
Leo L. Beranek	2005
William J. Cavanaugh	2005
Logan E. Hargrove	2005

David T. Blackstock	2004
William W. Lang	2004
Julian D. Maynard	2004
Rosalie M. Uchanski	2004
Mark C. Hedrick	2003
John W. Kopec	2003
Wesley L. Nyborg	2003
Richard J. Peppin	2003

Books

	Term to
David L. Bradley, *Chair* to 2005	2005
Nancy S. McGarr	2005
Jeffrey A. Nystuen	2005
Emily A. Tobey	2005
Stanley L. Chin-Bing	2004
Gordon E. Martin	2004
Victor W. Sparrow	2004
Christopher Feuillaude	2003
Jerry H. Ginsberg	2003
Philip L. Marston	2003
Joseph Pope	2003
Robert A. Walkling	2003
Stephen N. Wolf	2003

Ex officio:
Philip L. Marston, Associate Editor of JASA for Book Reviews

College of Fellows Steering

	Term to
William J. Cavanaugh, *Chair* to 2003	2003
Stanley L. Ehrlich	2005
E. Carr Everbach	2005
Robert D. Frisina	2005
Daniel R. Raichel	2005
Peter G. Cable	2004
M. David Egan	2004
Uwe J. Hansen	2004
Astrid Schmidt-Nielsen	2004
Joseph W. Dickey	2003
Judy R. Dubno	2003
Jiri Tichy	2003
Janet M. Weisenberger	2003

Education in Acoustics

	Term to
Uwe J. Hansen, *Chair* to 2003	2003
Courtney B. Burroughs	2005
Robin O. Cleveland	2005
Kenneth A. Cunefare	2005
D. Michael Daly	2005
Mary Florentine	2005
Logan E. Hargrove	2005
Mardi C. Hastings	2005
Peter L. Hoekje	2005
Darrell R. Jackson	2005
Michel T. T. Jackson	2005
Murray F. Korman	2005
Luc Mongeau	2005
John S. Robertson	2005
James M. Sabatier	2005
Neil A. Shaw	2005
Kevin B. Smith	2005
Ralph A. Stephen	2005
James E. West	2005
Wayne M. Wright	2005
David A. Brown	2004
Robert D. Collier	2004
Corinne M. Darvennes	2004
Margaritis S. Fourakis	2004
Carole E. Gelfer	2004
Douglas R. Jones	2004
Sharon Y. Manuel	2004
Philip L. Marston	2004
Ralph Muehleisen	2004
Andrew A. Piacsek	2004
Daniel R. Raichel	2004
Sally G. Revoile	2004
Thomas D. Rossing	2004
Ronald A. Roy	2004
Dawn R. Schuette	2004
Scott D. Sommerfeldt	2004
William Thompson, Jr.	2004
Robert A. Walkling	2004
George S. K. Wong	2004
William A. Ahroon	2003
Anthony A. Atchley	2003
Fredericka Bell-Berti	2003
Suzanne E. Boyce	2003
Robert D. Celmer	2003
Annabel J. Cohen	2003
E. Carr Everbach	2003
Thomas B. Gabrielson	2003
Katherine S. Harris	2003
Elizabeth S. Ivey	2003
Joie P. Jones	2003
Maria B. Mody	2003
Amy T. Neel	2003
P. K. Raju	2003
Deborah M. Rekart	2003
Daniel A. Russell	2003
M. Roman Serbyn	2003
Victor W. Sparrow	2003
Emily A. Tobey	2003

Ethics and Grievances

	Term to
Mardi C. Hastings, *Chair* to 2005	2005
Barbara Shinn-Cunningham	2005

Investments

	Term to
Ira Dyer, *Chair* to 2004	2004
Lawrence A. Crum	2005
Ira J. Hirsh	2003

David Feit, Treasurer, *ex officio*

Medals and Awards

		Term to
Mark F. Hamilton, *Chair* to 2003		2003
Ewart A. Wetherill	Architectural Acoustics	2003
David L. Bradley	Underwater Acoustics	2003
Charles R. Greene	Animal Bioacoustics	2003
Mauro Pierucci	Structural Acoustics and Vibration	2003
James W. Beauchamp	Musical Acoustics	2004
Michael J. Buckingham	Acoustical Oceanography	2004
Mahlon D. Burkhard	Engineering Acoustics	2004
Stanley L. Ehrlich	Signal Processing in Acoustics	2004
Gary G. Weismer	Speech Communication	2004
Marjorie R. Leek	Psychological and Physiological Acoustics	2003
Wesley L. Nyborg	Biomedical Ultrasound/ Bioresponse to Vibration	2003
James M. Sabatier	Physical Acoustics	2003
Michael R. Stinson	Noise	2003

Committee on Meetings—June 2002–December 2002

Dana S. Hougland, *Chair* to 2005
Anthony A. Atchley, Vice President-Elect
Sergio Beristain, Fall 2002, Cancun
Courtney B. Burroughs, Spring 2002, Pittsburgh
Joseph M. Cuschieri, Fall 2001, Ft. Lauderdale
Samir N. Y. Gerges, Fall 2002, Cancun
D. Wesley Grantham, Spring 2003, Nashville
Elaine Moran, ASA Office Manager, *ex officio*
Clark Penrod, Fall 2003, Austin
Scott D. Pfeiffer, Spring 2001, Chicago
Charles E. Schmid, Executive Director, *ex officio*
James E. West, Fall 2002, Cancun
William A. Yost, Vice President

Membership

		Term to
Joseph W. Dickey, *Chair* to 2003		2003
Anthony J. Brammer	Biomedical Ultrasound/ Bioresponse to Vibration	2005
Courtney B. Burroughs	Structural Acoustics and Vibration	2005
Burton G. Hurdle	Foreign Members	2005
John Erdreich	Noise	2005
Christopher Feuillade	Acoustical Oceanography	2004
Steven L. Garrett	Physical Acoustics	2004
David I. Havelock	Signal Processing in Acoustics	2004
Peter H. Rogers	Underwater Acoustics	2004
Gregory C. Tocci	Architectural Acoustics	2004
Uwe J. Hansen	Musical Acoustics	2003
Thomas R. Howarth	Engineering Acoustics	2003
Darlene R. Ketten	Animal Bioacoustics	2003
Maureen Stone	Speech Communication	2003
Ervin R. Hafter	Psychological and Physiological Acoustics	2003

Public Relations

	Term to
Paul A. Baxley, *Chair* to 2003	2003
Gerald D'Spain	2005
Geoffrey F. Edelmann	2005
Ellen S. Livingston	2005
James F. Lynch	2005
Duncan E. McGehee	2005
Barbara J. Sotirin	2005
E. Carr Everbach	2004
Charles Gaumond	2004
Christy K. Holland	2004
David Lubman	2004
Andrew A. Piacsek	2004
William J. Cavanaugh	2003
N. Ross Chapman	2003
Stanley E. Dosso	2003
Blas Espinoza-Varas	2003
Jack E. Randoff	2003
Thomas D. Rossing	2003
James C. Yu	2003

Allan D. Pierce, Editor-in-Chief, *ex officio*
Elaine Moran, ASA Office Manager, *ex officio*
Charles E. Schmid, Executive Director, *ex officio*
Thomas D. Rossing, Echoes Editor, *ex officio*

Publication Policy

	Term to
Floyd Dunn, *Chair* to 2003	2003
Robert C. Bilger	2005
James F. Lynch	2005
James H. Miller	2005
Allan J. Zuckerwar	2005
David T. Blackstock	2004
Mark F. Hamilton	2004
Patricia K. Kuhl	2004
Alan Powell	2004
Sigfrid D. Soli	2004
Jont B. Allen	2003

Ilene J. Busch-Vishniac, President-Elect, *ex officio*
Allan D. Pierce, Editor-in-Chief, *ex officio*

Regional Chapters

Elizabeth A. McLaughlin, *Chair* to 2005

Elmer L. Hixson	Austin
Angelo J. Campanella	Central Ohio
Dean E. Capone	Central Pennsylvania
John W. Kopec	Chicago
Ernest M. Weiler	Cincinnati
Edwin H. Toothman	Delaware Valley
Gary W. Siebein	Florida
Timothy J. Foulkes	Greater Boston
Michael J. Anderson	Inland Northwest
vacant	Los Angeles
Hari S. Paul	Madras, India
Marehalli G. Prasad	Metropolitan New York
Elizabeth A. McLaughlin	Narragansett
Noral Stewart	North Carolina
Peter F. Assmann	North Texas
James R. Angerer	Northwest
R. Dean Ayers	Orange County
Paul A. Baxley	San Diego
David Braslau	Upper Midwest
vacant	Washington, DC
Thomas M. Disch	Wisconsin

Uwe J. Hansen, Chair, Education in Acoustics, *ex officio*

Rules and Governance

	Term to
Tony F. W. Embleton, *Chair* to 2005	2005
William M. Hartmann	2005
Richard H. Lyon	2005
William J. Cavanaugh	2004
Floyd Dunn	2004
Elaine Moran	2003
Charles E. Schmid	2003

Prizes and Special Fellowships

	Term to
Wayne M. Wright, *Chair* to 2005	2005
Uwe J. Hansen	2005
Anthony A. Atchley	2004
Constantine Trahiotis	2004
Fredericka Bell-Berti	2003
James E. West	2003

Standards

Executive Committee
Paul D. Schomer, Chair (Standards Director)
vacant, Vice Chair
Susan B. Blaeser, Standards Manager, *ex officio*

S1 Representation
George S. K. Wong, Chair S1 and ASA representative on S1
John P. Seiler, Vice Chair S1 and ASA alternate representative on S1

S2 Representation
Richard J. Peppin, Chair S2
David J. Evans, Vice Chair
Sabih I. Hayek, ASA representative on S2
Bruce E. Douglas, ASA alternate representative on S2

S3 Representation
R. F. Burkard, Chair S3 and ASA representative on S3
J. Franks, Vice Chair S3 and ASA alternate representative on S3

S12 Representation
Paul D. Schomer, Chair S12
Robert D. Hellweg, Vice Chair S12
Bennett M. Brooks, ASA representative on S12
Paul D. Schomer, ASA alternate representative on S12

International TAGs (ex officio)
Paul D. Schomer, Chair, U.S. TAG for ISO/TC 43 and ISO/TC 43/SCI
David J. Evans, Chair, U.S. TAG for ISO/TC 108
Victor A. Nedzelnitsky, U.S. Technical Advisor for IEC/TC 29

ASA Technical Committee Representatives
William A. Yost, Chair of ASA Technical Council, *ex officio*
Er Chang Shang, Acoustical Oceanography
Ann E. Bowles, Animal Bioacoustics
George E. Winzer, Architectural Acoustics
E. Carr Everbach, Biomedical Ultrasound/Bioresponse to Vibration
Mahlon D. Burkhard, Engineering Acoustics
Ian M. Lindevald, Musical Acoustics
Richard J. Peppin, Noise
Sameer I. Madanshetty, Physical Acoustics
Peggy B. Nelson, Psychological and Physiological Acoustics
David J. Evans, Signal Processing in Acoustics
Shrikanth Narayanan, Speech Communication
Louis A. Herstein, Structural Acoustics and Vibration
Arnie L. Van Buren, Underwater Acoustics

ASA Officers
David Feit, Treasurer, *ex officio*
Charles E. Schmid, Executive Director, *ex officio*

Past Chair of ASACOS (ex officio)
Tony F. W. Embleton

Associate Editors for Standards News—JASA (ex officio)
Susan B. Blaeser
George S. K. Wong

Tutorials

	Term to
Yves H. Berthelot and Beverly A. Wright, *Cochairs* to 2005	2005
Ann E. Bowles	2005
James M. Chambers	2005
Gary W. Elko	2005
Fredericka Bell-Berti	2004
Robin O. Cleveland	2004
George V. Frisk	2004

Charles E. Schmid, Executive Director, *ex officio*

Women in Acoustics

	Term to
Nancy S. Timmerman, *Chair* to 2003	2003
Uwe J. Hansen	2005
Mardi C. Hastings	2005
Peggy B. Nelson	2005
Lily Wang	2005
Lisa Zurk	2005
Lawrence A. Crum	2004
Martha M. Larson	2004
Penelope Menounou	2004
Sandra L. Poliachik	2004
Brigitte Schulte-Fortkamp	2004
Elvira B. Viveiros	2004
Ina Rea Bicknell	2003
Margaret Cheesman	2003
Bozena Kostek	2003
Alexandra I. Tolstoy	2003
Peta White	2003

Anthony A. Atchley, *ex officio* as Vice President-Elect

JASA Editorial Board

June 2005
P. F. Assmann, Speech Perception
S. B. Blaeser, Acoustical News—Standards
K. A. Cunefare, Noise, Its Effects and Control
K. W. Grant—Speech Perception
P. L. Marston, Acoustical Reviews—Books
W. G. Mayer, Acoustical News—International
A. N. Norris, Structural Acoustics and Vibration
D. L. Rice, Acoustical Reviews—Patents
R. L. Diehl, Speech Perception
D. Keith Wilson, Noise, Its Effects and Controls

G. S. K. Wong, Acoustical Standards News
Allan J. Zuckerwar, Applied Acoustics; Transduction; Acoustical Measurements

June 2004
W. W. L. Au, Bioacoustics—Animal
Y. H. Berthelot, Ultrasonics; Physical Effects of Sound
J. C. Burgess, Acoustic Signal Processing
D. E. Chimenti, General Linear Acoustics
F. Dunn, Bioacoustics—Biomedical
J. H. Ginsberg, Structural Acoustics and Vibration
M. F. Hamilton, Nonlinear Acoustics
J. G. Harris, Ultrasonics; Physical Effects of Sound
M. S. Howe, Aeroacoustics
M. Kleiner, Architectural Acoustics
S. E. McAdams, Music and Musical Acoustics
M. Ochmann, General Linear Acoustics
D. D. O'Shaughnessy, Speech Processing
R. S. Raspet, Ultrasonics, Physical Effects of Sound
W. L. Siegmann, Underwater Acoustics
L. C. Sutherland, Aeroacoustics
A. J. Szeri, Ultrasonics; Physical Effects of Sound
L. L. Thompson, General Linear Acoustics
N. F. Viemiester, Psychological Acoustics
R. L. Weaver, Structural Acoustics and Vibration
E. R. Williams, Structural Acoustics and Vibration

June 2003
L. E. Bernstein, Psychological Acoustics
L. H. Carney, Physiological Acoustics
S. A. Chin-Bing, Underwater Sound
A. Lofqvist, Speech Production
B. L. Lonsbury-Martin, Physiological Acoustics
J. J. McCoy, Mathematical Acoustics
E. Moran, Acoustical News—USA
T. D. Rossing, Education in Acoustics
V. W. Sparrow, Computational Acoustics
R. Stern, Electronic Archives and References

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).]*

Microphone Arrays

M. Brandstein and D. Ward, Editors

Springer-Verlag, New York, 2001.
398 pp. Price: $99.00 (hardcover) ISBN: 3540419535.

Microphone Arrays is a collection of 18 chapters about microphone array research organized into four sections: speech enhancement, source localization, applications, and future directions. The editors have assembled contributions from 32 experts who broadly represent the research currently being conducted on microphone arrays and their applications.

In the interest of fairness to the editors and contributors, I will first admit to a certain amount of skepticism about the value of these edited survey books. While the underlying idea of presenting the collective wisdom of a particular research community is appealing, these books have generally struck me as incoherent assemblies of each of the authors' latest ideas—with minimal description of relevant related works. The organization is often topical with great variation in approach and notation. In addition, these surveys often exhibit substantial redundancy of basic ideas. I generally find books with one or two authors more concisely and efficiently present a particular topic.[1]

This particular collection covers an important research area and presents a wide range of views—from directional microphone analyses (Elko) to Kalman filtering applied to a combination of acoustic and visual localization cues (Strobel *et al.*). I found the chapters to be of varying quality—with most presenting complete references for a particular area and some (e.g., Greenberg and Zurek) providing insightful discussion of referenced work in the context of the chapter rather than simply citing the work. Much of the reported work is available in the open literature, but it is convenient to have such a wide collection in a single volume. My general complaints about this format were sometimes validated. However, given my disinclination to this type of collection, I was frequently surprised to find myself enjoying the book.

In the first section of the book the chapters are generally concerned with a variety of beamforming approaches. Included in this section are nice discussions of superdirective arrays (Bitzer and Simmer) and robust beamforming (Hoshuyama and Sugiyama). In addition, Elko presents an interesting analysis of a variety of combinations of differential microphones in isotropic noise fields. The final chapters of this section focus more explicitly on speech enhancement with a particularly interesting combination of speech modeling and speech enhancement by Brandstein and Griebel.

The second section of the book contains three chapters focusing on the use of microphone arrays for sound localization. DiBiase *et al.* do a nice job of describing why time delay estimation in realistic acoustic environments is a difficult problem. The extensive signal processing literature on multiple source localization is reviewed in the chapter by Di Claudio and Parisi. In the final chapter in this section, Strobel *et al.* present an interesting combination of audio-visual cues (as well as a review on Kalman filter theory). This is probably the most focused section of the book.

The third section of the book presents a variety of applications of microphone arrays. These include hearing aids, noise and echo reduction, hands-free automobile applications, speech recognition, and blind separation of acoustic signals. The hearing aid chapter (Greenberg and Zurek) contains a nice review of both existing research work and existing devices. It also addresses key considerations (e.g., signal cancellation and the relative importance of low frequencies) that hearing-impaired users place on the application of microphone arrays. Three chapters address the combination of echo reduction and noise cancellation. Martin focuses on arrays using post-filtering, while Kellermann explores the benefits and limitations of different combinations of echo cancellation and beamforming resources. Nordholm *et al.* present a nice subband application for an automobile array that also combines the tasks of echo suppression and noise reduction. Other applications discussed include speech recognition using microphone arrays (Omologo *et al.*) and the difficult problem of the blind separation of realistic acoustic signals (Douglas).

The final section contains two chapters about the future directions and applications of microphone arrays. Elko discusses fundamental limitations in a thoughtful way—asking researchers to think in new ways about what kind of processing is really ''optimal.'' Van Compernolle discusses a variety of applications that appear to be well suited to the use of microphone arrays. I must admit some concern over his decision to describe car phones as a ''killer application'' for microphone arrays given the inherent risks of driving while distracted.

Overall I was pleasantly surprised by the quality of the collection. This relatively small (400 page) collection presents a nice overview of microphone array research over the last two decades. The chapters also provide a good set of references for someone interested in engaging in microphone array work in a particular area. While the collection does have many of the drawbacks associated with its format, there are enough positives to justify its use as a resource.

MICHAEL W. HOFFMAN
Department of Electrical Engineering
University of Nebraska
Lincoln, Nebraska 68588-0511

[1]I must admit that I have found writing chapters for these edited books can be an onerous task as well. So I do bear some sympathy for the contributors.

Filter Design with Time Domain Mask Constraints: Theory and Applications

Ba-Ngu Vo, Antonio Cantoni, and Kok Lay Teo

Kluwer Academic Publishers, Boston, 2001.
350 pp. Price: $136.00 (hardcover) ISBN: 0792371380.

This monograph is concerned with the time-domain design of an envelope constrained (EC) filter whose response stays within a given tolerance band of the desired response and also minimizes the noise amplification. The objective is to present a unified approach to the problem that spans digital, analog, and hybrid (analog and digital) systems. The research results during the last three decades on the EC problem and its generalization, to handle the case of uncertain inputs (that is, inputs not specified exactly, but known to stay within an input mask), have been published in scattered journals and so the monograph should be a welcome addition to the literature. Readers with some background in optimization theory by vector space methods will benefit more from the perusal of the monograph, and the required math-

ematical background has been summarized in two appendices for quick reference.

Besides the two appendices, the monograph contains six chapters. Chapter 1, entitled ''Introduction,'' motivates and formulates the problem, and gives a brief account of the topics covered in the succeeding chapters. The shortcomings of the conventional least squares approach to filter design in selected applications are underscored and the relevance of the EC approach is pointed out.

In Chap. 2, entitled ''Filtering with convex response constraints,'' the EC filtering problem and its variants are shown to be special cases of a convex programming problem in Hilbert space which covers the analog, digital, and hybrid cases of interest. The issues of existence, uniqueness, and finite-dimensional approximants to the optimal solution are addressed here. The lemmas leading to the proof of the theorem on convergence to the optimal solution as the dimension of the underlying vector space tends to infinity are contained in an appendix to this chapter.

Chapter 3, entitled ''Analysis and problem characterization,'' examines the relationship between the primal and dual solutions to, respectively, the convex programming problem in Hilbert space and its dual in the space of regular Borel measures. The dual of the EC filtering problem for both the finite- and infinite-dimensional cases is considered. In particular, the practically important finite-dimensional case is investigated in the setting of the convex semi-infinite programming problem (optimization in finite-dimensional spaces with infinitely many constraints). The special case of semi-infinite quadratic programming and the primal–dual structure of this class of problems offer insight into the structure of the optimal EC filter. Again, an appendix to this chapter provides proof of results not included in the main body of the text for interested readers.

Chapter 4, entitled ''Discrete-time EC filtering algorithms,'' is devoted to computational aspects of finding solutions to the discrete-time EC filtering problem. Strategies suitable for off-line and also on-line implementations of algorithms are discussed. A tapped delay line FIR filter structure and a discrete-time orthonormal Laguerre network structure (as a low-order, more robust IIR alternative to the FIR structure) are presented for solving the problem of focus in this chapter and the performances of the two approaches are compared via a channel equalization example.

The continuous-time counterpart of the discrete-time EC filtering problem in Chap. 4 is presented in Chap. 5, entitled ''Numerical methods for continuous-time EC-filtering.'' Both analog and hybrid filters are considered in this chapter and some attention is given to the problem of approximation. The continuous-time orthonormal Laguerre network, originally due to Norbert Wiener and Yuk Wing Lee, is used here in the channel equalization problem in addition to FIR filters accompanied with various postfilters for the hybrid case.

Chapter 6, entitled ''Robust envelope constrained filtering,'' addresses the important issue of robustness to perturbations of the prescribed input or filter implementation errors.

The philosophy behind this book is filter design with the objective of restricting the response within prescribed upper and lower bounds while minimizing the effect of input noise. The synthesis is in the time domain and the mathematical tools used are nonlinear optimization and functional analysis. Another approach towards the design of filters so that the magnitude of the frequency response is constrained by greatest lower and least upper bounds has been advanced in Refs. 1 and 2 listed below. The analysis is conducted in the frequency domain and dwells on the deployment of primarily algebraic tools. The common feature in both approaches is the concern with the need for tackling a set of filters as opposed to a single filter for performance evaluation. A significant body of recent literature has also been devoted to the design of interval filters for robust stability (Ref. 3), but this book is not concerned with the parametric approach to tackling uncertainty documented there.

N. K. BOSE
Department of Electrical Engineering
Pennsylvania State University
University Park, Pennsylvania 16802

[1] N. K. Bose and K. D. Kim, ''Boundary implications for frequency response of interval FIR and IIR filters,'' IEEE Trans. Signal Process. **39**, 2167–2173 (1991).
[2] A. Levkovich, E. Zeheb, and N. Cohen, ''Frequency response envelopes of a family of uncertain continuous-time systems,'' IEEE Trans. **42**, 156–165 (1995).
[3] J. Kogan, *Robust Stability and Converxity: An Introduction*, Lecture Notes in Control and Info. Sci. (Springer, London, 1995), Vol. 201.

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Micro-Tech Hearing Instruments, 3500 Holly Lane No., Suite 10, Plymouth, Minnesota 55447*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
KEVIN P. SHEPHERD, *Mail Stop 463, NASA Langley Research Center, Hampton, Virginia 23681*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*

6,338,277

43.20.Hq FLOWMETER FOR ATTENUATING ACOUSTIC PROPAGATIONS

Andrew Stephen Diston *et al.*, assignors to G. Kromschroder Aktiengesellschaft
15 January 2002 (Class 73/861.28); filed 4 December 1998

The velocity of a fluid flowing along a cylindrical passage is determined from the time of flight of acoustic waves between two transducers located at the ends of the passage. Small diameter passages are undesirable because they induce pressure losses, but large diameter passages may permit the generation of nonaxial modes, which may interfere with the measurement. The latter problem is overcome in the present patent by the introduction of attenuation devices on the walls of the passage. Typical attenuation devices include axial slots that contain dissipative material or rough surfaces and/or that communicate with extended volumes.—EEU

6,347,293

43.20.Ye SELF-CHARACTERIZING VIBRATING CONDUIT PARAMETER SENSORS AND METHODS OF OPERATION THEREFOR

Timothy J. Cunningham and Stuart J. Shelley, assignors to Micro Motion, Incorporated
12 February 2002 (Class 702/189); filed 9 July 1999

This patent applies to Coriolis effect mass flowmeters, in which the vibration modes of a curved conduit are changed as the result of the flow of mass through that conduit. In such meters, vibrations of the conduit are produced by an actuator and measured by one or more sensors. In order to overcome errors that may be introduced by changes in the sensor and to overcome some effects of extraneous vibrations, this patent employs an automated approach for determining the conduit's modal characteristics.—EEU

6,348,968

43.35.Sx PHOTOACOUSTIC SPECTROSCOPY APPARATUS AND METHOD

Tom Autrey *et al.*, assignors to Battelle Memorial Institute
19 February 2002 (Class 356/432); filed 30 March 2001

A cell for stimulating a material sample by light and detecting the resulting sound waves that emanate from the sample is constructed of two blocks of a light-transmitting material, such as quartz. The two blocks may be visualized as having trapezoidal cross sections and as being connected along their larger faces via an annular gasket. The test sample is confined in the central space between the blocks and the gasket. A light beam is directed onto the sample via a lateral face of one of the blocks and may be either reflected or refracted by the sample, permitting different modes of excitation of the sample. A transducer attached to the base of one of the blocks is used to detect the sound.—EEU

6,350,238

43.35.Yb REAL-TIME DISPLAY OF ULTRASOUND IN SLOW MOTION

Bjorn Olstad and Hans Torp, assignors to GE Medical Systems Global Technology Company, LLC
26 February 2002 (Class 600/437); filed 2 November 1999

The current state of the art makes it possible to obtain ultrasound images with frame rates that exceed the limitations of the human eye and current video standards such as PAL and NTSC. The human eye recognizes 30–50 frames per second, but rates of 100–300 images per second can be achieved with current high performance systems. The patent covers a device and system for acquiring ultrasound information at a high acquisition rate, but with a slowed-down display rate so that images can be better perceived by the human eye. Acquisition and display are synchronized under specified conditions, which may be a predetermined time interval or a triggering event. The result is a slow-motion display that maintains real-time synchrony.—DRR

6,350,241

43.35.Yb METHOD AND APPARATUS FOR MULTIPLE ANGLE COMPOUND FLOW IMAGING

Ilan Lifshitz, assignor to GE Medical Systems Global Technology Company, LLC
26 February 2002 (Class 600/454); filed 27 December 1999

The patent describes a method of multi-angle compound flow imaging which receives at an ultrasound transducer multiple ultrasound reflections from a target. The reflections are oriented at multiple angles with respect to the transducer normal. The method entails jointly evaluating information derived during the various signal reflections to establish a display of the target and displaying the result as, for example, an angiogram.—DRR

6,353,277

43.38.Ar ACOUSTIC TRANSDUCER

Thomas Hahn-Jose, assignor to Fraunhofer-Gesellschaft zur Foerderung der Angewandten Forschung E.V.
5 March 2002 (Class 310/324); filed in Germany 23 August 1997

This transducer consists of a piezoceramic disc that is bonded to a diaphragm to form a monomorphous flexural vibrator. As the disc vibrates in its plane, the Poisson effect results in thickness vibration, and this vibration is coupled to the air by the diaphragm. The latter preferably consists of a mixture of epoxy and hollow glass spheres and has a thickness approximately equal to a quarter wavelength of the thickness vibration mode of the disc, so that it has low acoustic impedance.—EEU

6,354,146

43.38.Fx ACOUSTIC TRANSDUCER SYSTEM FOR MONITORING WELL PRODUCTION

James Robert Birchak *et al.*, assignors to Halliburton Energy Services, Incorporated
12 March 2002 (Class 73/61.79); filed 17 June 1999

The system described in this patent consists of one or more pairs of piezoelectric transducers that are mounted within the walls of thick pipes in oil wells. The properties of well products flowing through these are evaluated by means of signals that propagate perpendicular to the flow direction. The design is claimed to be suitable for reliable multi-year long operation in a downhole environment.—EEU

6,338,395

43.38.Ja KAPOK LOUDSPEAKER ENCLOSURE DAMPING MATERIAL

J. Dwidjaja Setiabudi *et al.*, assignors to P. T. Hartono Istana Electronics
15 January 2002 (Class 181/151); filed 29 September 1999

Kapok is a fluffy, fibrous, organic material somewhat similar to cotton puffs or unspun wool. In the 1930s it was commonly employed as furniture stuffing. Is it possible that until now no one has ever lined or stuffed a loudspeaker box with kapok?—GLA

6,343,133

43.38.Ja AXIALLY PROPAGATING MID AND HIGH FREQUENCY LOUDSPEAKER SYSTEMS

Alan Brock Adamson, Scarborough, Ontario, Canada
29 January 2002 (Class 381/340); filed 22 July 1999

Contemporary concert sound systems may employ large line arrays made up of identical two-way or three-way modules. Known prior art (Heil's United States patent 5,163,167) can convert sound from a conventional high-frequency driver into a coherent vertical line source. This patent describes an improved coaxial waveguide that is intended to combine sound from a high-frequency and a mid-range driver into a single, cylindrical wavefront without the drawbacks of earlier designs.—GLA

6,345,102

43.38.Ja VEHICLE LOUDSPEAKERS

Gillian Margaret Davis *et al.*, assignors to New Transducers Limited
5 February 2002 (Class 381/86); filed in the United Kingdom 16 October 1996

One or more piezoelectric wafers drive a thin, lightweight panel. The result is a talking sun visor.—GLA

6,347,823

43.38.Ja VEHICLE FRONT END PANEL WITH HORN COVER

Ikuo Ozawa *et al.*, assignors to Denso Corporation
19 February 2002 (Class 296/1.1); filed in Japan 20 October 1999

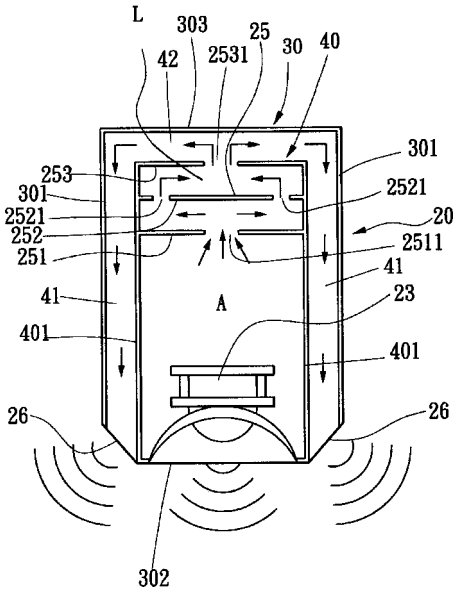
An automobile horn assembly is proposed in which the acoustic driver is attached to a horn that is integral to the front body member which houses the radiator. Reduced part count and cost are the motivating factors behind this design.—KPS

6,339,649

43.38.Ja LOUDSPEAKER SYSTEM WITH STACKABLE LOUDSPEAKER UNITS

Waterson Chen, Taichung City and Chin-Lung Lin, Kaohsiung City, both of Taiwan, Province of China
15 January 2002 (Class 381/335); filed 12 April 1999

Although this appears to be a transmission line loudspeaker cabinet with a few extra twists and turns, the inventor has other ideas: (a) multiple modules can be separated or stacked, and (b) radiation from openings **26** is intended to cover ''...the whole area of the audiovisual room.'' Woofer mod-



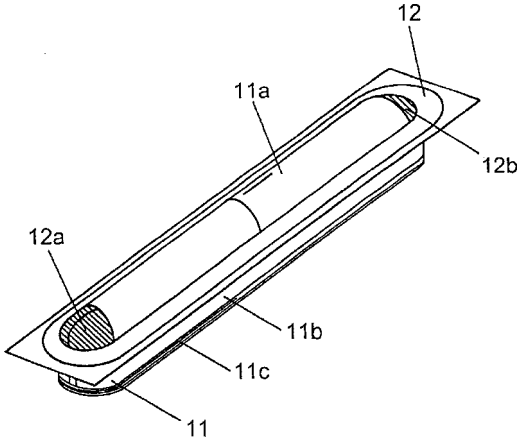
ules, two-way modules (woofer plus tweeter), and a subwoofer module can be arranged in various groupings.—GLA

6,341,167

43.38.Ja SPEAKER

Masatoshi Okuyama *et al.*, assignors to Matsushita Electric Industrial Company, Incorporated
22 January 2002 (Class 381/407); filed in Japan 18 June 1998

An elongated panel or dome **11a** is edge-driven by voice coil **11c**. This configuration, which may have originated with Doschek around 1956, can



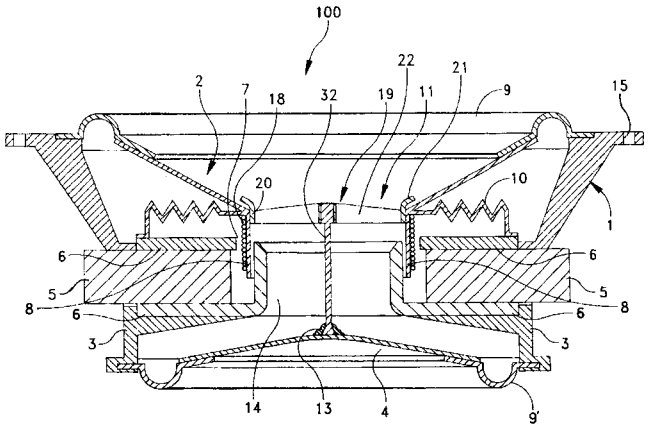
utilize a simple magnetic structure employing two parallel linear gaps. The patent describes an improved design which is intended to increase electroacoustic efficiency while reducing manufacturing cost.—GLA

6,343,128

43.38.Ja DUAL CONE LOUDSPEAKER

C. Ronald Coffin, Topsfield, Massachusetts
29 January 2002 (Class 381/186); filed 17 February 1999

In the inventor's earlier dual-cone loudspeaker, cones **2** and **4** are mechanically linked to increase the volume of air pumped for a given voice coil displacement. One might question whether the added complexity is



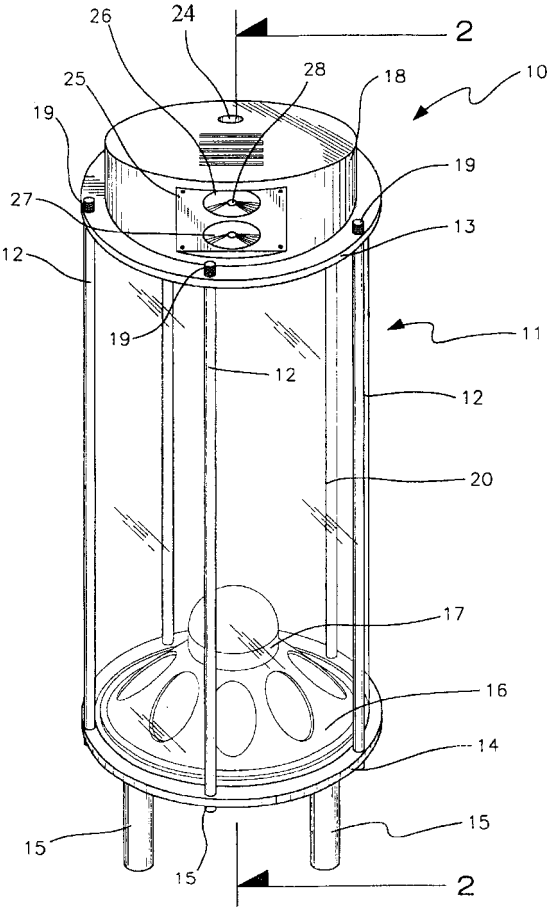
worth the effort, but the design is clever. The patent describes a number of practical improvements intended to improve performance and reliability.—GLA

6,345,685

43.38.Ja LOUDSPEAKER SYSTEM

Leigh D. Wells and Emma-Jane Smith, both of Trellech Monmouthshire, Wales, the United Kingdom
12 February 2002 (Class 181/153); filed in the United Kingdom 26 January 2000

Woofer **17** and tweeters **26** and **27** are mounted at opposite ends of translucent cylinder **20**. The top cap also contains a concealed light source. Light is directed downward, bounces off spherical mirror **17**, and illuminates the inner surface of the cylinder. This is not the first self-contained sound and light device to be patented, but it is said to incorporate the advantages of



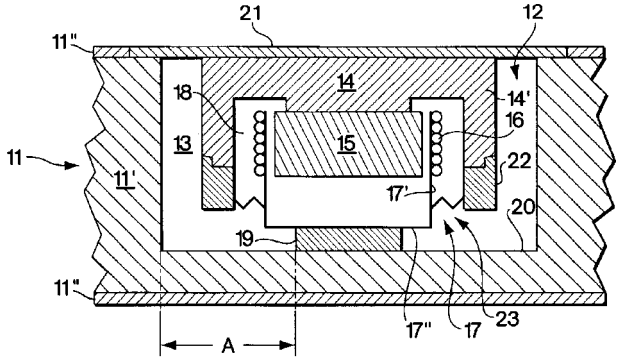
prior art "...while simultaneously overcoming some of the disadvantages normally associated therewith."—GLA

6,347,149

43.38.Ja DRIVER FOR A FLAT ACOUSTIC PANEL

Wolfgang Bachmann *et al.*, assignors to Harman Audio Electronic Systems GmbH
12 February 2002 (Class 381/396); filed in Germany 15 May 1998

Magnetic circuit **14** and **15** and voice coil **16** are typical of moving coil loudspeakers except for centering spider **23**, which appears to fit into a space no larger than magnetic gap **18**. The entire motor assembly sits in a cavity in "acoustic panel" **11**, the core of which is rigid expanded foam. Mechanical



coupling member **19** is smaller than the voice coil diameter to facilitate the generation of bending waves in panel **11**. The design appears to be more like a tuning fork than a conventional panel loudspeaker and one would expect low-frequency efficiency to suffer accordingly.—GLA

6,347,147

43.38.Kb HIGH NOISE SUPPRESSION MICROPHONE

Edward F. Downs, Jr. and Kevin M. Venturella, assignors to the United States of America as represented by the Secretary of the Navy
12 February 2002 (Class 381/114); filed 7 December 1998

The invention is intended for use in high-noise environments such as helicopters, tanks, speedboats, and the like. A thin piezoelectric film is sandwiched between two conductive layers. The sandwich is embedded in a mechanical sound filter except for one exposed face which is placed in contact with the user's body (i.e., forehead). The patent text asserts that this arrangement is superior to boom microphones or air-conduction helmet-mounted microphones. However, no test data is referenced and no mention is made of prior helmet-mounted contact microphones designed for military use.—GLA

6,349,223

43.38.Lc UNIVERSAL HAND-FREE SYSTEM FOR CELLULAR PHONES IN COMBINATION WITH VEHICLE'S AUDIO STEREO SYSTEM

Tonny Chen, assignor to E. Lead Electronic Company, Limited
19 February 2002 (Class 455/569); filed 8 March 1999

A hands-free system for use with cellular phones is designed to operate in combination with a vehicle's audio stereo system. The system can automatically turn the stereo system off and on in response to the presence or absence of an incoming cellular phone signal. The design functions for various types of phones produced by different manufacturers.—KPS

6,360,187

43.38.Lc AMBIENT ADJUSTED VOLUME CONTROL FOR IN-VEHICLES MESSAGES

Thomas J. Hermann, assignor to Ford Global Technologies, Incorporated
19 March 2002 (Class 702/191); filed 28 June 1999

Within an automobile interior, the sound level of high-priority messages such as navigation instructions and warning signals is automatically adjusted based on the measured ambient noise, including that from the entertainment system.—KPS

6,356,185

43.38.Md CLASSIC AUTOMOBILE SOUND PROCESSOR

Jay Sterling Plugge, Sunnyvale, California and Jason Carl Plugge, Stillwater, Minnesota
12 March 2002 (Class 340/384.3); filed 5 July 2000

An automobile sound processor containing prerecorded or synthesized sound signatures of vintage automobiles is integrated with the stereo sound system. Signals from sensors that measure engine speed and manifold vacuum pressure are used to change the reproduced sound so that it reflects the automobile's state, be it constant speed, acceleration, or deceleration. Thus a modern car, or a reproduction, can be made to sound like a vintage car to the vehicle occupants.—KPS

6,343,130

43.38.Vk STEREOPHONIC SOUND PROCESSING SYSTEM

Yasushi Yamazaki, assignor to Fujitsu Limited
29 January 2002 (Class 381/309); filed in Japan 3 July 1997

The invention is concerned with improved localization of computer game sound images. The processing system includes a number of stereo filter units, each including a FIR filter and a method of selecting the most appropriate filter setting for the particular computer in use. Depending on the computer CPU, performance may be sacrificed for speed and vice versa.—GLA

6,343,131

43.38.Vk METHOD AND A SYSTEM FOR PROCESSING A VIRTUAL ACOUSTIC ENVIRONMENT

Jyri Huopaniemi, assignor to Nokia OY
29 January 2002 (Class 381/310); filed in Finland 20 October 1997

Parametrized filters are used to represent reflective and absorptive surfaces in a virtual acoustic environment. Filter parameters are stored and recalled rather than using lookup tables of filter coefficients. The patent includes interesting information and is generally easy to follow.—GLA

6,343,132

43.38.Vk LOUDSPEAKER

Suemei Fukuhara *et al.*, assignors to Matsushita Electric Industrial Company, Limited
29 January 2002 (Class 381/339); filed in Japan 28 February 1997

In cinema sound mixing, the front-center loudspeaker is normally used to reproduce dialog. However, in a home TV installation the center speaker is most often located above the picture rather than behind it. This patent

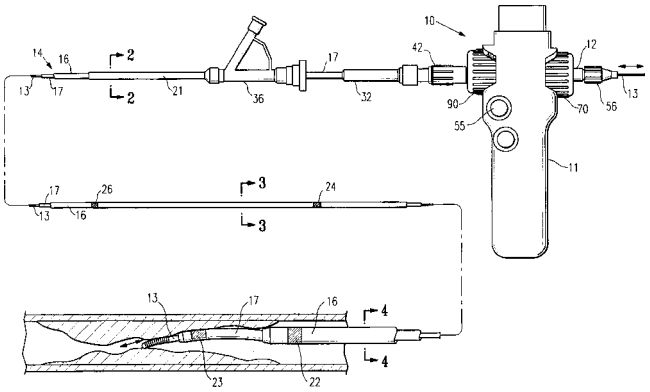
describes an unusual design intended to delocalize the center sound image just enough to avoid conflict with the visual image. A narrow band of frequencies centered near 6 kHz is electrically removed from the main loudspeaker and fed to a tweeter fitted with a special horizontal dispersing device.—GLA

6,348,040

43.40.Ng VIBRATING GUIDEWIRE

Kent C. B. Stalker and Edward J. Nance, assignors to Advanced Cardiovascular Systems, Incorporated
19 February 2002 (Class 600/585); filed 6 April 1999

This is a handheld vibrating device that imparts reciprocating motion to a guidewire while leaving the guidewire free to rotate about its longitudinal axis and facilitating advancement of the guidewire through a highly occluded blood vessel. The vibrating device is equipped with a lock ring for



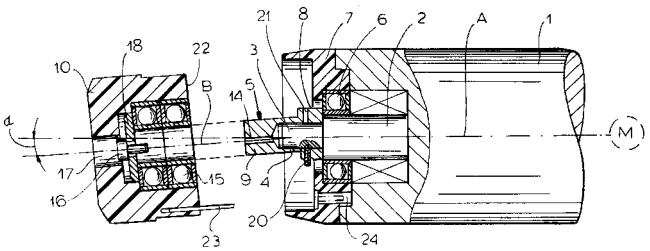
reversibly preventing guidewire rotation without affecting guidewire reciprocation. The device may be used with or without a delivery catheter fitted with a distal end, the curvature of which can be varied while the catheter is inside the patient.—DRR

6,352,518

43.40.Ng APPLIANCE FOR VIBRATION THERAPY WITH MOTOR HOUSING AND ECCENTRIC HEAD DRIVE

Helga Ruf nee Wolf, Griesheim, Germany
5 March 2002 (Class 601/89); filed 15 November 1999

This is an appliance for carrying out vibration and percussive therapy for the treatment of bone, musculature, and lung pathologies, or for prophylaxis against a disorder, or simply serving as a device for health maintenance. The device uses an oscillation generator that is affixed to a stub of a motor shaft. The oscillation generator contains a pin that is inclined at an angle to the axis of its shoulder that bears against a shoulder between the shaft and a stub thereof and is secured to the stub by a setscrew.—DRR



6,332,511

43.50.Gf SILENCER ASSEMBLY HAVING SINGLE STRAND FIBERGLASS ACOUSTIC PACK MATERIAL

Michael P. Parlato *et al.*, assignors to Burgess-Manning, Incorporated
25 December 2001 (Class 181/282); filed 7 December 1999

This muffler design is intended for high-velocity, high-temperature, exhaust flows, as found in high-pressure steam vents and gas turbine engines. The design consists of the common arrangement of two concentric cylinders, with acoustical absorption material placed in the space between the cylinders. The fill material is single strand fiberglass, which is said to prevent fragmentation and migration of the fill material.—KPS

6,341,663

43.50.Gf SILENCER WITH A SHUNT RESONATOR

Matthias Alex and Rolf Fuesser, assignors to Filterwerk Mann & Hummel GmbH
29 January 2002 (Class 181/249); filed in Germany 27 November 1997

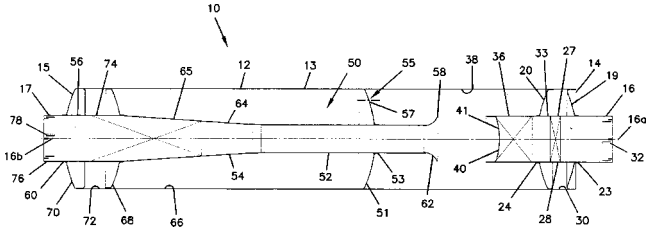
This silencer, intended for intakes of internal combustion engines and the like, consists of a pipe that is surrounded by an essentially cylindrical volume. The pipe communicates with this volume via a number of radial tubular elements. These elements, in conjunction with the volume, form Helmholtz resonators. Selection of the diameters, lengths, and locations of these elements permits one to design mufflers that are effective in desired frequency ranges.—EEU

6,334,506

43.50.Gf MUFFLER ARRANGEMENTS AND METHODS

John E. Hamrin and Matthew W. Jones, assignors to Donaldson Company, Incorporated
1 January 2002 (Class 181/249); filed 10 August 2000

This muffler is intended for use with heavy-duty trucks. In contrast to most designs, this one is aimed at reducing noise in the sleeping quarters rather than drive-by noise. (Team driving requires that one person rests while the other drives.) The muffler seeks to reduce noise at frequencies below 350 Hz, where resonances in the sleeping cab are found. The inlet



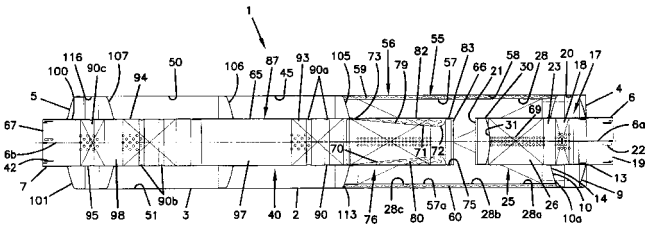
tube **16** is closed at **40;** gases flow through perforated sections **28** and **36**. The outlet tube **50** consists of constant-area section **52** and divergent-area section **65**. This arrangement is said to reduce noise below 350 Hz. Several other geometrical configurations are described.—KPS

6,354,398

43.50.Gf MUFFLERS FOR USE WITH ENGINE RETARDERS; AND METHODS

Theodore G. Angelo *et al.*, assignors to Donaldson Company, Incorporated
12 March 2002 (Class 181/256); filed 16 May 2000

A muffler intended for use with heavy trucks is specifically designed to attenuate noise under both positive power and compression braking conditions. Inlet **6** is perforated and held in place by perforated baffles. The outlet



7 has a convergent–divergent throat and arrangements of perforations. The outer shell is lined for about one-third of its length, near the inlet, with acoustical absorption material **57** and **60** which provides both absorption and damping of shell vibrations.—KPS

6,353,789

43.50.Nm PREDICTING BROADBAND NOISE FROM A STATOR VANE OF A GAS TURBINE ENGINE

Donald B. Hanson, assignor to United Technologies Corporation
5 March 2002 (Class 701/100); filed 13 December 1999

Stator vanes are located downstream of the fan in a typical turbofan aircraft engine. Flow from the fan, through the stators, generates tonal and broadband noise, both of which can be influenced by the geometry of the stator blades. In contrast to a simple radial arrangement, blade sweep (blade tip is downstream of the root) and blade lean (circumferential displacement of the blade tip relative to its root) can be optimized for noise benefits. This patent presents a mathematical model for the prediction of broadband noise as a function of stator blade lean and sweep.—KPS

6,358,592

43.55.Ev MELTBLOWN FIBROUS ACOUSTIC INSULATION

Larry Leroy Vair, Jr. and Kenneth Andrew Clocksin, assignors to Johns Manville International, Incorporated
19 March 2002 (Class 428/74); filed 6 February 2001

This insulation mat is treated by melting fibers immediately adjacent to the surface in order to form a thin integral skin.—CJR

6,352,134

43.55.Ti ACOUSTIC BOARD

Chao Hsiang Wang, Taipei, Taiwan, Province of China
5 March 2002 (Class 181/292); filed 18 July 2000

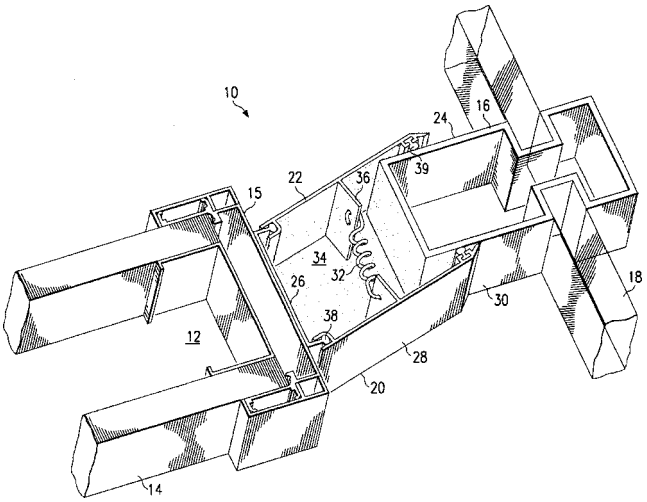
This acoustic board has large and small diameter conical cups imbedded in a panel. The cups have an open end that is meant to collect sound and provide noise reduction from one side of the panel to the other, in defiance of any laws of physics for blocking sound transmission.—CJR

6,351,915

43.55.Ti COUPLING APPARATUS WITH PARALLEL MEMBERS JOINED BY AN ELASTIC OR SPRING ELEMENT PIVOTALLY INTERFACING A DRYWALL AND CURTAIN WALL MULLION

Steven A. Puckett, assignor to S&P Resources, Incorporated
5 March 2002 (Class 52/167.1); filed 28 April 2000

This device couples the end of a partition to an exterior curtain wall mullion. The coupling device has a slip joint **30** and **39** to allow inward–outward movement of the curtain wall during windy conditions, and a



hinged arrangement **38** to allow side-to-side movement of the curtain wall during an earthquake. Even with these movements, the device maintains airtight closure between the curtain wall and the partition.—CJR

6,356,639

43.60.Gk AUDIO DECODING APPARATUS, SIGNAL PROCESSING DEVICE, SOUND IMAGE LOCALIZATION DEVICE, SOUND IMAGE CONTROL METHOD, AUDIO SIGNAL PROCESSING DEVICE, AND AUDIO SIGNAL HIGH-RATE REPRODUCTION METHOD USED FOR AUDIO VISUAL EQUIPMENT

Tsukuru Ishito *et al.*, assignors to Matsushita Electric Industrial Company, Limited
12 March 2002 (Class 381/17); filed in Japan 11 April 1997

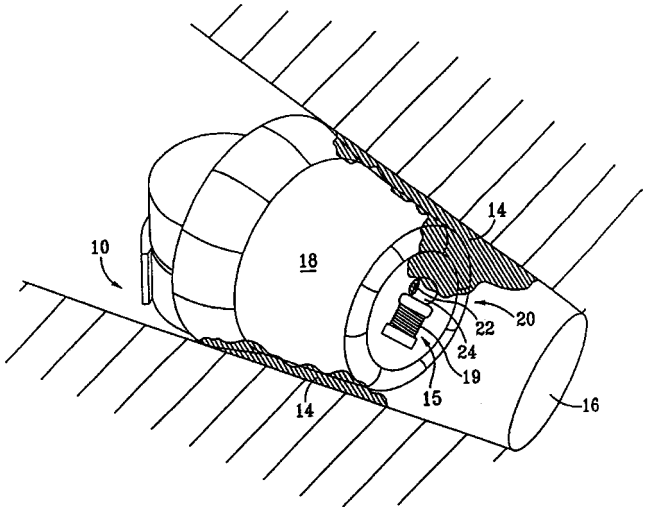
This patent, with the longest title in this reviewer's memory, covers an allegedly ''do-all'' device, a sound image localization device that includes a signal source and a divider for dividing the audio signal into two digital channels. One of the two digital signals is filtered so as to localize a virtual sound image. The filtered and unfiltered digital signals are then converted to analog signals and reproduced over control speakers into an assigned space area.—DRR

6,349,790

43.66.Ts SELF-CLEANING CERUMEN GUARD FOR A HEARING DEVICE

Owen D. Brimhall, assignor to Sonic Innovations, Incorporated
26 February 2002 (Class 181/135); filed 12 June 2000

When an in-the-ear type hearing aid is removed from the wearer's ear canal, a piece of thermally activated material (e.g., bi-metallic element)



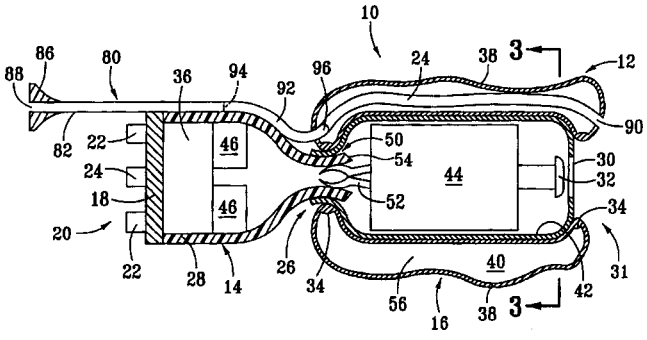
cools and thereby expands. In one embodiment, this expansion produces an extension of the material, causing it to move over the sound outlet of the hearing aid when the device is not being used. When the hearing aid is again placed in the wearer's ear canal, the material warms from body temperature and retracts to expose the sound outlet. The expansion and contraction of the material is said to remove ear wax and other debris blocking the hearing aid sound outlet each time the hearing aid is removed from the ear canal.—DAP

6,359,993

43.66.Ts CONFORMAL TIP FOR A HEARING AID WITH INTEGRATED VENT AND RETRIEVAL CORD

Owen D. Brimhall, assignor to Sonic Innovations
19 March 2002 (Class 381/328); filed 15 January 1999

A completely-in-the-canal (CIC) hearing aid is described that consists of an elastic conformal tip surrounding a module that contains the hearing aid receiver and connects to the main body of the hearing aid via an articulating joint. This type of construction may form an air-tight seal of the hearing aid to the wearer's ear canal which helps to prevent acoustic feedback problems. If such hearing aids do not have a pressure equalization



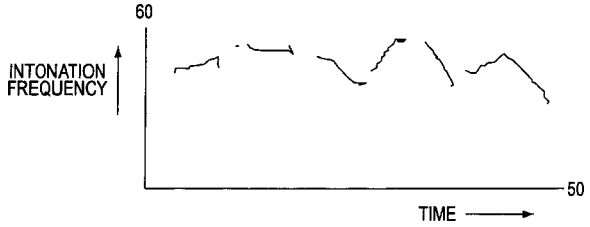
vent, considerable discomfort may occur to the wearer when inserting and removing the hearing aids due to air pressure build-up and vacuum being created during insertion and removal, respectively. To alleviate these problems, the CIC device has a removal cord that is a hollow tube, which forms a pressure equalization vent to allow the air trapped in the wearer's ear canal to escape.—DAP

6,358,054

43.70.Dn METHOD AND APPARATUS FOR TEACHING PROSODIC FEATURES OF SPEECH

Martin Rothenberg, assignor to Syracuse Language Systems
19 March 2002 (Class 434/185); filed 6 June 2000

This speech diagnostic and training station provides prosodic structure feedback to the user in the form of a speechlike signal from which vowel and consonant articulation details have been removed. The signal may take



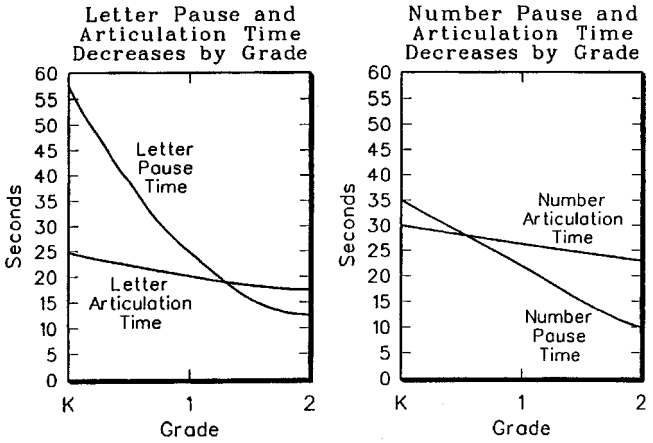
the form of a sinewave, a pulse stream, or a static vowel sound, with varying pitch and loudness. The audio playback may be accompanied by a related video display.—DLR

6,350,128

43.70.Ep RAPID AUTOMATIZED NAMING METHOD AND APPARATUS

Graham Neuhaus, Houston, Texas
26 February 2002 (Class 434/178); filed 5 September 2000

This speech processing system is designed for the diagnosis of reading difficulties. Based on frame-by-frame speech power levels, the input signal is segmented into articulations and pause periods. Pause period durations are



measured as the primary reading-level indicator. Several standardized tests are described which rate the reading competence based on the pause durations.—DLR

6,351,214

43.72.Ar GLASS BREAKAGE DETECTOR

Kenneth G. Eskildsen *et al.*, assignors to Pittway Corporation
26 February 2002 (Class 340/550); filed 9 April 2001

This glass breakage detector processes an input audio signal with increased gain at high frequencies and samples that signal at 44.1 Ksamples/s. Detected features include several filter bands with amplitudes and zero-crossings measured before and after filtering. A fairly elaborate decision tree tests the resulting features and includes false alarm tests, such as the ''microwave door'' rule, the ''balloon pop'' rule, and the ''hand clap'' rule.

Various decision parameters may be adjusted to optimize performance during or after installation.—DLR

6,353,810

43.72.Fx SYSTEM, METHOD AND ARTICLE OF MANUFACTURE FOR AN EMOTION DETECTION SYSTEM IMPROVING EMOTION RECOGNITION

Valery A. Petrushin, assignor to Accenture LLP
5 March 2002 (Class 704/236); filed 31 August 1999

This patent describes several methods for analyzing speech to determine emotional content and a workstation system for evaluating recognition systems with respect to the recognition of emotional content. During the data gathering phase, a person may be fitted with electrodes to measure various biological responses during the recording of test utterances. Various techniques are described for eliciting specific emotions from the speaker during such recording sessions.—DLR

6,351,679

43.72.Ja VOICE ANNOUNCEMENT MANAGEMENT SYSTEM

Gordon Douglas Ainslie, assignor to Telefonaktiebolaget LM Ericsson (publ)
26 February 2002 (Class 700/94); filed in Australia 20 August 1996

This large, distributed system offers the means to operate and manage a network of local and remote audio announcement stations. Each announcement location is provided with a monitor microphone and the resulting audio signals are processed by a recognition system to confirm station operation. Announced items may include speech synthesized from text, prerecorded speech, tones, or other audio signals.—DLR

6,349,280

43.72.Ne METHOD AND APPARATUS FOR SPEAKER RECOGNITION

Hiroaki Hattori, assignor to NEC Corporation
19 February 2002 (Class 704/250); filed in Japan 4 March 1998

This device reportedly determines a speaker's identity using the same feature extraction technology as is used for word recognition. The feature computation details are not presented here, but in a cited publication. Apparently, a typical hidden Markov model system is used to determine distances between input and stored word patterns. Once a word match has been located for an input token, the feature distances are again examined to determine the speaker identity.—DLR

6,349,282

43.72.Ne COMPOUND WORDS IN SPEECH RECOGNITION SYSTEMS

Filip Van Aelten *et al.*, assignors to Lernout & Hauspie Speech Products N.V.
19 February 2002 (Class 704/257); filed 20 April 1999

This speech recognition system offers an effective increase in the size of the phonetic lexicon by combining root word forms with potential affixes to form inflected and compound words. Derived and compounded forms are lumped under the term ''compounds''. For example, the word ''postfixing'' could be constructed of the elements ''post,'' ''fix,'' and ''ing.'' A compound lexicon includes compounding probability values to prevent combining when the combination has a lower probability than the separate items. Phonetic alternations, such as often occur with suffixes like ''tion'' (as in ''repeat''/''repetition''), are not considered.—DLR

6,353,809

43.72.Ne SPEECH RECOGNITION WITH TEXT GENERATION FROM PORTIONS OF VOICE DATA PRESELECTED BY MANUAL-INPUT COMMANDS

Hidetaka Takahashi and Takafumi Onishi, assignors to Olympus Optical, Limited
5 March 2002 (Class 704/235); filed in Japan 6 June 1997

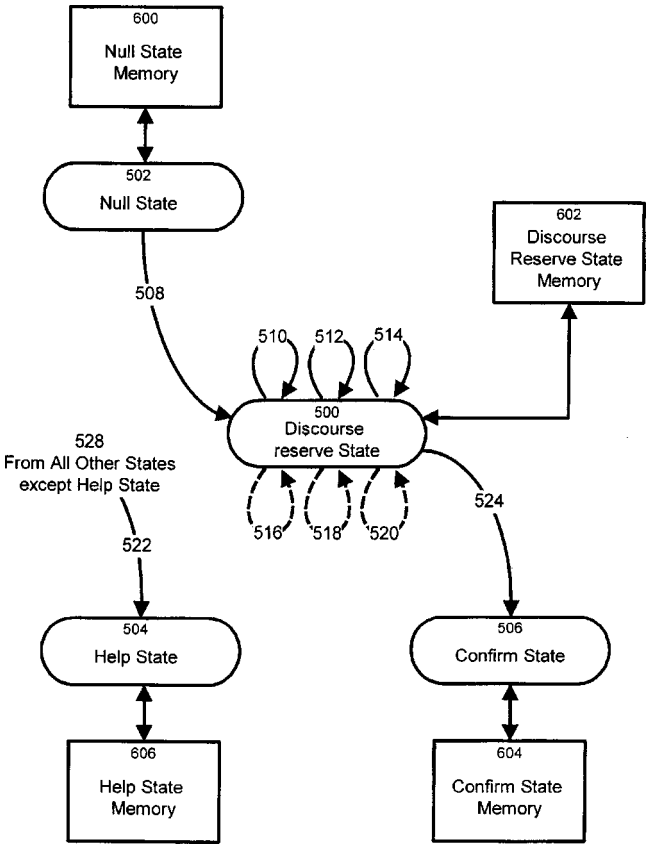
This patent describes a workstation for use in developing and testing speech recognition systems which might be used in consumer devices, such as cell phones or pagers. Alternative recognition programs may be loaded and run to test recognition of prerecorded test materials. Recognizer performance is rated automatically, with or without operator assistance.—DLR

6,356,869

43.72.Ne METHOD AND APPARATUS FOR DISCOURSE MANAGEMENT

Nicolas Chapados *et al.*, assignors to Nortel Networks Limited
12 March 2002 (Class 704/275); filed 30 April 1999

This dialogue management system includes the logic to generate dynamic states in what would otherwise be described as a finite state control structure. Based on words recognized in the incoming speech signal and the current system state, the state transition network is dynamically modified. A



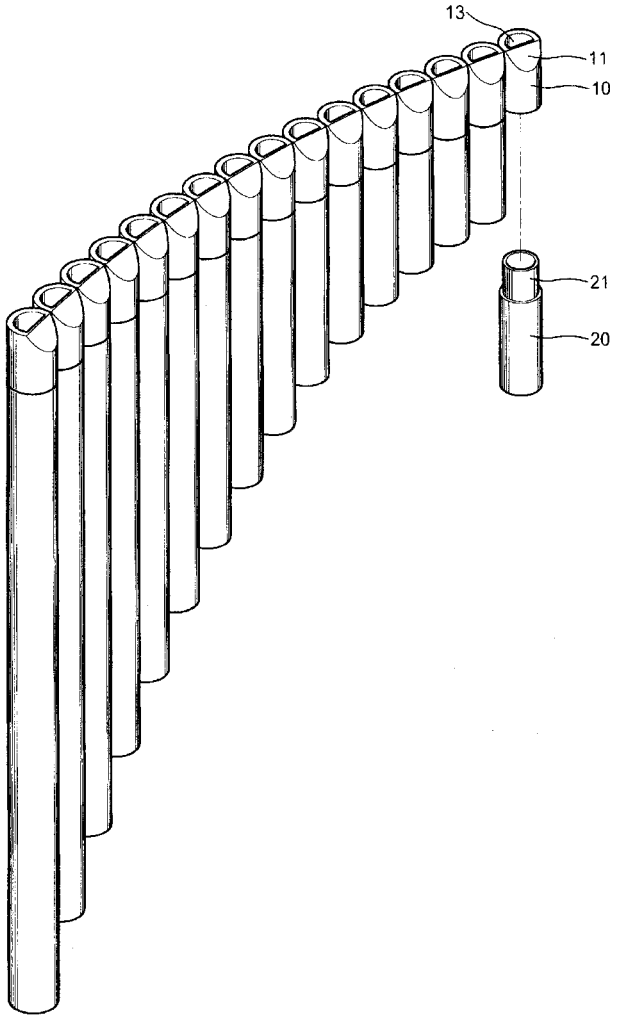
software problem solver continually makes decisions based on available information on the state of the dialogue. The result is said to allow more natural handling of incomplete utterances, shifts of context, and other such effects, providing a more natural speech interaction.—DLR

6,342,662

43.75.Ef TUNE CHANGEABLE PANPIPE WITHOUT HARMING THE LIPS OF A PLAYER

Yin-Yao Chang, Feng-Yuan City 420, Taiwan, Province of China
29 January 2002 (Class 84/402); filed 30 January 2001

Because ''Many students who study to play this panpipe are stopped on halfway because their lips are always harmed by the sharp edge''...the



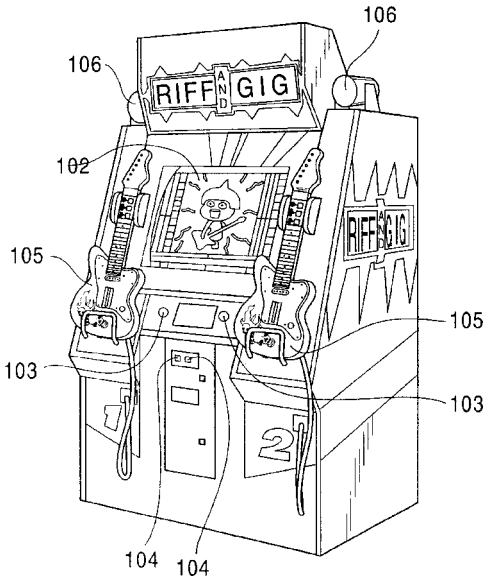
patent proposes a smoother recorderlike edge and insertable transposition pipes for playing in a different key.—MK

6,342,665

43.75.Wx MUSIC GAME SYSTEM, STAGING INSTRUCTIONS SYNCHRONIZING CONTROL METHOD FOR SAME, AND READABLE RECORDING MEDIUM RECORDED WITH STAGING INSTRUCTIONS SYNCHRONIZING CONTROL PROGRAM FOR SAME

Katsunori Okita *et al.*, assignors to Konami Company, Limited
29 January 2002 (Class 84/609); filed in Japan 16 February 1999

A fascinating look at a Japanese arcade rock and roll game machine. The gamer's drum or guitar performance can be compared against a reference, thereby giving a score. Furthermore, the drum and guitar game machines can be networked together for competition and ensemble playing. Sketches of the hardware and software architecture are given.—MK

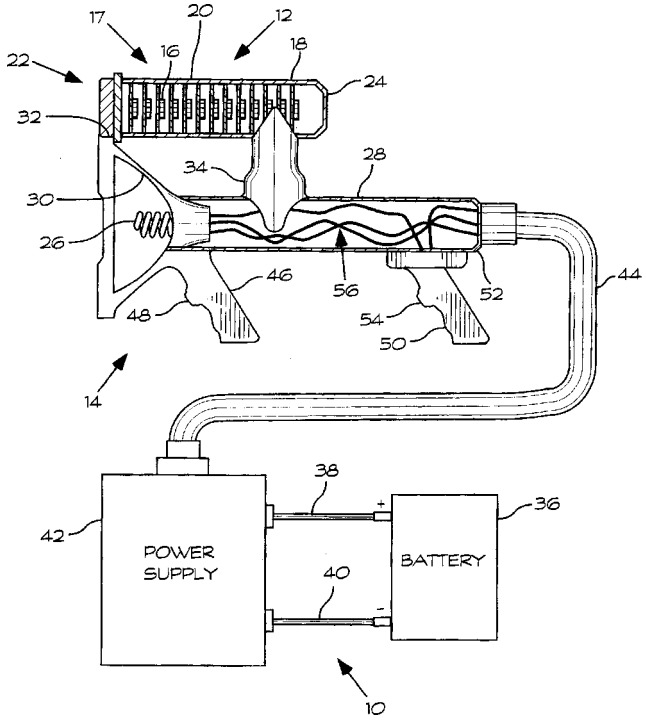


6,359,835

43.80.Nd HIGH INTENSITY DIRECTED LIGHT AND SOUND CROWD DISPERSION DEVICE

Franz J. Gayl, assignor to the United States of America as represented by the Secretary of the Navy
19 March 2002 (Class 367/139); filed 20 March 2001

The device is described as ''a portable, combined arms, audio-visual system intended for humane deterrence, riot control and the defusing of possible escalation of civil violence.'' The audio subsystem generates a directional acoustic beam by means of a linear arrangement of piezoelectric transducers, phased to increase power in the forward direction. The longitudinal end-fired array is contained within an insulated, internally reflective tube, using resonance to increase sound pressure levels, while shielding the operator and his teammates from effects. The intended effect of the acoustic



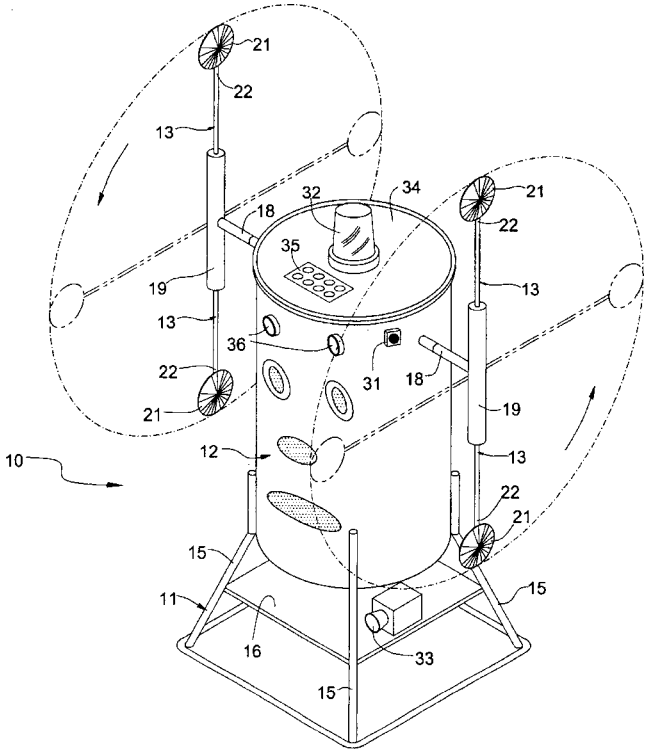
energy is to produce aural pain in an adversary, hopefully sufficient to deter or to cut off threatening confrontation. The audio subsystem can be operated in a variety of modes, such as heightened single-frequency output or a randomized, intermittent noise. The audio subsystem operates in combination with a pulsed, diffusive laser or flash device.—DRR

6,351,908

43.80.Pe AUTOMATED DEER SCARECROW

James Wendell Thomas, Jay, Florida
5 March 2002 (Class 43/1); filed 3 August 1999

This is another of those rather amusing sound-generating devices designed to scare off deer and other undesired animals from agricultural crop



areas. The device consists of a substantially hollow housing with a removable top portion, mounted over a rigid support frame member. The housing includes a number of light- and sound-emitting devices in addition to at least two rotating elongated arms. These arms have a hollow interior containing flowable metal particulate which produces ''unnatural'' sounds during the course of rotation. An electronic controller connected to a power relay energizes the aforementioned electrical components and a motor assembly rotates the arms. The apparatus is powered by a solar-charged battery.—DRR

6,349,720

43.80.Qf APPARATUS FOR ACOUSTICALLY DETERMINING POSITION OF AN ENDOTRACHEAL TUBE

Walter Dennis Clark, assignor to Integrated Medical Systems, Incorporated
26 February 2002 (Class 128/200.26); filed 25 June 1998

A preferred embodiment of this endotracheal tube (ETT) consists of a cuffed tube, the proximal end of which incorporates a noise-generating apparatus that provides an audible signal when air passes through the tube. A second embodiment is a combination of a cuffed ETT and an insertion rod for stiffening the ETT during insertion. The insertion rod is coupled to an electrically powered noise generator mounted on the distal end to produce an audible sound before any air flow through the ETT is established. In both versions, the proper placement is determined by the unaided ear listening to noise coming from both sides of the chest. If the noise is from one side only, the insertion is too deep. If sound is heard from the stomach, the ETT is in the esophagus instead of the trachea and should be pulled out and reinserted in such a manner that it presses more forcefully in the ventral direction.—DRR

6,352,860

43.80.Qf LIQUID AND SOLID TISSUE MIMICKING MATERIAL FOR ULTRASOUND PHANTOMS AND METHOD OF MAKING THE SAME

Ernest L. Madsen and Gary R. Frank, assignors to Wisconsin Alumni Research Foundation
5 March 2002 (Class 436/8); filed 17 November 2000

A tissue mimicking material for ultrasound phantoms has ultrasound speed and attenuation characteristics like those of human tissues and is intended for use in measuring and calibrating the potential biological effects of ultrasound equipment. The material is formed of an aqueous mixture of large, organic water-soluble molecules condensed from skim milk with a total solids content in the range of 10% to 30% by weight. The total fat content is less than 1% by weight, with the residual lipid particles being sufficiently small to remain in suspension without agglomerating and separating from the mixture over extended periods of time.—DRR

6,358,208

43.80.Qf ASSESSMENT OF CARDIOVASCULAR PERFORMANCE USING ULTRASOUND METHODS AND DEVICES THAT INTERROGATE INTERSTITIAL FLUID

Philipp Lang, San Francisco and John D. Mendlein, Encinitas, both of California
19 March 2002 (Class 600/438); filed 21 November 1998

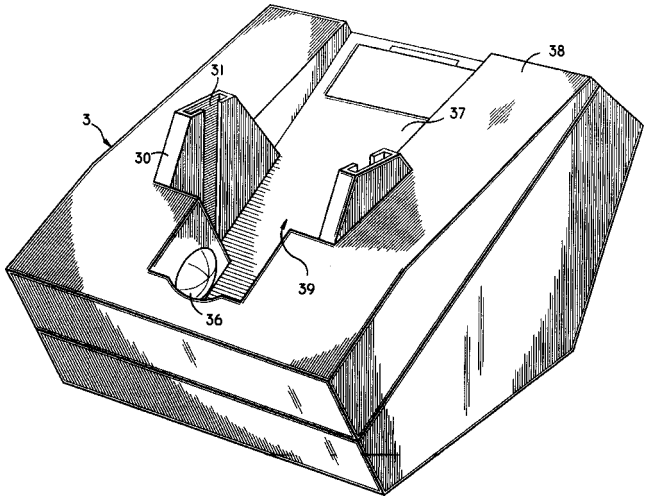
The device and associated methodology purport to measure capillary-related interstitial fluid and to monitor clinically rapid shifts in capillary-related interstitial fluid distribution in tissues through ultrasonic means. Cardiovascular performance is monitored by interrogating interstitial fluid in a subject by using ultrasound waves. The device is intended for either continuous or intermittent monitoring.—DRR

6,352,512

43.80.Qf BONE ANALYSIS APPARATUS AND METHOD FOR CALIBRATION AND QUALITY ASSURANCE OF AN ULTRASONIC BONE ANALYSIS APPARATUS

Kevin E. Wilson *et al.*, assignors to Hologic, Incorporated
5 March 2002 (Class 600/449); filed 26 March 1999

The method involves calibrating an ultrasound bone analysis apparatus containing at least two transducer assemblies. The figure shows such an apparatus for the analysis of human foot bones. Each transducer assembly is adjustable so that a face of each pad can be placed in contact with a phantom simulation body part (not shown) and also such that the faces contact a body



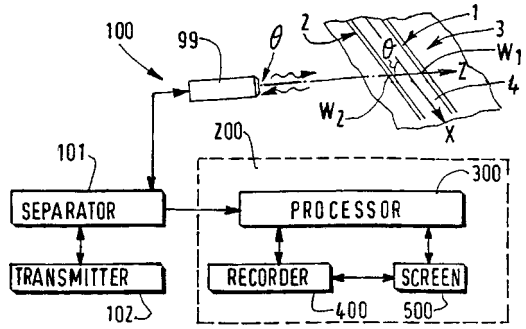
part. An ultrasound signal is transmitted from one transducer to the other. The time interval for the ultrasound to pass through the body part is determined and then, from time and pulse width values, the speed of the ultrasound passing through the body part with squish compensation is calculated.—DRR

6,358,206

43.80.Qf ULTRASOUND PROCESS FOR THE DETERMINATION OF THE LOCATION OF A PARIETAL SURFACE IN A TISSUE AND OF THE ABSOLUTE RADIUS OF AN ARTERY, AND ULTRASOUND APPARATUS FOR CARRYING OUT SUCH PROCESS

Claude Cohen-Bacrie, assignor to U.S. Philips Corporation
19 March 2002 (Class 600/437); filed in the European Patent Office 15 December 1998

The subject device executes an ultrasound process to locate a local discontinuity of homogeneity forming a parietal surface situated in a tissue. The process includes a step for the acquisition of a radio-frequency signal representing the amplitudes of echoes returned by tissue scatterers as functions of their digital depths along an excitation line crossing the parietal



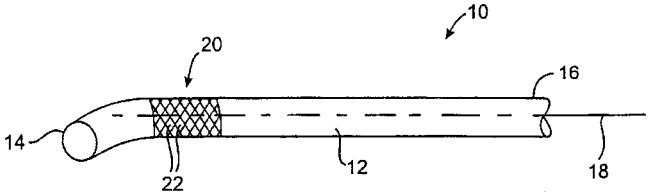
surface. The signals are transmitted through a transducer that is coupled to the tissue and connected to an ultrasound analysis apparatus. The process is also used to estimate an artery radius from the digital depths of the arterial walls on the excitation line.—DRR

6,358,211

43.80.Qf ULTRASOUND LUCENT APPARATUS AND METHODS OF USING

Donald S. Mamayek, assignor to Scimed Life Systems, Incorporated
19 March 2002 (Class 600/459); filed 12 October 1999

The device is directed to providing a guidewire having a greater reflectivity to increase the likelihood of correct guidewire location and positioning. In one embodiment the apparatus consists of an elongated wire body having an outer surface and a longitudinal axis. The wire body contains a number of corner reflectors deployed along an embossed portion of



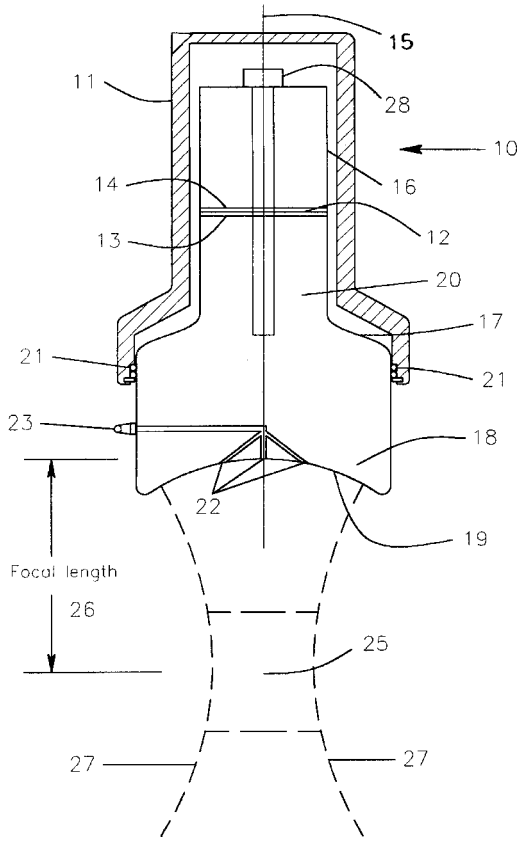
the outer surface. The wire body is designed to be inserted into a body lumen or a patient vasculature. In this manner, the use of corner reflector technology effectively increases the acoustical reflectivity of the wire body.—DRR

6,350,245

43.80.Sh TRANSDERMAL ULTRASONIC DEVICE AND METHOD

William W. Cimino, Louisville, Colorado
26 February 2002 (Class 601/2); filed 22 December 1998

The device is a hand-held ultrasonic surgical device with a focusing lens for fragmenting or emulsifying a predetermined volume of a medium located generally near the lens' focal point without significant heating of the



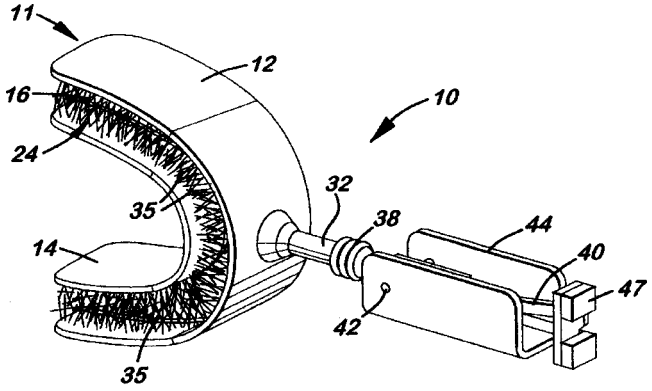
medium. The acoustic assembly has a resonant frequency that is primarily-determined by the assembly length. The preferred range for the resonant frequency, between 100 and 250 kHz, achieves effective focusing and sufficient ultrasonic intensity to fragment or emulsify tissue. The acoustic assembly includes an ultrasonic motor **12**, a rear driver **16**, a front driver **17**, a compression fastener **28**, and a focusing lens **18**.—DRR

6,353,956

43.80.Sh COMBINED ULTRASONIC TOOTHBRUSH MODULE

Jason Berge, Kirkland, Washington
12 March 2002 (Class 15/22.1); filed 31 August 2000

This gadget is a candidate for advertising on late night TV shows. It is a U-shaped toothbrush module to be used with currently available ultrasonic toothbrushes. It consists of a soft thermoplastic, silicone, or latex material. Upper and lower bridge receiving spaces are configured to enclose the bridges when the piece is placed in the user's mouth. On the inside surfaces of the U-shaped attachment are bristles that contact the exposed surfaces of



the teeth. The module includes a pivoting T-shape member and magnets attached at one end that connect to a standard ultrasonic toothbrush hand piece. Motion generated by the hand piece is transmitted through the T-shaped connector, to the U-shaped bristle component, and to the surfaces of the teeth, causing the brushes to move over the upper and lower bridges.—DRR

6,352,509

43.80.Vj THREE-DIMENSIONAL ULTRASONIC DIAGNOSIS APPARATUS

Tetsuya Kawagishi and Naohisa Kamiyama, assignors to Kabushiki Kaisha Toshiba
5 March 2002 (Class 600/443); filed in Japan 16 November 1998

This three-dimensional diagnostic apparatus transmits ultrasonic beams three-dimensionally to a diagnostic site, such as a left ventricle of the heart. Three-dimensional data received from the diagnostic site establishes a cardiac cavity region and generates a display image such that information on a heart myocardial site is identified by different data values. This feature is said to enable the apparatus to provide information on three-dimensional myocardial muscle blood flow that should be useful for clinical diagnoses.—DRR

6,352,510

43.80.Vj ULTRASOUND TRANSDUCERS FOR REAL TIME TWO AND THREE DIMENSIONAL IMAGE ACQUISITION

Leonid S. Barabash, Higley, Arizona *et al.*
5 March 2002 (Class 600/443); filed 22 June 2000

The patent covers an apparatus and method for rapid two- and three-dimensional ultrasound image acquisition. A transducer design with a phased array placement cross geometry is supposed to reduce the level of sidelobe amplitudes and extend the usable frequency range of the transducers. An objective is to provide a simple cross transducer display with one transmit array and one receive array, eliminating the shared control element. The patent includes versions of cross transducers with a shared central element, but the design of the central crossing area is said to allow separation of transmit and receive returns, with an increase in signal-to-noise ratio. The method provides for use of cross transducers for acquisition of real time two- and three-dimensional images.—DRR

6,352,511

43.80.Vj MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD FOR POST PROCESSING

John A. Hossack *et al.*, assignors to Acuson Corporation
5 March 2002 (Class 600/443); filed 2 November 2000

A send/receive transducer generates ultrasonic data. Image processing is applied to the ultrasound data for presentation of an image. A number of the ultrasound image processing steps may be reversed. For example, persistence processing may be reversed in order to obtain ultrasound data associated with data prior to persistence processing. The recovered data may be used to generate an image or for changing the amount of persistence. Other processes that may be reversed to recover ultrasound data include focal and depth gain compensation, dynamic range compression, intensity or color mapping, and various filters, such as persistence or spatial filtering.—DRR

Loudness enhancement: Induced loudness reduction in disguise? (L)

Bertram Scharf
INSERM EPI 9902, Laboratoire d'Otologie et de NeuroOtologie, Université d'Aix-Marseille II, Marseille, France and Institute for Hearing, Speech, & Language and Department of Psychology (125 NI), Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Søren Buus[a)]
Institute for Hearing, Speech, & Language and Communications & Digital Signal Processing Center, Department of Electrical & Computer Engineering (440 DA), Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Bärbel Nieder
Institute for Hearing, Speech, & Language, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

(Received 24 January 2002; accepted for publication 21 June 2002)

Two opposite sequential loudness effects concern the effect of a stronger Tone 1 on the loudness of a subsequent weaker Tone 2, as assessed by loudness matches with Tone 3. Loudness enhancement is reported when Tone 1 precedes Tone 2 by 50 to 100 ms. Loudness recalibration (or *induced loudness reduction*) is obtained for delays of about 1 s. This letter argues that what appears as an enhancement of Tone 2's loudness is, in fact, an induced reduction of Tone 3's loudness, which occurs because Tones 1 and 3 are at the same frequency. Preliminary experiments support this analysis. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1500755]

PACS numbers: 43.66.Cb, 43.66.Mk [MRL]

According to a number of investigators (e.g., Elmasian and Galambos, 1975; Plack, 1996; Zeng, 1994; Zwislocki and Sokolich, 1974), a Tone burst, Tone 2, increases in loudness level by as much as 10 to 15 phons when it follows—with a delay not exceeding 100 ms—a stronger Tone burst, Tone 1, at the same frequency. This loudness enhancement was measured by introducing a third Tone burst, Tone 3, at least 500 ms after Tone 2. Elmasian and Galambos (1975) and Zwislocki and Sokolich (1974) had the listener adjust the level of Tone 3 so that it sounded as loud as Tone 2. Plack (1996) and Zeng (1994) had the listener indicate which tone was louder—the second one or third one—in a 2IFC two-track adaptive procedure, which varied the level of Tone 2. Both procedures gave the same result: In the presence of the stronger Tone 1, the level of Tone 3 had to be higher than that of Tone 2 to make them equal in loudness. This apparent enhancement of the loudness of Tone 2 usually diminished when Tone 1 became much more intense than Tone 2 and disappeared when it was at the same level as Tone 2. Zwislocki and Sokolich (1974) also showed that Tone 1 had to be at or near the frequency of Tones 2 and 3 for maximum enhancement; at separations much greater than a critical band, enhancement was absent.

Given these and other reports of measurements made at various sound frequencies and levels, loudness enhancement appeared to be a robust if unexplained phenomenon. However, the interpretation of the results described above as an enhancement of loudness needs to be reconsidered in the light of recent experiments showing that a stronger tone often *reduces* the loudness of a weaker tone that follows it with a delay of the order of a second. This effect has been referred to as loudness recalibration (e.g., Mapes-Riordan and Yost, 1999; Marks, 1994). We prefer the term induced loudness reduction (ILR). [Not only is ILR a more neutral and descriptive term, but also it is reminiscent of induced loudness adaptation with which it shares many characteristics (cf. Scharf, 2001).] The stimulus conditions that lead to ILR are also found in the experiments that reported loudness enhancement. Hence, ILR must have taken place in all the reported enhancement experiments. It is possible that what looked like an increase in the loudness of Tone 2 was, in fact, a decrease in the loudness of Tone 3. To examine this possibility, we compare the experimental paradigms and procedures used in the experiments on enhancement and ILR.

Figure 1 shows the basic stimulus sequences used in the measurement of enhancement and ILR. In both paradigms, a stronger Tone 1 is followed by Tone 2 whose loudness is assessed by comparison to Tone 3. In enhancement, Tone 2 follows Tone 1 after about 100 ms, whereas in ILR, Tone 2 is almost always delayed more than 500 ms. In enhancement, Tone 3 is always at the same frequency as Tone 1; in ILR, it is at a distant frequency. [In enhancement, Tones 2 and 3 last less than 30 ms; in ILR, they usually last 200 ms or more, but this difference probably is unimportant because tone bursts that last only 5 ms undergo nearly as much ILR as tones that last 200 ms (Nieder *et al.*, 2002).]

We compare typical results for enhancement from Zeng (1994) and for ILR from Mapes-Riordan and Yost (1999). Both studies used a 2IFC two-track adaptive procedure to determine equal loudness. Table I provides procedural and

[a)]Author to whom correspondence should be addressed. Electronic mail: buus@neu.edu

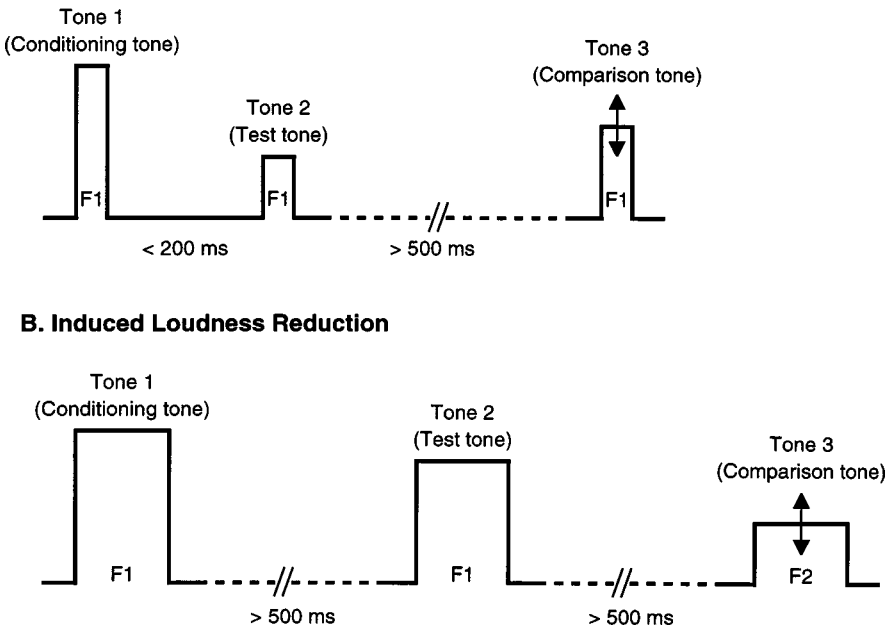


FIG. 1. Schematic illustration of stimulus sequences used in the measurement of enhancement and ILR. (a) For loudness enhancement, a stronger conditioning tone (Tone 1) precedes a weaker test tone (Tone 2) by less than 200 ms. The loudness of Tone 2 is assessed by matching it in loudness to a comparison tone (Tone 3) that follows it with a delay of at least 500 ms. (b) For induced loudness reduction, the stronger conditioning tone (Tone 1) precedes the weaker test tone (Tone 2) by at least 500 ms. Again, the loudness of Tone 2 is assessed by matching it in loudness to a comparison tone (Tone 3) that follows it with a delay of at least 500 ms. The frequency (F2) of Tone 3 is different from that (F1) of Tones 1 and 2.

stimulus details for the two experiments. The two critical differences are highlighted: (1) Zeng delayed Tone 2 by 100 ms, Mapes-Riordan and Yost by 1000 ms; and (2) Zeng presented all three tones at 1000 Hz whereas Mapes-Riordan and Yost had the first two tones at 500 Hz and Tone 3 at 2500 Hz. (Other relatively small differences between the two studies should have had little effect on the results.)

The enhancement paradigm showed that Tone 3 had to be at a higher level than Tone 2 to maintain equal loudness in the presence of the stronger Tone 1. The ILR paradigm gave just the opposite outcome. When the stronger Tone 1 was introduced, the level of Tone 3 always had to be lowered to maintain equal loudness. We take as an example the case in which the tone that underwent ILR (Tone 3 in the enhancement study, Tone 2 in the ILR study) was approximately 20 dB weaker than Tone 1. To achieve equal loudness with Tone 2, in the enhancement study the level of Tone 3 had to be 15 dB *higher* than that required in the absence of Tone 1, whereas in the ILR study, Tone 3 had to be 11 dB *lower*.

Is it the case that enhancement turns into ILR as the delay between Tones 1 and 2 is lengthened from near 100 ms to over 500 ms? If so, ILR ought to have been apparent in an experiment in which Zwislocki and Sokolich (1974) delayed Tone 2 by 500 ms, but their data seemed to show neither enhancement nor ILR. This finding can be explained as follows. Because Tones 1, 2, and 3 all had the same frequency and were 500 ms apart, Tone 1 surely reduced the loudness of both Tones 2 and 3, an ILR that could not be revealed by matching Tones 2 and 3 in loudness. It must be that the frequency relations among the tones play an essential role. Enhancement is measured when the delay is short and Tones 1 and 3 are at the same frequency. Induced loudness reduction is measured when the delay is longer and Tones 1 and 3 are at different frequencies. We propose that loudness enhancement reflects, at least in part, the ILR caused by Tone 1 on Tone 3. For example, in Zeng's (1994) study, the delay from the end of the strong Tone 1 to the onset of *Tone 3* was 775 ms which must have led to a decline in the loudness level of Tone 3. Even if the loudness of Tone 2 did not change, Tone 3 would have had to be set higher in level to compensate for the decline in its loudness. The decline would be expected to be approximately 10 phons (e.g., Mapes-Riordan and Yost, 1999). Since Zeng measured average ''enhancement'' of 8 to 15 dB, most of the enhancement must have reflected ILR. Hence what appears to be enhancement of the loudness of Tone 2 is more likely a reduction of the loudness of Tone 3.

A straightforward test of this hypothesis is to use the enhancement paradigm of Fig. 1(a), but with Tone 3 set to a very different frequency from Tones 1 and 2. According to the enhancement hypothesis, the loudness of Tone 2 should be enhanced without regard to the frequency of Tone 3. Ac-

TABLE I. Comparison of enhancement experiment by Zeng (1994) and ILR experiment by Mapes-Riordan and Yost (1999).

			Stronger tone				Test tone				Comparison tone			
Study	Procedure	Participants	Frequency (Hz)	SPL (dB)	Duration (ms)	Delay (ms)	Frequency (Hz)	SPL (dB)	Duration (ms)	Interval (ms)	Frequency (Hz)	SPL (dB)	Duration (ms)	Intertrial interval
Zeng (1994)	2IFC	3	**1000**	90	100	**100**	**1000**	varied	25	650	**1000**	30–90	25	Resp. +2 s
Mapes-Riordan and Yost (1999)	2IFC	4	**500**	80	1000	**1000**	**500**	60, 70	500	500	**2500**	varied	500	Resp. +1 s

cording to the ILR hypothesis, the loudness of Tone 2 will not appear to be enhanced because the loudness of Tone 3 will not be reduced owing to the large frequency separation between Tones 1 and 3. Preliminary tests were carried out on 14 undergraduates who received course credit for their participation. It turned out that comparing a third 20-ms tone burst to a second one that came only 100 ms after a stronger tone (in a 2IFC adaptive procedure) was very difficult for these untrained, unmotivated listeners. Hence, both intra- and interlistener variability was very large. Nonetheless, the average results generally agree with the ILR hypothesis outlined above. Statistically significant enhancement was obtained when Tone 3 had the same frequency (500 Hz) as the first two tones, but not when it had a very different frequency (2500 Hz). The implication is that when Tone 3 was at the same frequency as the stronger Tone 1, presented 600 ms earlier, the loudness of Tone 3 was reduced. To compensate for this reduction, the level of Tone 3 had to be increased. The results of this preliminary study also suggest that the stronger Tone 1 had little effect on the loudness of Tone 2; introducing Tone 1 did not appear to affect the level at which Tone 3 (at 2500 Hz) was matched in loudness to Tone 2. The conclusion that the loudness of Tone 2 was changed little if at all by the presence of Tone 1 agrees with the observation that the change of loudness level obtained in enhancement experiments is roughly the same as that obtained in ILR experiments.

Another test of the hypothesis that enhancement reflects mainly ILR is based on Zwislocki and Sokolich's (1974) finding that enhancement is absent when the frequency of the stronger Tone 1 is very different from that of Tones 2 and 3. They interpreted this result to mean that Tone 2 had to be at the same frequency as Tone 1. We interpret it to mean that it is Tone 3 that must be at the same frequency as Tone 1. We tested our interpretation on another 11 undergraduates by delaying Tone 2 by 100 ms [see Fig. 1(a)] and setting it to 2500 Hz in one block of trials and to 500 Hz in another. (Tones 1 and 3 were at 500 Hz throughout.) Tone 2 was varied according to the adaptive procedure while Tone 3 was fixed at either 60 or 70 dB SPL. The ILR prediction is that the introduction of the stronger Tone 1 (at 80 dB SPL) will cause Tone 2—whether 500 or 2500 Hz—to be set lower to match Tone 3. The enhancement prediction is that the stronger Tone 1 will affect the match only when Tone 2 is at the same frequency as Tone 1. The outcome was that the 2500-Hz Tone 2 was set 8 to 10 dB lower with the stronger Tone 1 present. This result fully supports the ILR hypothesis and contradicts the enhancement hypothesis. However, no significant effect of the stronger Tone 1 was found with Tone 2 at 500 Hz. This result contradicts both hypotheses and is unlike the results of Zeng (1994). As noted above, in a very similar set of measurements on three listeners, Zeng measured a considerable decrease in the level of Tone 2, just as expected since Tone 3 was always at the same frequency as the stronger Tone 1. Our listeners appear to have had more trouble in making their judgments when all three frequencies were the same than when the second one was different. Although we cannot consider our results as more than suggestive, they certainly tend to support our interpretation of loudness enhancement as the result of a reduction in the loudness of Tone 3 in the enhancement triad.

There remains the nagging question of why a stronger sound should have a marked effect on the loudness of a sound hundreds of milliseconds later and apparently little or no effect on the loudness of an immediately following sound. We can only guess that the proximate sound, sharing the same interval of temporal loudness summation, is processed together with the stronger sound. This integration appears also to be effective for a weaker sound that *precedes* a stronger sound. Elmasian *et al.* (1980) and Plack (1996) found enhancement when a weaker tone preceded a stronger tone by 100 ms. A third comparison tone had to be at a higher level than the weaker first tone to match its loudness. The conditions were just right for ILR because the third comparison tone was presented 500 or 1500 ms after the stronger tone and at the same frequency. Hence, the loudness of the third tone must have been reduced by the preceding stronger tone. Again, the apparent enhancement can be explained by ILR and there is little reason to believe that a change occurred in the loudness of the initial weak tone. Thus, we are left with the implication that a weak tone presented in close temporal proximity to a stronger tone somehow is protected from ILR. The protection occurs whether the weak tone is presented closely before or closely after the stronger tone, but the nature of the mechanism underlying this effect remains an open question.

CONCLUSIONS

Previous studies of loudness enhancement showed that two tone bursts (Tones 2 and 3) of the same frequency, level, and duration were not equal in loudness when a stronger Tone 1 was present. Tone 3 had to be higher in level to match the loudness of Tone 2. The straightforward interpretation of this finding was that this increase in the level of Tone 3 reflected an increase or enhancement of the loudness of Tone 2. This interpretation was perfectly reasonable on the basis of what was known at the time about sequential effects in loudness. There was no reason to consider the possibility that the stronger Tone 1 had, in fact, reduced the loudness of Tone 3. However, many recent experiments demonstrate that stronger sounds reduce the loudness of temporally remote, weaker sounds. Hence, what appeared to be an enhancement of the loudness of Tone 2 was, in fact, a reduction of the loudness of Tone 3. Preliminary experiments support this interpretation; they also indicate that the loudness of Tone 2 appears to be neither enhanced nor reduced by the immediately preceding stronger Tone 1 or by the stronger tones on earlier trials, but it remains unclear why that is so.

ACKNOWLEDGMENTS

The authors are grateful to Robert and Martha Teghtsoonian, Rhona Hellman, Scott Parker, and Erv Hafter for helpful comments. We thank Lilian Moriarty for testing some of the listeners. Part of this research was supported by NIH/NIDCD Grant No. R01DC02241.

Elmasian, R., and Galambos, R. (**1975**). ''Loudness enhancement: monaural, binaural, and dichotic,'' J. Acoust. Soc. Am. **58**, 229–234.
Elmasian, R., Galambos, R., and Bernheim, A. (**1980**). ''Loudness enhancement and decrement in four paradigms,'' J. Acoust. Soc. Am. **67**, 601–607.
Mapes-Riordan, D., and Yost, W. A. (**1999**). ''Loudness recalibration as a function of level,'' J. Acoust. Soc. Am. **106**, 3506–3511.
Marks, L. E. (**1994**). '' 'Recalibrating' the auditory system: The perception of loudness,'' J. Exp. Psychol. **20**, 382–396.
Nieder, B., Buus, S., Florentine, M., and Scharf, B. (**2002**). ''How the duration and level of intense tones affect the loudness reduction (recalibration) of subsequent weaker tones,'' in Proceedings 25th Annual Mid-Winter Meeting of the Association for Research in Otolaryngology, p. 194.
Plack, C. J. (**1996**). ''Loudness enhancement and intensity discrimination under forward and backward masking,'' J. Acoust. Soc. Am. **100**, 1024–1030.
Scharf, B. (**2001**). ''Sequential effects in loudness,'' in *Fechner Day 2001: Proceedings of the Seventeenth Annual Meeting of the International Society for Psychophysics*, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Pabst, Berlin), pp. 254–259.
Zeng, F.-G. (**1994**). ''Loudness growth in forward masking: Relation to intensity discrimination,'' J. Acoust. Soc. Am. **96**, 2127–2132.
Zwislocki, J. J., and Sokolich, W. G. (**1974**). ''On loudness enhancement of a tone burst by a preceding tone burst,'' Percept. Psychophys. **16**, 87–90.

Inverse calculation of material parameters for a thin-layer system using transient elastic waves

Chien-Ching Ma[a)]
Department of Mechanical Engineering, National Taiwan University, Taipei, Taiwan 10617, Republic of China

Shaw-Wen Liu
Department of Electrical Engineering, Chin Min College, 110 Shyue Fu Road, Tou-Fen, Taiwan 351, Republic of China

Chen-Ming Chang
Department of Mechanical Engineering, National Taiwan University, Taipei, Taiwan 10617, Republic of China

(Received 27 April 2001; revised 15 April 2002; accepted 23 May 2002)

The inverse calculation of material parameters of a thin-layer system is investigated using transient elastic waves. The inverse problem is formulated as an optimization problem in which the norm of the discrepancies between the calculated and measured normal surface displacements is minimized through the simplex algorithm. The theoretical result is first solved using the Laplace transform and the transient response is then implemented analytically by Cagniard's method. In the experiment, the source time function is generated by the brittle fracture of a pencil lead on the surface of the thin-layer system, and a National Bureau of Standards (NBS) conical transducer is used to record the surface responses. To obtain reliable inverse results for material parameters, a two-step inverse calculation procedure is proposed. The recovered material parameters of the specimens agree well with the theoretical values and experimental results. [DOI: 10.1121/1.1496763]

PACS numbers: 43.20.Bi, 43.20.El, 43.38.Fx [ANN]

I. INTRODUCTION

Elastic waves carry both amplitude and polarization information and travel at distinct wave velocities. In an inhomogeneous elastic body, dilatational waves convert to shear waves and vice versa at each location in the medium at which the mechanical parameters undergo a change. Measured surface displacements resulting from a dynamic loading functionally depend upon a complete set of subsurface material parameters such as density, elastic moduli, and geometry. This implies the possibility of using measured surface responses to develop a nondestructive test to determine the dynamic elastic parameters. The objective of this study is to recover the parameters of the thin layer in a layered medium using the normal responses of the medium subjected to a normal dynamic loading.

Wave propagation in layered medium has been the focus of considerable studies for many years, mainly because of its significance in various technological and geophysical problems. A large amount of forward problems for wave propagation in elastic medium have been investigated extensively since 1904, when Lamb (1904) studied the surface motions of an elastic, homogeneous, and isotropic half-space generated by the localized surface source. Since then, there have been numerous contributions as summarized by Achenbach (1973), Miklowitz (1978), and van der Hijden (1987). Recently, an excellent review of the progress in this area was given by Every *et al.* (2001). Roughly speaking, there have been three types of approaches to studying these problems: the transform method, the method of generalized rays, and the method of normal modes. In fact, all these methods are intimately connected, even though they are different in spirit. By using Fourier transform, the method for calculating the dynamic response at surface and interior points of a semi-infinite anisotropic elastic continuum to a point load suddenly applied at the surface was established by Every *et al.* (1997). The surface response is reduced to a one-dimensional integral for numerical evaluation. This technique can be extended for the problem of layered media by using Green's function approach derived by Zhang *et al.* (1998). Recently, Lee and Ma (2000) developed an effective analytical method to solve transient full-field solutions in the layered medium using the transform method. Then, the experimental results for the transient responses of a thin layer overlaying a half-space were given to verify the numerical results based on the analytical method (Ma and Lee, 2000).

Compared with the forward problems, inverse problems are more difficult because they usually involve a lot of computations of forward problems. Therefore, effectiveness and precision are prerequisites for the forward problem before further study of the inverse problem. The forward solution is well understood today. Recently, attention has been focused on the inverse problem due to its practical applications. The determination of the elastic parameters of the layers making up a stratified half-space remains a fundamental problem in theoretical seismology resulting from the needs of the seismic prospecting industry to delineate underground structural

a)Electronic mail: cma@w3.me.ntu.edu.tw

features with great precision. The understanding of coated layer-material parameters is also of considerable importance for the practical applications in electronic and mechanical engineering. Clarke (1984) developed a method of reconstructing the elastic parameters as functions of depth, for a horizontally stratified, isotropic elastic half-space. The method makes use of *P-SV* data at a single stacking parameter to obtain the density and both longitudinal and shear wave speeds. The data required are the elements of the full reflection matrix at the surface. A fast algorithm that recursively generates the parameter profiles of density and Lamé constants as functions of depth in an elastic medium was presented by Yagle and Levy (1985). A deconvolution method for multiple Green's functions was applied to experimental data to determine the orientation and time history of an oblique force exerted on the surface of an elastic plate (Michaels and Pao, 1986). Sacks and Symes (1987) studied the problem of recovering the elastic parameters of a layered half-space from single-component measurements of reflected waves. The Rayleigh–Ritz technique was chosen by Deobald and Gibson (1988) to model the vibrations of rectangular orthotropic plates. The natural frequencies were then used to determine the four apparent elastic constants including two Young's moduli, the in-plane shear modulus, and a Poisson's ratio. An ultrasonic method for simultaneous determination of the layer thickness, density, elastic moduli, and attenuation from normal and oblique incidence reflection (transmission) frequency spectra was proposed by Lavrentyev and Rokhlin (1997). The sensitivity of the algorithm to the layer parameters and stability against random noise were studied. By making use of the resonant frequencies obtained from electronic speckle pattern interferometry, an inverse evaluation of the material constants for composite plates using the Rayleigh–Ritz technique and the simplex method was developed by Ma and Lin (1999).

In this study, the inverse problem is formulated as an optimization problem in which the norm of the discrepancies between the calculated and measured vertical surface displacements is minimized using the simplex method. The analytical result is first solved using a double-Laplace transform and the transient solution is then implemented analytically by Cagniard's method (Cagniard, 1939). The transient solutions are exact, and are expressed in an infinite series form of the ray integrals. Since the transient response for the thin-layer medium is exact up to the arrival of the next ray, only a finite number of rays will be involved in the numerical calculation. In the experiment, the source time function is generated by the breakage of a pencil lead on the surface of the thin-layer system, and an NBS conical transducer is used to record the vertical surface displacements. It is noted in this study that the theoretical solution is derived for the case of line-source/line-receiver, while the experimental setup is a point-source/point-receiver arrangement. The errors versus the variations of the parameters were studied before performing inverse calculations. The results show that we are able to reconstruct the material parameters, all quantities being in close agreement with the original specimens.

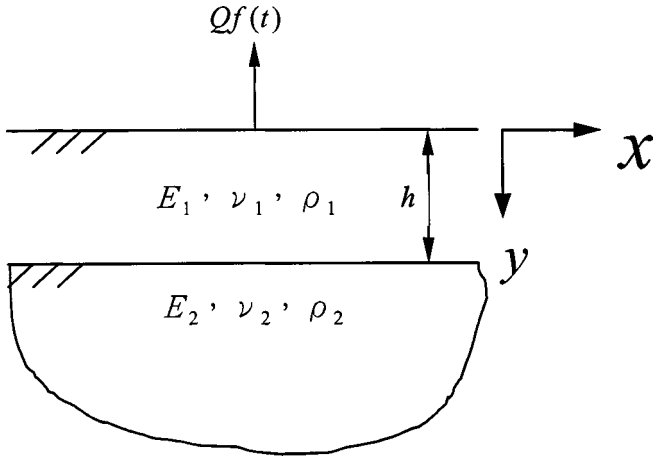


FIG. 1. Schematic representation of a thin-layer system subjected to a normal force at the free surface.

II. FORMULATION FOR THE FORWARD PROBLEM

A forward solution is prerequisite to the inverse calculation studied here. Thus, an effective method is presented in this section. The forward problem is solved using the Laplace transform; the inverse transform is then evaluated by means of Cagniard's method.

Consider a thin-layer system that is composed of a half-space of density ρ_2, Young's modulus E_2, and Poisson's ratio ν_2 overlaying by a layer of thickness h, density ρ_1, Young's modulus E_1, and Poisson's ratio ν_1, as shown in Fig. 1. The quantities related to the layer and the half-space are suffixed by subscript 1 and 2, respectively. A Cartesian coordinate system is oriented so that the y axis is normal to the surface. The surface lies in the plane $y=0$. Since only in-plane motion is considered here, the two-dimensional elastodynamic equation without body forces can be expressed in terms of two scalar potentials, longitudinal wave potential ϕ and shear wave potential ψ, as follows:

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}=S_L\frac{\partial^2\phi}{\partial t^2}, \tag{1}$$

$$\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}=S_T\frac{\partial^2\psi}{\partial t^2}, \tag{2}$$

where

$$S_L=\sqrt{\frac{\rho}{\lambda+2\mu}}=\frac{1}{C_L},\quad S_T=\sqrt{\frac{\rho}{\mu}}=\frac{1}{C_T}.$$

Here, λ and μ are Lamé's constants, S_L and S_T are the slowness of longitudinal and shear waves, and C_L and C_T are the longitudinal and shear wave velocities, respectively. The relations between displacements and potentials are given by

$$u=\frac{\partial\phi}{\partial x}+\frac{\partial\psi}{\partial y}, \tag{3}$$

$$v=\frac{\partial\phi}{\partial y}-\frac{\partial\psi}{\partial x}, \tag{4}$$

where u and v are the displacements in the x- and y directions, respectively. The stresses can be expressed in terms of

two potentials by means of Hook's law. The stresses are given by

$$\sigma_{xx}=\lambda\left(\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}\right)+2\mu\left(\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\psi}{\partial x\partial y}\right), \tag{5}$$

$$\sigma_{yy}=\lambda\left(\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}\right)+2\mu\left(\frac{\partial^2\phi}{\partial y^2}-\frac{\partial^2\psi}{\partial x\partial y}\right), \tag{6}$$

$$\sigma_{xy}=\mu\left(2\frac{\partial^2\phi}{\partial x\partial y}-\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}\right). \tag{7}$$

A normal point loading applied at the free surface is considered. Thus, the boundary conditions at the surface are

$$\sigma_{yy}(x,0,t)=Q\delta(x)f(t), \tag{8}$$

$$\sigma_{xy}(x,0,t)=0. \tag{9}$$

In addition to these boundary conditions, the continuity conditions for displacements and stresses at the interface ($y=h$) between the layer and half-space have to be taken into account

$$\begin{Bmatrix}\sigma_{yy}\\ \sigma_{xy}\\ u\\ v\end{Bmatrix}_1=\begin{Bmatrix}\sigma_{yy}\\ \sigma_{xy}\\ u\\ v\end{Bmatrix}_2. \tag{10}$$

As mentioned before, the problem will be solved using the integral transformation. The one-sided Laplace transform over time t and the bilateral Laplace transform on the spatial variable x for a function f is defined as

$$\overline{f}^*(y;\eta,p)=\int_{-\infty}^{\infty}e^{-p\eta x}\int_0^{\infty}f(x,y,t)e^{-pt}dt\,dx, \tag{11}$$

where p is a positive real number, large enough to ensure the convergence of the integral, and η is a complex variable. The equations of motion [Eqs. (1) and (2)] in the Laplace transform domain are two ordinary differential equations with the following general solutions:

$$\overline{\phi}^*=A(p,\eta)e^{-p\gamma_L y}+B(p,\eta)e^{p\gamma_L y}=\overline{\phi}_b^*+\overline{\phi}_t^* \tag{12}$$

$$\overline{\psi}^*=C(p,\eta)e^{-p\gamma_T y}+D(p,\eta)e^{p\gamma_T y}=\overline{\psi}_b^*+\overline{\psi}_t^*, \tag{13}$$

where

$$\gamma_L=\sqrt{S_L^2-\eta^2},\quad \gamma_T=\sqrt{S_T^2-\eta^2},$$

$$\overline{\phi}_b^*=A(p,\eta)e^{-p\gamma_L y},\quad \overline{\phi}_t^*=B(p,\eta)e^{p\gamma_L y},$$

$$\overline{\psi}_b^*=C(p,\eta)e^{-p\gamma_T y},\quad \overline{\psi}_t^*=D(p,\eta)e^{p\gamma_T y}.$$

The unknown coefficients A, B, C, and D can be determined by boundary and continuity conditions. The condition $\mathrm{Re}\,\gamma_L\geqslant 0$ ($\mathrm{Re}\,\gamma_T\geqslant 0$) is satisfied by providing a branch cut along $S_L\leqslant|\mathrm{Re}\,\eta|\leqslant\infty$ ($S_T\leqslant|\mathrm{Re}\,\eta|\leqslant\infty$), $\mathrm{Im}\,\eta=0$, and choosing the branch of positive square root. Note that the quantities $\overline{\phi}_b^*$ and $\overline{\psi}_b^*$ with subscript b denote the waves propagating along the $+y$ direction and those with subscript t are the waves propagating along the $-y$ direction. Then, the displacement and stress fields in the transformed domain can be expressed as

$$\overline{u}^*=p(\eta\overline{\phi}_t^*+\gamma_T\overline{\psi}_t^*)+p(\eta\overline{\phi}_b^*-\gamma_T\overline{\psi}_b^*), \tag{14}$$

$$\overline{v}^*=p(\gamma_L\overline{\phi}_t^*-\eta\overline{\psi}_t^*)-p(\gamma_L\overline{\phi}_b^*+\eta\overline{\psi}_b^*), \tag{15}$$

$$\overline{\sigma}_{xx}^*=\mu p^2[(S_T^2-2S_L^2+2\eta^2)\overline{\phi}_t^*+2\eta\gamma_T\overline{\psi}_t^*]+\mu p^2[(S_T^2-2S_L^2+2\eta^2)\overline{\phi}_b^*-2\eta\gamma_T\overline{\psi}_b^*], \tag{16}$$

$$\overline{\sigma}_{yy}^*=\mu p^2[(S_T^2-2\eta^2)\overline{\phi}_t^*-2\eta\gamma_T\overline{\psi}_t^*]+\mu p^2[(S_T^2-2\eta^2)\overline{\phi}_b^*+2\eta\gamma_T\overline{\psi}_b^*], \tag{17}$$

$$\overline{\sigma}_{xy}^*=\mu p^2[2\eta\gamma_L\overline{\phi}_t^*+(S_T^2-2\eta^2)\overline{\psi}_t^*]+\mu p^2[-2\eta\gamma_L\overline{\phi}_b^*+(S_T^2-2\eta^2)\overline{\psi}_b^*]. \tag{18}$$

The expressions shown from Eqs. (14)–(18) are the fundamental solutions in the transformed domain. Once the unknown coefficients are determined from the boundary conditions, the transient solutions can be obtained by applying the Cagniard method of Laplace inversion. In order to solve the layer half-space system considered in this study, the wave propagation of the problem can be classified into the following three categories.

A. Waves caused by a point loading at the surface of a homogeneous half-space

According to Eqs. (12) and (13), the wave potentials for a homogeneous half-space can be written as follows:

$$\overline{\phi}_0^*=A_0e^{-p\gamma_{L1}y},\quad \overline{\psi}_0^*=C_0e^{-p\gamma_{T1}y}, \tag{19}$$

where A_0 and C_0 are unknown coefficients determined by the transformed boundary conditions from Eqs. (8) and (9)

$$\overline{\sigma}_{yy}^*=Q\overline{f}(p),\quad \overline{\sigma}_{xy}^*=0. \tag{20}$$

Using Eqs. (17), (18), (19), and (20), one gets

$$\begin{bmatrix}A_0\\ C_0\end{bmatrix}=\frac{Q\overline{f}(p)}{\mu_1 p^2R(\eta)}\begin{bmatrix}S_{T1}^2-2\eta^2\\ 2\eta\gamma_{L1}\end{bmatrix}, \tag{21}$$

where $R(\eta)=(S_{T1}^2-2\eta^2)^2+4\eta^2\gamma_{L1}\gamma_{T1}$ is the well-known Rayleigh wave equation.

B. Waves propagating from layer to interface ($y=h$)

Both reflected and refracted waves can be created for waves incident upon an interface. For the sake of simplicity, the incident longitudinal and shear waves are considered separately.

First, for incident longitudinal waves, the wave potential can be written as

$$\overline{\phi}_i^*=A_i(p,\eta)e^{-p\gamma_{L1}y}. \tag{22}$$

The corresponding reflected and refracted wave potentials, which consist of both longitudinal and shear waves, are expressed as

$$\overline{\phi}_r^*=B_re^{p\gamma_{L1}y},\quad \overline{\psi}_r^*=D_re^{p\gamma_{T1}y}, \tag{23}$$

and

$$\overline{\phi}_f^*=A_fe^{-p\gamma_{L2}y},\quad \overline{\psi}_f^*=C_fe^{-p\gamma_{T2}y}, \tag{24}$$

respectively. The unknown coefficients B_r, D_r, A_f, and C_f are determined by the continuity conditions at the interface, and given by

$$\frac{B_r}{A_i}=\frac{\beta_r}{\Delta}e^{-2p\gamma_{L1}h}, \tag{25}$$

$$\frac{D_r}{A_i}=\frac{\delta_r}{\Delta}e^{-p(\gamma_{L1}+\gamma_{T1})h}, \tag{26}$$

$$\frac{A_f}{A_i}=\frac{\alpha_f}{\Delta}e^{-p(\gamma_{L1}-\gamma_{L2})h}, \tag{27}$$

$$\frac{C_f}{A_i}=\frac{\chi_f}{\Delta}e^{-p(\gamma_{L1}-\gamma_{T2})h}, \tag{28}$$

where

$$\begin{aligned}
\Delta&=4\eta^2(r-1)^2\gamma_{L1}\gamma_{T1}\gamma_{L2}\gamma_{T2}+r(2\eta^2+k_1)(2\eta^2+k_2)\\
&\quad\times(\gamma_{L1}\gamma_{T2}+\gamma_{L2}\gamma_{T1})+(2\eta^2+k_1)^2\gamma_{L2}\gamma_{T2}\\
&\quad+(k_2r+2\eta^2)^2\gamma_{L1}\gamma_{T1}+(k_2r-k_1)^2\eta^2,\\
\beta_r&=4\eta^2(r-1)^2\gamma_{L1}\gamma_{T1}\gamma_{L2}\gamma_{T2}+r(2\eta^2+k_1)\\
&\quad\times(2\eta^2+k_2)(\gamma_{L1}\gamma_{T2}-\gamma_{L2}\gamma_{T1})-(2r\eta^2+k_1)^2\\
&\quad\times\gamma_{L2}\gamma_{T2}+(k_2r+2\eta^2)^2\gamma_{L1}\gamma_{T1}-(k_2r-k_1)^2\eta^2,\\
\delta_r&=4\eta(1-r)(k_1+2r\eta^2)\gamma_{L1}\gamma_{L2}\gamma_{T2}+2\eta(k_2r+2\eta^2)\\
&\quad\times(k_1-k_2r)\gamma_{L1},\\
\alpha_f&=2(k_1+2r\eta^2)(k_1+2\eta^2)\gamma_{L1}\gamma_{T2}+2(k_2r+2\eta^2)\\
&\quad\times(k_1+2\eta^2)\gamma_{L1}\gamma_{T1},\\
\chi_f&=4\eta(r-1)(k_1+2\eta^2)\gamma_{L1}\gamma_{T1}\gamma_{L2}+2\eta(k_1-k_2r)\\
&\quad\times(k_1+2\eta^2)\gamma_{L1},\\
r&=\mu_2/\mu_1,\quad k_1=\gamma_{T1}^2-\eta^2,\quad k_2=\gamma_{T2}^2-\eta^2.
\end{aligned}$$

Second, the incident shear wave potential and its reflected and refracted wave potentials can be written, respectively, in the forms

$$\overline{\psi}_i^*=C_i(p,\eta)e^{-p\gamma_{T1}y}, \tag{29}$$

$$\overline{\phi}_r^*=B_r^*e^{p\gamma_{L1}y},\quad \overline{\psi}_r^*=D_r^*e^{p\gamma_{T1}y}, \tag{30}$$

$$\overline{\phi}_f^*=A_f^*e^{-p\gamma_{L2}y},\quad \overline{\psi}_f^*=C_f^*e^{-p\gamma_{T2}y}. \tag{31}$$

Similarly, the coefficients B_r^*, D_r^*, A_f^*, and C_f^* can be obtained from the continuity conditions

$$\frac{B_r^*}{C_i}=\frac{\beta_r^*}{\Delta}e^{-p(\gamma_{L1}+\gamma_{T1})h}, \tag{32}$$

$$\frac{D_r^*}{C_i}=\frac{\delta_r^*}{\Delta}e^{-2p\gamma_{T1}h}, \tag{33}$$

$$\frac{A_f^*}{C_i}=\frac{\alpha_f^*}{\Delta}e^{-p(\gamma_{T1}-\gamma_{L2})h}, \tag{34}$$

$$\frac{C_f^*}{C_i}=\frac{\chi_f^*}{\Delta}e^{-p(\gamma_{T1}-\gamma_{T2})h}, \tag{35}$$

where

$$\begin{aligned}
\beta_r^*&=4\eta(r-1)(k_1+2r\eta^2)\gamma_{L2}\gamma_{T1}\gamma_{T2}-2\eta(k_1-k_2r)\\
&\quad\times(k_2r+2\eta^2)\gamma_{T1},\\
\delta_r^*&=4\eta^2(r-1)^2\gamma_{L1}\gamma_{T1}\gamma_{L2}\gamma_{T2}+r(2\eta^2+k_1)\\
&\quad\times(2\eta^2+k_2)(\gamma_{L1}\gamma_{T1}-\gamma_{L1}\gamma_{T1})-(2r\eta^2+k_1)^2\\
&\quad\times\gamma_{L2}\gamma_{T2}+(k_2r+2\eta^2)^2\gamma_{L1}\gamma_{T1}-(k_2r-k_1)^2\eta^2,\\
\alpha_f^*&=4\eta(1-r)(k_1+2\eta^2)\gamma_{L1}\gamma_{T1}\gamma_{T2}+2\eta(k_2r+k_1)\\
&\quad\times(k_1+2\eta^2)\gamma_{T1},\\
\chi_f^*&=2(k_1+2r\eta^2)(k_1+2\eta^2)\gamma_{L2}\gamma_{T1}+2\eta(k_2r-k_1)\\
&\quad\times(k_1+2\eta^2)\gamma_{T1}.
\end{aligned}$$

From the derivation shown above, the waves generated from the interface can be written in a more compact form through the transfer matrices. The reflected and refracted transfer matrices are defined as

$$\begin{aligned}
H^r&=\begin{bmatrix}H_{pp}^r & H_{sp}^r\\ H_{ps}^r & H_{ss}^r\end{bmatrix}\\
&=\begin{bmatrix}\frac{\beta_r}{\Delta}e^{-2p\gamma_{L1}h} & \frac{\beta_r^*}{\Delta}e^{-p(\gamma_{L1}+\gamma_{T1})h}\\ \frac{\delta_r}{\Delta}e^{-p(\gamma_{L1}+\gamma_{T1})h} & \frac{\delta_r^*}{\Delta}e^{-2p\gamma_{T1}h}\end{bmatrix},
\end{aligned} \tag{36}$$

$$\begin{aligned}
H^f&=\begin{bmatrix}H_{pp}^f & H_{sp}^f\\ H_{ps}^f & H_{ss}^f\end{bmatrix}\\
&=\begin{bmatrix}\frac{\alpha_f}{\Delta}e^{-p(\gamma_{L1}-\gamma_{L2})h} & \frac{\alpha_f^*}{\Delta}e^{-p(\gamma_{T1}-\gamma_{L2})h}\\ \frac{\chi_f}{\Delta}e^{-p(\gamma_{L1}-\gamma_{T2})h} & \frac{\chi_f^*}{\Delta}e^{-p(\gamma_{T1}-\gamma_{T2})h}\end{bmatrix},
\end{aligned} \tag{37}$$

where H_{pp}^r and H_{sp}^r are the amplitude transfer functions of reflected longitudinal waves caused by incident longitudinal and shear waves, respectively. The other elements in the transfer matrices have similar definitions.

C. Waves propagating from layer to free surface ($y=0$)

The incident longitudinal and shear wave potentials can be written as

$$\overline{\phi}_i^*=Be^{p\gamma_{L1}y},\quad \overline{\psi}_i^*=De^{p\gamma_{T1}y}. \tag{38}$$

Due to the traction-free condition at the surface, there are no refracted waves in this case. The reflected longitudinal and shear potentials are

$$\overline{\phi}_r^*=Ae^{-p\gamma_{L1}y},\quad \overline{\psi}_r^*=Ce^{-p\gamma_{T1}y}. \tag{39}$$

Taking into account the boundary conditions ($\overline{\sigma}_{yy}^*=0, \overline{\sigma}_{xy}^*=0$) at the free surface, the amplitudes of reflected wave potentials (A and C) can be expressed in terms of the amplitudes of incident wave potentials (B and D) and transfer matrix (H^t) at the free surface

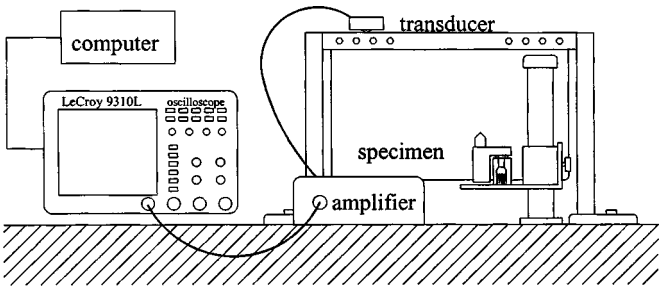


FIG. 2. The thin-layer system of two-dimensional acoustical emission experiment.

$$\begin{bmatrix} A \\ C \end{bmatrix} = H^t \begin{bmatrix} B \\ D \end{bmatrix}, \tag{40}$$

where

$$H^t = \frac{1}{R(\eta)} \begin{bmatrix} 4\eta^2 \gamma_{L1}\gamma_{T1} - k_1^2 & 4\eta\gamma_{T1}k_1 \\ -4\eta\gamma_{L1}k_1 & 4\eta^2\gamma_{L1}\gamma_{T1} - k_1^2 \end{bmatrix}. \tag{41}$$

Based on the transfer matrices at the free surface and interface, the entire wave potentials in the layer half-space can be obtained easily. The potentials corresponding to waves propagating in the $+y$ and $-y$ directions in the layer are given by

$$\begin{bmatrix} \bar{\phi}_{2n}^* \\ \bar{\psi}_{2n}^* \end{bmatrix} = (H^t H^r)^n \begin{bmatrix} \bar{\phi}_0^* \\ \bar{\psi}_0^* \end{bmatrix}, \quad n = 0,1,2,\ldots, \tag{42}$$

and

$$\begin{bmatrix} \bar{\phi}_{2n+1}^* \\ \bar{\psi}_{2n+1}^* \end{bmatrix} = H^r (H^t H^r)^n \begin{bmatrix} \bar{\phi}_0^* \\ \bar{\psi}_0^* \end{bmatrix}, \quad n = 0,1,2,\ldots, \tag{43}$$

respectively. Thus, the complete wave potentials for waves propagating in the layer are obtained by the superposition of the down-going and up-going waves shown in Eqs. (42) and (43)

$$\begin{bmatrix} \bar{\phi}^* \\ \bar{\psi}^* \end{bmatrix} = \sum_{n=0}^{\infty} \begin{bmatrix} \bar{\phi}_{2n}^* + \bar{\phi}_{2n+1}^* \\ \bar{\psi}_{2n}^* + \bar{\psi}_{2n+1}^* \end{bmatrix}. \tag{44}$$

Similarly, the wave potentials in the half-space are given by

$$\begin{bmatrix} \bar{\phi}^* \\ \bar{\psi}^* \end{bmatrix} = \sum_{n=0}^{\infty} (H^t H^r)^n H^f \begin{bmatrix} \bar{\phi}_0^* \\ \bar{\psi}_0^* \end{bmatrix}. \tag{45}$$

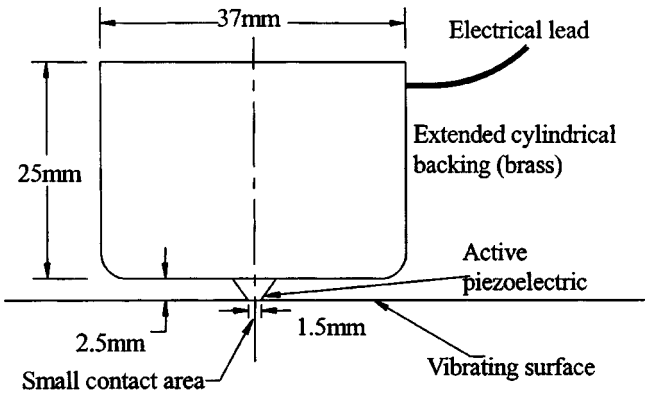


FIG. 3. Schematic diagram of the NBS conical transducer.

TABLE I. Dimensions of the specimens.

Specimen	Dimension (mm)
Brass/aluminum	300×15×1/300×300×1
Acrylic/aluminum	300×13.7×1.8/300×300×1

Once the transformed wave potentials are calculated, the transient responses can be obtained by applying the Cagniard method of Laplace inversion. The Cagniard method is surveyed in detail in the textbook written by Achenbach (1973) and in the review article by Pao and Gajewski (1977). The transient responses of the thin-layer medium are determined by summing all the contribution of waves, and the solution is exact up to the arrival of the next wave.

III. EXPERIMENTAL SETUP

The experimental setup for the measurement of surface responses on a thin-layer medium is shown in Fig. 2. The measured surface responses were used to recover the material parameters of the thin layer. An NBS conical transducer as shown in Fig. 3 was used to measure the transient normal surface displacement of a specimen. A detailed description of the transducer can be found in Proctor (1982) and Greenspan (1987). The basic transducer consists of an active piezoelectric element of conical shape and a large cylindrical backing block made of brass. The truncated end of the conical element is about 1.5 mm in diameter, small compared to a wavelength over the frequency range of 50 kHz to 1 MHz. Thus, small contact size makes the transducer an approximation to a point receiver. The voltage signals received from the NBS transducer were fitted in a preamplifier and then recorded by a digital oscilloscope (LeCroy 9310L) with the fastest sampling rate of 100 Ms/s and vertical resolution of 8 bits.

Experiments were performed by generating normal forces with step-like time functions on the surface of a thin-layer medium. The normal forces were generated by the breakage of pencil leads. Then, the normal surface responses 3 cm away from the source were measured by using the NBS conical transducer. Two specimens, brass/aluminum and acrylic/aluminum, with quite different surface responses, were used to investigate the inverse calculation of material parameters of the layer. The dimensions and material properties of the specimens, are listed in Tables I and II, respectively. The brass and the acrylic, with thickness of 15 and 13.7 mm, respectively, are considered to be the thin layers of the specimens. In order to simulate the thin-layer system, the thickness of the lower plate (aluminum) has been made much larger compared to that of the layer. The interface between the thin layer and lower plate was bonded with epoxy resin (PC-10 adhesive) for a curing time of about 1 h under the

TABLE II. Material properties of the specimens.

Material	ρ (kg/m^3)	E (Gpa)	ν
Brass	8512	110	0.34
Acrylic	1190	6.32	0.316
Aluminum	2700	68.9	0.34

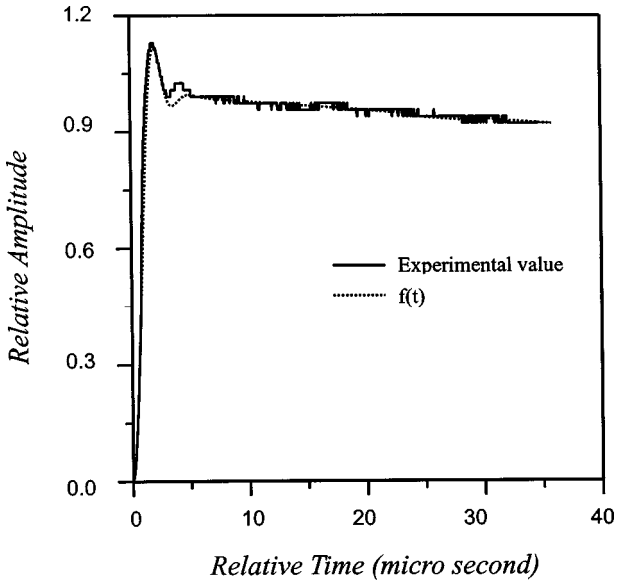


FIG. 4. The source-time function generated by breaking the pencil lead and the corresponding fitting function.

pressure produced by a vise. Both bonding surfaces of the specimens were polished carefully to ensure perfect contact at the interface. Moreover, in order to check the bonding effects caused by epoxy resin, an additional two specimens were prepared to perform a control experiment. The first is composed of two aluminum plates bonded with epoxy resin. The other is an entire aluminum plate. The experimental results for these two specimens and the numerical results from analytical solution are found to be in good agreement. This verifies that not only can the effects of epoxy resin present at the interface be neglected, but also that the numerical accuracy for analytical solutions is reliable.

The experimental value shown in Fig. 4 was obtained by breaking the pencil lead directly on the NBS conical transducer. The step-like time function produced by the breakage of pencil leads has a short rising time of about 1 μs and is reproducible without damaging the sample. Note that the step-like time function received by the transducer decays exponentially as time increases. This exponential decay is caused by the dynamic characteristic of the NBS conical transducer. However, a function can be used to fit the step-like time function as

$$f(t)=e^{-\alpha t}(1-(1-\beta t^2)e^{-\zeta t^2}\cos\chi t)H(t), \tag{46}$$

where the parameter α dominates the effect of exponential decay caused by the NBS conical transducer, ζ and χ characterize the overshoot and ripples, β controls the minor characteristic of the curve, and $H(t)$ is the Heaviside step-time function. The parameters used for $f(t)$ shown in Fig. 4 are given by $\alpha=2.35\times10^3$, $\beta=1.10\times10^{11}$, $\chi=1.05\times10^5$, and $\zeta=2.00\times10^{11}$. It is found that the function $f(t)$ with the given parameters can fit the step-like time function well up to 250 μs. Note that the function of Eq. (46) is twice differentiable, and the first derivative evaluated at the origin is zero. Therefore, this function was used instead of the Heaviside step-time function in the numerical calculation to reduce the error induced by the time function of applied force together with the sensor system.

IV. SIMPLEX METHOD

The objective of this study is to reconstruct the parameters of a thin-layer medium, such as the Poisson's ratio, density, Young's modulus, and thickness of the layer, using the normal transient responses of the medium subjected to a normal surface point load. This study is formulated as an optimization problem in which the discrepancy between the computed and measured normal responses is minimized. The non-negative error function, which is a function of Poisson's ratio, density, Young's modulus, and thickness of the thin layer, can be defined as

$$\varepsilon(\nu,\rho,E,h)=\sum_{i=1}^{n}(v_c(i\Delta t)-v_e(i\Delta t))^2, \tag{47}$$

where Δt is the sampling period, n is the number of samples, and v_c and v_e denote the normal displacements obtained by computed results and experimental measurement, respectively.

There are various optimization schemes that can be used to solve Eq. (47). The success of recovering material parameters is partially dependent on the choice of algorithm. It was shown from past experience that the algorithms based on gradients often face the problem of being trapped in local minimum points. Simplex optimization, based neither on gradients (first-order derivatives) nor on quadratic forms (second-order derivatives), is a well-known method for sequentially optimizing systems having multivariables. Although the problem of being trapped in local minimum points may still exist, it can be remedied by some proper techniques. The simplex method makes use of the information of the error function only, and was first described by Nelder and Mead (1964). Since then, numerous modifications have been made to improve the efficiency with which the algorithm converges to an optimum. A systematic inversion scheme based on the simplex algorithm of curve fitting for piecewise smooth functions was used by Karim *et al.* (1990) to determine certain material properties of plate specimens from leaky Lamb wave data. The scheme has been found to be very efficient and simple compared to the traditional methods of inversion. By careful application of the simplex technique to the acoustic material signature of layered solids, Kundu (1992) obtained the thickness and material properties of the layer. The simplex inversion algorithm was used by Castaings *et al.* (2000) for the initial estimates of plate thickness and plate properties, the Newton–Raphson inversion technique was then adopted to improve the accuracy of convergence. The simplex optimization technique was applied to the acoustic microscopy analysis by Kundu *et al.* (1991, 2000) to obtain the mechanical properties of biological cells. Therefore, the simplex method is adopted in this study. The Poisson's ratio, density, Young's modulus, and thickness of the thin layer as well as an error term are used as coordinates to describe an information space. A simplex is a geometric figure, with one more vertex than variables, which lies within this information space. The vertices

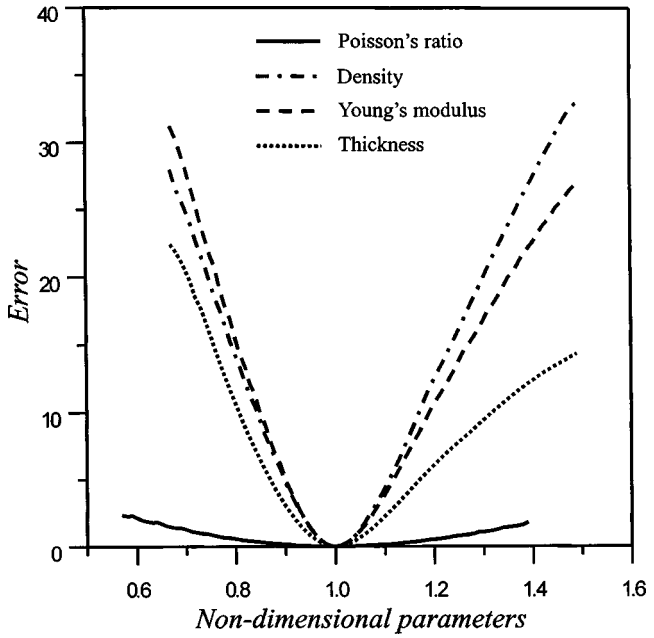


FIG. 5. The errors caused by the variation of material parameters.

of the simplex each represent a particular set of material parameters, thickness of the layer, and the associated error term. If N unknowns will be determined from the simplex optimization scheme, then at least $N+1$ sets of initial guesses should be used in the inverse calculation. The algorithm moves the simplex about the information space, by defining new vertices, such that the simplex response term is maximized and the error term is minimized. Although the simplex method may be less efficient compared to gradient-type schemes, it has the advantage that one can easily restart the iterations from the final point.

It is well known that the results of the inverse calculation strongly depend on the initial guesses and convergence criterion. Sometimes the optimization process may be trapped in a local minimum that results in unacceptable solutions. Hence, one has little chance to get accurate results if only one inverse calculation is performed. The development of a reliable methodology is needed to obtain accurate results and a two-step inverse calculation procedure is proposed in this study. First, $N+1$ sets of inverse calculations are performed. Their inverse results are then used as initial guesses to construct the final inverse results. Usually the results obtained by the second step are better than that in the first step. Even if the optimization process converges to local minimum point in the first step, it can escape from that point automatically in the second step.

V. RESULTS OF INVERSE CALCULATIONS

The interest of this study is to evaluate the parameters of the thin layer in a layer medium, including the Poisson's ratio (ν), density (ρ), Young's modulus (E), and thickness (h). In order to examine and understand the effect of the variations of these parameters on surface normal displacements, the errors versus the variations of nondimensional parameters are shown in Fig. 5, in which the nondimensional parameter equal to 1 denotes the exact value. It is observed that basically all of the errors behave well. This is suitable for the simplex method to recover the material parameters. However, the local minimum points are found at the error function for the Poisson's ratio. Also note that a large variation of Poisson's ratio from exact value induces a relatively small error compared to the other three parameters. This indicates that the Poisson's ratio is insensitive to the normal surface displacements. One may expect that the inverse value of Poisson's ratio will not be as accurate as that of the other three parameters. Before using experimental data for the inversion of material parameters, some theoretical inversions were carried out to investigate the performance of the proposed method. The distance between the source and the receiver is 3 cm for both theoretical and experimental inversions.

A. Theoretical inversion

In the theoretical inversion, only the specimen of brass/aluminum is considered. A forward calculation of the normal displacement base on the material parameters listed in Tables I and II are performed and the transient result generated by the numerical calculation is used as the experimental measured data. The theoretical inversion of the parameters is based on the numerically generated data. First, the inverse calculations were examined under the variation of one parameter each time. It was found that the results of inversion were pretty good, and the local minimum points did not cause much difficulty for the simplex method. Then, the variations of two and more parameters were examined for the inverse calculations. The errors of the inversions are acceptable, but they have a tendency to become larger as the number of parameters used in the inversion increases. As was expected, the difference between the inverse result and the theoretical value for Poisson's ratio is the largest, and that for thickness is the next. If the initial guess deviates far from the exact value, one may encounter difficulty with convergence.

Based on the preceding examination, the reasonable restrictions on the values of parameters are recommended to increase the accuracy of inverse calculation. The velocities of Rayleigh and longitudinal waves can be estimated by their arrival times from the surface responses. Because the Poisson's ratio is a function of Rayleigh and longitudinal wave velocities, the value of Poisson's ratio can be confined to a reasonable range as long as the velocities of longitudinal and Rayleigh waves are restricted to reasonable ranges. In this study, the velocities of longitudinal and Rayleigh waves are restricted according to their arrival times, which are within the ranges of 5–11 μs and 13–17 μs, respectively. Similarly, the thickness of the layer can be confined in a range by estimating the arrival time of the first reflected longitudinal wave from the interface. In addition, a rough estimation can be made for the density and Young's modulus according to the waveform of the surface response, which is strongly related to the acoustic impedances of the layer and bottom medium. By using the reasonable restrictions on the parameters, the results of theoretical inversion are excellent, as shown in Table III, in which six sets of inverse results and one set of theoretical values are listed. The initial guesses were chosen at random within the reasonable ranges for first

TABLE III. The results of theoretical inverse calculations for brass/aluminum.

	ν	ρ (kg/m^3)	E (Gpa)	h (mm)
1st inverse results	0.320	8393.5	108.4	15.4
2nd inverse results	0.362	8594.4	111.5	14.7
3rd inverse results	0.380	9237.6	123.6	16.1
4th inverse results	0.382	9255.8	121.9	15.5
5th inverse results	0.365	8731.7	115.3	15.4
Final inverse results	0.339	8512.5	111.0	15.1
Theoretical values	0.340	8512.0	110.0	15.0
Error (%)	0.29	0.01	0.91	0.67

to fifth inverse calculations. Then, the final (sixth) inverse results were obtained by using the previous five sets of inverse results as the initial guesses. One can see that the five inverse results for the first step deviate from the exact values from 1% to 12%. However, the errors between theoretical values and final inverse results for all of the parameter are within 1%.

B. Experimental inversion

The specimens of brass/aluminum and acrylic/aluminum were used in the experimental inversion. In order to have reliable inverse results for the material parameters, six inverse calculations have been carried out for each set of experimental data. The initial guesses were chosen at random for five of the inverse calculations. These inverse results were then used as the initial guesses for the last inverse calculation to obtain the final inverse results.

For the specimen of brass/aluminum, the acoustic impedance of brass is higher than that of aluminum. This is the case of a stress wave traveling in a high impedance material and striking a boundary with a lower impedance material. Therefore, the waveform measured at the surface would be analogous to that of a stress wave striking a free boundary. Both experimental and theoretical values for the normal surface displacements on the specimen of brass/aluminum are shown in Figs. 6 and 7. It is observed that the theoretical waveforms are not identical to the experimental waveforms. There could be some reasons that cause the difference between theoretical and experimental waveforms: the theoretical values of material parameters are not exactly equal to the real ones, the step-like time function of Eq. (46) is not identical to the experimental one, and the noises and errors may be produced in the experimental measurements.

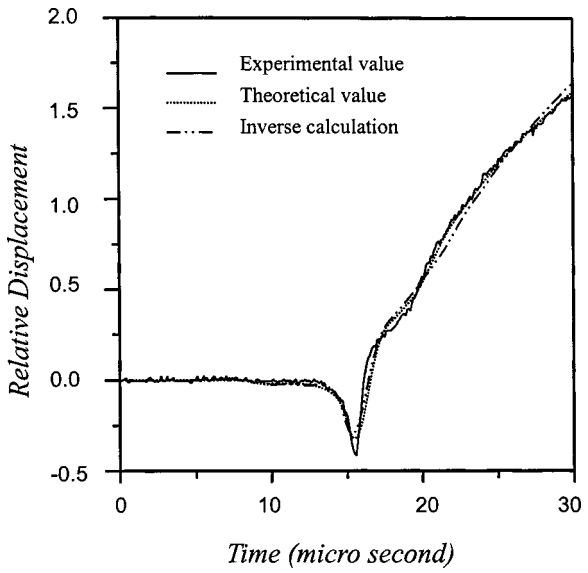


FIG. 6. The waveforms of fifth inverse results corresponding to Table VIII for brass/aluminum.

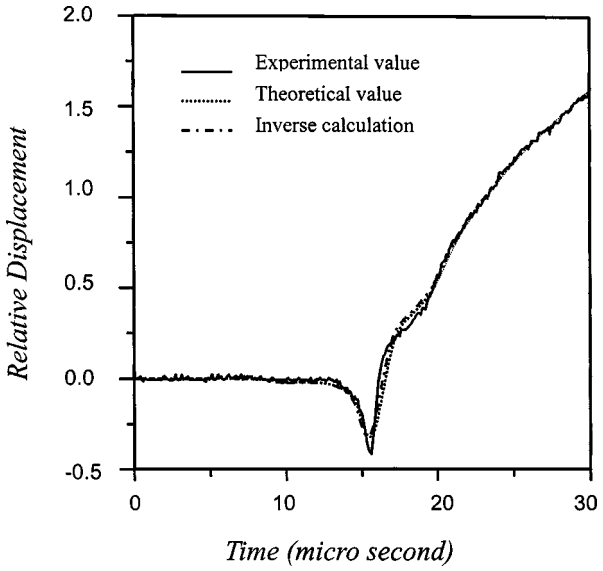


FIG. 7. The waveforms of final (sixth) inverse results corresponding to Table IX using the preceding five inverse results as initial guesses for brass/aluminum.

Since four parameters of the layer are determined simultaneously using the simplex method, five sets of the initial guesses of the four parameters are used for each inverse calculation. The results of five inverse calculations for the specimen of brass/aluminum are given in Tables IV–VIII. Five initial guesses for each inverse calculation are also listed in the tables. Figure 6 shows the waveforms of the vertical displacements corresponding to the results of Table VIII. The final inverse results as shown in Table IX and Fig. 7 are obtained by using the previous five inverse results as initial guesses. Note that the difference between final inverse results and theoretical values is not the smallest one among all of the inverse calculations, because the inverse calculations are based on the experimental results, not on the theoretical values. However, the difference between final inverse results and experimental results is the smallest one. The final

TABLE IV. The first inverse calculation for brass/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Initial guesses	0.32	10 040	76	17.3
	0.41	8 280	85	22.2
	0.25	10 120	106	24.1
	0.42	7 930	78	18.6
	0.40	6 030	116	14.6
Inverse results	0.33	8 027.9	105.7	14.2
Theoretical values	0.34	8512	110	15.0
Error (%)	2.94	5.69	3.91	5.30

TABLE V. The second inverse calculation for brass/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Initial guesses	0.32	10 240	97	13.3
	0.41	10 380	105	19.6
	0.25	11 820	79	18.1
	0.32	7 130	80	22.2
	0.40	6220	116	16.0
Inverse results	0.368	8 322.5	112	14.9
Theoretical values	0.34	8 512	110	15.0
Error (%)	8.24	2.22	1.82	0.67

TABLE VI. The third inverse calculation for brass/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Initial guesses	0.23	10 500	78	16.3
	0.34	7 800	79	37.2
	0.43	11 210	96	24.1
	0.32	10 540	111	16.4
	0.41	11 480	105	14.2
Inverse results	0.279	8 946.6	117.9	18.5
Theoretical values	0.34	8 512	110	15.0
Error (%)	17.9	5.1	7.18	23.3

TABLE VII. The fourth inverse calculation for brass/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Initial guesses	0.32	12 000	86	13.2
	0.40	10 100	97	26.1
	0.32	10 300	117	18.2
	0.28	10 500	99	35.2
	0.26	9 000	87	13.1
Inverse results	0.312	8 196.6	108.2	15.7
Theoretical values	0.34	8 512	110	15.0
Error (%)	8.24	3.71	1.64	4.67

TABLE VIII. The fifth inverse calculation for brass/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Initial guesses	0.25	10 120	99	13.9
	0.42	7 830	88	16.6
	0.40	7 420	76	24.1
	0.23	10 300	94	25.4
	0.34	8 800	113	37.1
Inverse results	0.376	9 718.2	132.8	17.9
Theoretical values	0.34	8 512	110	15.0
Error (%)	10.59	14.17	20.73	19.3

TABLE IX. The final inverse calculation for brass/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Inverse results	0.371	8350.2	112.2	14.9
Theoretical values	0.34	8512	110	15.0
Error (%)	9.12	1.90	2.00	0.7

inverse result should be that closest to the real one as long as the experimental result is accurate enough. It is found in the final inverse results that the error of Poisson's ratio is the largest one compared to other parameters. This is consistent with the results of the error function shown in Fig. 5. It was mentioned in theoretical inverse calculation that the value of Poisson's ratio could be restricted in a reasonable range based on the arrival times of longitudinal and Rayleigh waves. However, the arrival time of longitudinal wave was not obvious in the experimental signals. Therefore, there were no restrictions on the value of Poisson's ratio to avoid introducing artificial errors.

Next, the specimen of acrylic/aluminum is considered. The process of inverse calculation for acrylic/aluminum is similar to that for brass/aluminum. However, the acoustic impedance of acrylic is smaller than those of brass and aluminum. This is the case of a wave propagating from a layer with low impedance to a medium with higher impedance. Therefore, the patterns of vertical displacements at the surface of acrylic/aluminum as shown in Figs. 8 and 9 are quite different from those of brass/aluminum as shown in Figs. 6 and 7. The time used for inverse calculations of acrylic/aluminum is 40 μs instead of 30 μs because the wave velocities in acrylic are slower.

The initial guesses and results of five inverse calculations for the specimen of acrylic/aluminum are shown in Tables X–XIV. The waveforms of vertical displacements corresponding to the results of Table XIV are also shown in Fig. 8. Table XV and Fig. 9 show the final inverse results obtained by using the previous five inverse results as initial guesses. It is observed that the waveforms of final results match very well with those of the experimental values except the region before Rayleigh waves arrive. The error for the inverse result of Poisson's ratio is not the largest one among all of the inverse parameters. The value of Poisson's ratio was restricted in a reasonable range because the arrival times of longitudinal and Rayleigh waves could be easily identified from the signals. Overall, compared with brass/aluminum,

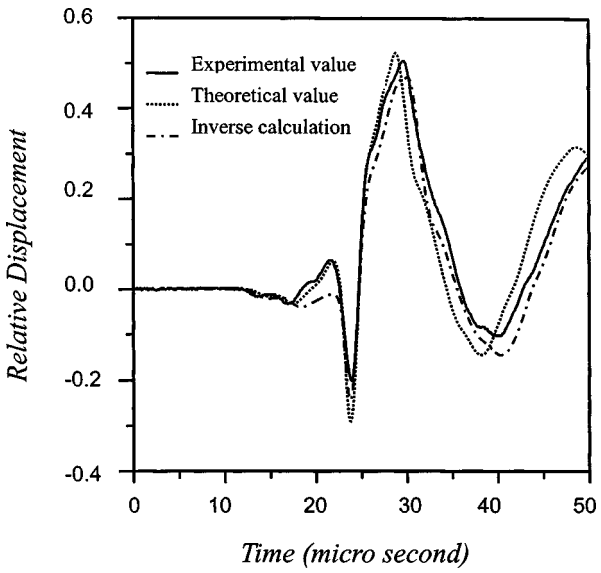


FIG. 8. The waveforms of fifth inverse results corresponding to Table XIV for acrylic/aluminum.

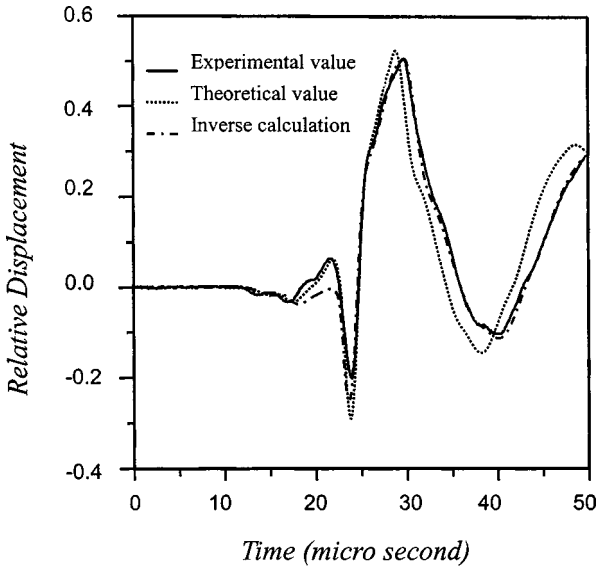


FIG. 9. The waveforms of final (sixth) inverse results corresponding to Table XV using the preceding five inverse results as initial guesses for acrylic/aluminum.

the inverse results for acrylic/aluminum have larger errors. This could be caused by several factors. The decay of stress waves is more serious in acrylic than in brass. This has been found by experiment (results are not shown here). However, the decay of stress waves in medium is not taken into account in the analytical solutions used for inverse calculations. In addition, the theoretical values of material properties for acrylic may not be close enough to the real ones. This can be explained from Figs. 8 and 9, that the waveform calculated by theoretical values does not coincide with the waveform measured by experiment.

VI. CONCLUSIONS

An exact analytical solution using Laplace transform and Cagniard's method has been given in detail for a thin-layer system subjected to a normal point loading. Based on the analytical solutions and experimental results, the inverse calculations of material parameters were carried out for the specimens of brass/aluminum and acrylic/aluminum. The simplex method was used to minimize the norm of the discrepancies between the calculated and measured vertical surface displacements. To obtain reliable inverse results for material parameters, a two-step inverse calculation procedure is proposed in this study. Overall, the recovered material pa-

TABLE X. The first inverse calculation for acrylic/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Initial guesses	0.28	900	7	26
	0.32	1300	6	18
	0.41	880	15	17
	0.29	1400	4	14
	0.34	1000	3	23
Inverse results	0.301	1397.9	7.3	14.4
Theoretical values	0.316	1190.0	6.32	13.7
Error (%)	4.75	17.47	15.51	5.11

TABLE XI. The second inverse calculation for acrylic/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Initial guesses	0.26	900	7	23
	0.32	1200	5	12
	0.40	1100	8	16
	0.32	1300	14	28
	0.28	1050	4	15
Inverse results	0.290	1118.3	5.9	14.8
Theoretical values	0.316	1190.0	6.32	13.7
Error (%)	8.23	6.03	6.65	8.03

TABLE XII. The third inverse calculation for acrylic/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Initial guesses	0.41	1300	6	13
	0.31	850	5	25
	0.36	900	11	25
	0.27	1130	9	14
	0.32	1090	4	16
Inverse results	0.299	1276.6	6.7	14.4
Theoretical values	0.316	1190.0	6.32	13.7
Error (%)	5.38	7.28	6.01	5.11

TABLE XIII. The fourth inverse calculation for acrylic/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Initial guesses	0.38	1400	9.5	16.2
	0.32	1320	6.3	18.1
	0.31	1100	7.1	13.2
	0.34	800	6.8	13.4
	0.41	1480	5.4	17.2
Inverse results	0.347	1356.5	7.5	14.9
Theoretical values	0.316	1190	6.32	13.7
Error (%)	9.81	13.99	18.67	8.76

TABLE XIV. The fifth inverse calculation for acrylic/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Initial guesses	0.41	1000	4	12
	0.31	1100	7	32
	0.34	800	8	14
	0.38	1400	5	16
	0.32	930	3	18
Inverse results	0.360	1062.0	5.8	15.2
Theoretical values	0.316	1190.0	6.32	13.7
Error (%)	13.92	10.76	8.23	10.95

TABLE XV. The final inverse calculation for acrylic/aluminum.

	ν	ρ (kg/m^3)	E (GPa)	h (mm)
Inverse results	0.344	1298.6	7.1	15.0
Theoretical values	0.316	1190.0	6.32	13.7
Error (%)	8.86	9.13	12.34	9.49

rameters including Poisson's ratio, Young's modulus, density, and the thickness of the layer are in good agreement with the experimental results. For a given thin-layer medium with unknown material parameters for the layer, the accuracy of inverse results is strongly dependent on the accuracy of the experiment. Based on the waveforms of the surface responses, the accuracy can be improved further by means of reasonable restrictions on the initial guesses of material parameters for the simplex method. Therefore, the proposed method in this study has the potential for practical applications.

ACKNOWLEDGMENT

The research support of the National Science Council, Republic of China, through Grant No. NSC 89-2212-E002-018 at National Taiwan University is gratefully acknowledged.

Achenbach, J. D. (**1973**). *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam).

Cagniard, L. (**1939**). *Reflexion et Refraction des Ondes Seismiques Progressives* (Cauthiers-Villars); Translated into English and revised by E.A. Flinn and C. H. Dux, 1962. Reflection and Refraction of Progressive Seismic Waves (McGraw Hill, New York).

Castaings, M., Hosten, B., and Kundu, T. (**2000**). ''Inverse of ultrasonic, plane-wave transmission data in composite plates to infer viscoelastic material properties,'' NDT & E Int. **33**, 377–392.

Clarke, T. J. (**1984**). ''Full reconstruction of a layered elastic medium from *P-SV* slant-stack data,'' Geophys. J. R. Astron. Soc. **78**, 775–793.

Deobald, L. R., and Gibson, R. F. (**1988**). ''Determination of elastic constants of orthotropic plates by a modal analysis/Rayleigh–Ritz technique,'' J. Sound Vib. **124**, 269–283.

Every, A. G., Kim, K. Y., and Maznev, A. A. (**1997**). ''The elastodynamic response of a semi-infinite anisotropic solid to sudden surface loading,'' J. Acoust. Soc. Am. **102**, 1346–1355.

Every, A. G., Kim, K. Y., and Sachse, W. (**2001**). ''Point-Source/Point-Receiver Methods,'' in *Handbook of Elastic Properties of Solids, Liquids, and Gases. Volume I: Dynamic Methods for Measuring the Elastic Properties of Solids*, edited by M. Levy, H. Bass, R. Stern, and V. Keppens (Academic, New York), pp. 87–108.

Greenspan, M. (**1987**). ''The NBS conical transducer: Analysis,'' J. Acoust. Soc. Am. **81**, 173–183.

Karim, M. R., Mal, A. K., and Bar-Cohen, Y. (**1990**). ''Inversion of leaky Lamb wave data by simplex algorithm,'' J. Acoust. Soc. Am. **88**, 482–491.

Kundu, T., Bereiter-Hahn, J., and Hillmann, K. (**1991**). ''Measuring elastic properties of cells by evaluation of scanning acoustic microscopy $V(z)$ values using simplex algorithm,'' Biophys. J. **59**, 1194–1207.

Kundu, T. (**1992**). ''Inversion of acoustic material signature of layered solids,'' J. Acoust. Soc. Am. **91**, 591–600.

Kundu, T., Bereiter-Hahn, J., and Karl, I. (**2000**). ''Cell property determination from the acoustic microscope generated voltage versus frequency curves,'' Biophys. J. **78(5)**, 2270–2279.

Lamb, H. (**1904**). ''On the propagation of tremors over the surface of an elastic solid,'' Philos. Trans. R. Soc. London, Ser. A **203**, 1–42.

Lavrentyev, A. I., and Rokhlin, S. I. (**1997**). ''Determination of elastic moduli, density, attenuation, and thickness of a layer using ultrasonic spectroscopy at two angles,'' J. Acoust. Soc. Am. **102**, 3467–3477.

Lee, G. S., and Ma, C. C. (**2000**). ''Transient elastic waves propagating in a multilayered medium subjected to inplane dynamic loadings. I. Theory,'' Proc. R. Soc. London, Ser. A **456**, 1355–1374.

Ma, C. C., and Lin, C. C. (**1999**). ''Inverse evaluation of material constants for composite plates by optical interferometry method,'' AIAA J. **37**, 947–953.

Ma, C. C., and Lee, G. S. (**2000**). ''Transient elastic waves propagating in a multilayered medium subjected to inplane dynamic loadings. II. Numerical calculation and experimental measurement,'' Proc. R. Soc. London, Ser. A **456**, 1375–1396.

Michaels, J. E., and Pao, Y. H. (**1986**). ''Determination of dynamic forces from wave motion measurements,'' Trans. ASME, J. Appl. Mech. **53**, 61–68.

Miklowitz, J. (**1978**). *The Theory of Elastic Waves and Wave Guides* (North-Holland, Amsterdam).

Nelder, J. A., and Mead, R. (**1964**). ''A simplex method for function minimization,'' Comput. J. (UK) **7**, 308–313.

Pao, Y. H., and Gajewski, R. (**1977**). ''The generalized ray theory and transient responses of layered elastic solids,'' in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York), Vol. 13, Chap. 6, pp. 184–266.

Proctor, Jr., T. M. (**1982**). ''Some details on the NBS conical transducer,'' J. Acoust. Emiss. **1**, 173–178.

Sacks, P., and Symes, W. (**1987**). ''Recovery of the elastic parameters of a layered half-space,'' Geophys. J. R. Astron. Soc. **88**, 593–620.

van der Hijden, J. H. M. T. (**1987**). *Propagation of Transient Elastic Waves in Stratified Anisotropic Media* (North-Holland, Amsterdam).

Yagle, A. E., and Levy, B. C. (**1985**). ''A layer-stripping solution of the inverse problem for a one-dimensional elastic medium,'' Geophysics **50**, 425–433.

Zhang, X., Comins, J. D., Every, A. G., Stoddart, P. R., Pang, W., and Derry, T. E. (**1998**). ''Surface Brillouin scattering study of the surface excitations in amorphous silicon layers produced by ion bombardment,'' Phys. Rev. B **58**(20), 13677–13685.

Stable recursive algorithm for elastic wave propagation in layered anisotropic media: Stiffness matrix method

S. I. Rokhlin[a] and L. Wang
The Ohio State University, Nondestructive Evaluation Program, Edison Joining Technology Center, 1248 Arthur E. Adams Drive, Columbus, Ohio 43221

(Received 6 April 2001; revised 2 September 2001; accepted 6 June 2002)

An efficient recursive algorithm, the stiffness matrix method, has been developed for wave propagation in multilayered generally anisotropic media. This algorithm has the computational efficiency and simplicity of the standard transfer matrix method and is unconditionally computationally stable for high frequency and layer thickness. In this algorithm, the stiffness (compliance) matrix is calculated for each layer and recursively applied to generate a stiffness (compliance) matrix for a layered system. Next, reflection and transmission coefficients are calculated for layered media bounded by liquid or solid semispaces. The results show that the method is stable for arbitrary number and thickness of layers and the computation time is proportional to the number of layers. It is shown both numerically and analytically that for a thick structure the solution approaches the solution for a semispace. This algorithm is easily adaptable to laminates with periodicity, such as multiangle lay-up composites. The repetition and symmetry of the unit cell are naturally incorporated in the recursive scheme. As an example the angle beam time domain pulse reflections from fluid-loaded multilayered composites have been computed and compared with experiment. Based on this method, characteristic equations for Lamb waves and Floquet waves in periodic media have also been determined. [DOI: 10.1121/1.1497365]

PACS numbers: 43.20.Bi, 43.20.Fn, 43.20.Gp [ANN]

I. INTRODUCTION

The topic of elastic wave propagation in layered media has enjoyed significant attention for many decades due to the wealth of its applications to seismology, acoustics, and NDE.[1–20] These applications have been facilitated by the development of the very simple and computationally efficient transfer matrix method;[1–3] however, it was soon discovered that it is inherently computationally unstable for layer thicknesses of several wavelengths. The delta matrix method[4] has been proposed to resolve the instability problem for isotropic systems. For the latest improvements and reviews see Refs. 7 and 8.

The problem of layered generally anisotropic media is much more complicated. A transfer matrix method using the exponential Stroh representation was developed by Adler and co-workers.[5,9] The transfer matrix method in standard form was developed for general anisotropy by Nayfeh.[12,13] Analytic expressions for transfer matrix coefficients for a symmetry plane of a transverse isotropic system immersed in fluid were given in Ref. 14 and for a plane of symmetry of a layered orthotropic medium between solids in Ref. 15. All the above-mentioned results have the same type of instability as for the isotropic case. Castaings and Hosten[16,17] successfully generalized the delta matrix method for multilayered media with monoclinic layers (orthotropic layers arbitrarily rotated about one symmetry axis) and overcame the instability problem. However, the need for several third-order delta matrix operators of 20×20 size with analytically derived elements makes realization of the method difficult. It was shown by Potel and Belleval[18] that for a periodic anisotropic system, a Floquet wave formulation can reduce the computational instability.

A different approach was developed by Kennett[19,20] and Kennett and Kery[21] using the wave reverberation method for each layer (sometimes called the reflection matrix method or the invariant embedding method). This method is widely used in seismology for wave propagation in layered media. It is based on building reflection and transmission matrices for individual interfaces and combining them by a recursive scheme into global reflection and transmission matrices for the layered system. It is numerically stable and efficient. It was originally developed for isotropic cases and its extension to wave propagation in the plane of symmetry of orthotropic layered systems was done by Booth and Crampin[22,23] and for generally anisotropic layered media by Fryer and Frazer.[24,25] Further applications of this method to geophysical problems of elastic wave propagation in layered anisotropic media were described in Refs. 26 and 27. The reflection/transmission matrices for a generally anisotropic layer between anisotropic solids have also been obtained by this method by Rokhlin and Huang.[28]

An alternative approach to obtaining a computationally stable solution is the global matrix method. Schmidt *et al.*[29,30] demonstrated that this method, if properly implemented, is unconditionally computationally stable. Due to the band structure of the global matrix its computational intensity is proportional to the number of layers N and it is faster than the delta matrix algorithm. This method is

[a]Author to whom correspondence should be addressed; electronic mail: rokhlin.2@osu.edu

implemented[29] in a way similar to the implementation of large-scale finite element programs and is usable for large number of layers N. The global matrix method for anisotropic multilayered media has been developed by Mal.[31] A reformulation of the global matrix method for isotropic cases using stiffness matrices was done by Kausel and Roesset.[32] As in the global matrix method, the problem reduces to numerical solution of a global banded system of equations. It is different from the global matrix method in that the interface total displacements and stresses are found instead of the wave amplitudes. This method was applied by Wang and Rajapakse[33] for a plane of symmetry in orthotropic media and by Wang and Rokhlin[34,35] to the monoclinic cross-ply composites. The main effort in the development of the method was expended on the thin layer approximation (the so-called ''discrete formulation'') of the stiffness matrix[36,37] by using the finite element representation in the vertical direction.

Solutions of acoustic and geophysical problems are usually obtained by both time and space domain numerical integral transform techniques. Therefore computationally efficient and stable algorithms are required to solve plane wave harmonic problems as functions of frequency and angle to obtain the kernel for the integral transform. All the above-mentioned methods give exact solutions for wave propagation in layered media. Each has advantages and disadvantages and different degrees of computational stability and efficiency. For a recent review of the different methods for isotropic case refer to Jensen *et al.*[38]

In this paper, we developed the recursive stiffness matrix algorithm for wave propagation in an arbitrarily anisotropic layered structure bounded by liquid or solid anisotropic media. Instead of building the total stiffness matrix as a global banded matrix, we utilize a recursive algorithm, somewhat similar to the reflection/transmission matrix method of Kennet[20] and Fryer and Frazer.[24] Our method is different in that we operate with total stresses and displacements on the system interfaces instead of reflection/transmission matrices. For this reason it is more convenient for many ultrasonic problems, especially composites, where often imperfect interfaces must be considered.[28] A stable matrix formulation based directly on the interfacial displacement/stress vectors makes it easier to incorporate imperfect interfaces using an asymptotic or spring interfacial stiffness matrix. Also different boundary value problems are easily addressed due to the global stiffness matrix formulation in the form of interfacial stresses and displacements. Our method's computational efficiency is the same as that of the standard transfer matrix method (computation time proportional to N). Moreover, the method's stability is as unconditional as that for the optimized global matrix method[29] and the reflection matrix method.[20,24] In particular, we show analytically and numerically that for a very thick layered structure, or a thick layer, the solution approaches that for a semispace. The method is cast in a concise matrix form and is simple to program and implement. To find for each layer the wave and polarization vectors, we utilize the algorithm described in Ref. 6 for arbitrary anisotropic layers which is applicable to an arbitrarily oriented anisotropic layer or to planes of symmetry, which is very convenient in many applications.

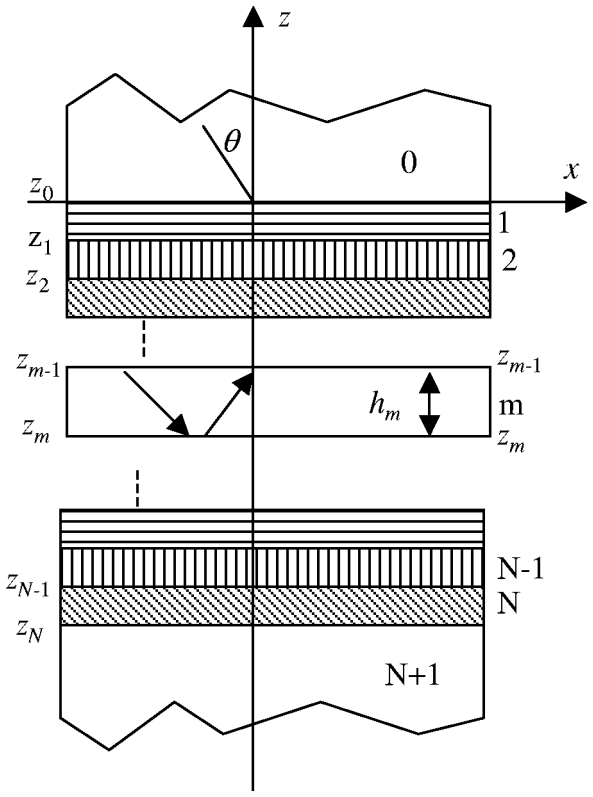


FIG. 1. A multilayered structure and coordinate system.

We have also adapted our algorithm to periodic anisotropic structures. This is relevant for structural composites since most fiber-reinforced composites include several repetitions of an identical unit element (cell) and are symmetric about the centerline of the laminate. Composite laminates are constructed by stacking a number of laminas in the thickness direction and most, such as cross-ply and quasi-isotropic composites, include a repetitive layout of a unidirectional orthotropic lamina, and are considered as an anisotropic periodic medium. This application has recently stimulated much investigation of wave propagation in anisotropic periodic solids.[39–43]

II. STIFFNESS MATRIX METHOD

A. Stiffness matrix for layer in anisotropic layered system

In this section we will formulate the stiffness matrix solution for generally anisotropic layers in a form suitable for multilayered system analysis. Let us consider a multilayered plate, consisting of N arbitrarily anisotropic layers as illustrated in Fig. 1. In the mth layer, the displacement vector $\mathbf{u}^m$ may be written as the summation of six partial waves

$$\mathbf{u}^m = \sum_{j=1}^{3} \left(a_j^+ \mathbf{p}_j^+ e^{ik_z^{+j}(z-z_m)} + a_j^- \mathbf{p}_j^- e^{ik_z^{-j}(z-z_{m-1})}\right)_m \times e^{i(k_x x + k_y y - \omega t)}, \quad (1)$$

where $\mathbf{u}^m=(u_x^m,u_y^m,u_z^m)^T$, T represents transpose. j $(=1,2,3)$ denotes jth partial wave. The positive and negative superscripts represent wave propagation in $+z$ or $-z$ directions, respectively; $\mathbf{p}_j^+$ and $\mathbf{p}_j^-(\mathbf{p}_j^\pm=(p_x^\pm,p_y^\pm,p_z^\pm)_j^T)$ are the unit displacement polarization vectors corresponding to waves with $(k_z^+)_j$ and $(k_z^-)_j$ wave vectors, respectively. The coordinate system is selected so that the x,z plane coincides with the incident plane[6] and thus $k_y=0$ in Eq. (1).

For each layer, we select the local coordinate origin at the top of the mth layer $(z=z_{m-1})$ for waves propagating along the $-z$ direction (a^-) and at the bottom of the mth layer $(z=z_m)$ for waves propagating along the $+z$ direction (a^+). The sorting of wave numbers and identification of the wave propagation direction is done from the energy flow (group velocity) direction. Such a selection of the local coordinate system is very important for eliminating the numerical overflow of the exponential terms in Eq. (1) when the waves become nonhomogeneous. This selection of coordinates assures that the exponential terms are normalized and that the nonhomogeneous exponentials equal unity on the interface, where they are initiated, and decay toward the opposite surface of the layer. Such coordinate selection has proven to be very useful for improving computational results[8,29,30] and differs from that used in most previous studies. However, appropriate coordinate selection is not sufficient to eliminate the well-known computational instability for thicker layers and higher frequencies.[29,30] As we will discuss in Secs. II C and II D, the method proposed here operates with matrices with dominating diagonal terms assuring numerical stability.

The displacement polarization vectors $\mathbf{p}_j^\pm$ and wave numbers $k_z^{\pm j}$ are determined by solving the Christoffel equation (2) (see Ref. 6) and applying Snell's law $k_x^m=k_x^0$ where $m=1,...,N$ and k_x^0 is the x projection of the wave number for the incident wave, as briefly outlined in Appendix A,

$$(c_{ijln}k_jk_n-\rho\omega^2\delta_{il})p_l=0, \tag{2}$$

where c_{ijln} represents the layer elastic constants and ρ density.

Equation (1) for the displacement vector on the upper $(z=z_{m-1})\mathbf{u}_{m-1}$ and lower surfaces $(z=z_m)\mathbf{u}_m$ of the layer m can be represented in the matrix form

$$\begin{bmatrix}\mathbf{u}_{m-1}\\ \mathbf{u}_m\end{bmatrix}_m=\begin{bmatrix}\mathbf{P}^- & \mathbf{P}^+\mathbf{H}^+\\ \mathbf{P}^-\mathbf{H}^- & \mathbf{P}^+\end{bmatrix}_m\begin{bmatrix}\mathbf{A}_m^-\\ \mathbf{A}_m^+\end{bmatrix}=\mathbf{E}_m^u\mathbf{A}_m, \tag{3}$$

where $\mathbf{P}^\pm(3\times3)=[\mathbf{p}_1^\pm,\mathbf{p}_2^\pm,\mathbf{p}_3^\pm]$, $\mathbf{A}_m^\pm=[a_1^\pm,a_2^\pm,a_3^\pm]^T$, and $\mathbf{H}^+(3\times3)=\mathrm{Diag}[e^{ik_z^{+1}h_m},e^{ik_z^{+2}h_m},e^{ik_z^{+3}h_m}]$, $\mathbf{H}^-(3\times3)=\mathrm{Diag}[e^{-ik_z^{-1}h_m},e^{-ik_z^{-2}h_m},e^{-ik_z^{-3}h_m}]$, $h_m=z_{m-1}-z_m$ being the thickness of the mth lamina. Due to the above-described local coordinate selection, the elements in both $\mathbf{H}^+$ and $\mathbf{H}^-$ will decay for complex wave numbers as the thickness increases. If the z axis is a symmetry axis, we have $k_z^{-j}=-k_z^{+j}$, therefore, $\mathbf{H}^+=\mathbf{H}^-$. This will be the case of monoclinic symmetry corresponding to arbitrary rotation of the orthotropic layer around the symmetry axis z.

The stress component vector $\boldsymbol{\sigma}=(\sigma_{31},\sigma_{32},\sigma_{33})^T$ on the $x-y$ plane parallel to the layer surface can be related to each of the plane wave displacement fields using Hooke's law,

$$\boldsymbol{\sigma}_m=\sum_{j=1}^{3}(a_j^+\mathbf{d}_j^+e^{ik_z^{+j}(z-z_m)}+a_j^-\mathbf{d}_j^-e^{ik_z^{-j}(z-z_{m-1})})_me^{i(k_xx+k_yy-\omega t)}, \tag{4}$$

where the components $(d_i^\pm)_j$ of the vector $\mathbf{d}_j^\pm$ are related to the polarization vector $\mathbf{p}_j^\pm$ by $(d_i^\pm)_j=(c_{i3ln}k_np_l^\pm)_j$. The stresses on the top $\boldsymbol{\sigma}_{m-1}(z=z_{m-1})$ and bottom surfaces $\boldsymbol{\sigma}_m(z=z_m)$ of the mth layer are related to the wave amplitudes $a_j^\pm$ in the matrix form as

$$\begin{bmatrix}\boldsymbol{\sigma}_{m-1}\\ \boldsymbol{\sigma}_m\end{bmatrix}_m=\begin{bmatrix}\mathbf{D}^- & \mathbf{D}^+\mathbf{H}^+\\ \mathbf{D}^-\mathbf{H}^- & \mathbf{D}^+\end{bmatrix}_m\begin{bmatrix}\mathbf{A}_m^-\\ \mathbf{A}_m^+\end{bmatrix}=\mathbf{E}_m^\sigma\mathbf{A}_m, \tag{5}$$

where $\mathbf{D}^\pm=[\mathbf{d}_1^\pm,\mathbf{d}_2^\pm,\mathbf{d}_3^\pm]$. Equations (3) and (5) relate the displacements and stresses on the layer surfaces to the wave displacement amplitudes $\mathbf{A}_m$. Substituting into Eq. (5) the amplitude vector $\mathbf{A}_m$ from Eq. (3), we obtain

$$\begin{bmatrix}\boldsymbol{\sigma}_{m-1}\\ \boldsymbol{\sigma}_m\end{bmatrix}_m=\mathbf{E}_m^\sigma\mathbf{E}_m^{u^{-1}}\begin{bmatrix}\mathbf{u}_{m-1}\\ \mathbf{u}_m\end{bmatrix}_m. \tag{6}$$

Equation (6) defines the layer stiffness matrix

$$\mathbf{K}_m(6\times6)=\mathbf{E}_m^\sigma\mathbf{E}_m^{u^{-1}}=\begin{bmatrix}\mathbf{K}_m^{11} & \mathbf{K}_m^{12}\\ \mathbf{K}_m^{21} & \mathbf{K}_m^{22}\end{bmatrix}. \tag{7}$$

One can also write Eq. (6) in the form

$$\begin{bmatrix}\mathbf{u}_{m-1}\\ \mathbf{u}_m\end{bmatrix}_m=\mathbf{E}_m^u\mathbf{E}_m^{\sigma^{-1}}\begin{bmatrix}\boldsymbol{\sigma}_{m-1}\\ \boldsymbol{\sigma}_m\end{bmatrix}_m. \tag{8}$$

Equation (8) defines the layer compliance matrix

$$\mathbf{S}_m(6\times6)=\mathbf{E}_m^u\mathbf{E}_m^{\sigma^{-1}}=\begin{bmatrix}\mathbf{S}_m^{11} & \mathbf{S}_m^{12}\\ \mathbf{S}_m^{21} & \mathbf{S}_m^{22}\end{bmatrix}. \tag{9}$$

Let us consider the stiffness matrix for a lower half space with the coordinate origin at the surface. In this case, only three waves $(\mathbf{A}^-)$ propagate in the $-z$ direction from the surface to infinity. The displacements and stresses can be obtained from Eqs. (3) and (5), respectively: $\mathbf{u}_{m-1}=\mathbf{P}^-\mathbf{A}^-$ and $\boldsymbol{\sigma}_{m-1}=\mathbf{D}^-\mathbf{A}^-$. Therefore the stiffness matrix $\mathbf{K}_h$ for the half space is given by

$$\boldsymbol{\sigma}_{m-1}=\mathbf{D}^-(\mathbf{P}^-)^{-1}\mathbf{u}_{m-1}=\mathbf{K}_h\mathbf{u}_{m-1}, \tag{10}$$

where $\mathbf{K}_h$ is the (3×3) stiffness matrix.

B. Relation to transfer matrix and asymptotic stability

Let us reorganize Eqs. (3) and (5), representing the displacements and stress on the top and bottom surfaces of the mth layer as

$$\begin{bmatrix}\mathbf{u}_{m-1}\\ \boldsymbol{\sigma}_{m-1}\end{bmatrix}_m=\begin{bmatrix}\mathbf{P}^- & \mathbf{P}^+\mathbf{H}^+\\ \mathbf{D}^- & \mathbf{D}^+\mathbf{H}^+\end{bmatrix}_m\begin{bmatrix}\mathbf{A}_m^-\\ \mathbf{A}_m^+\end{bmatrix}=\mathbf{W}_m^-\begin{bmatrix}\mathbf{A}_m^-\\ \mathbf{A}_m^+\end{bmatrix}, \tag{11}$$

$$\begin{bmatrix}\mathbf{u}_m\\ \boldsymbol{\sigma}_m\end{bmatrix}_m=\begin{bmatrix}\mathbf{P}^-\mathbf{H}^- & \mathbf{P}^+\\ \mathbf{D}^-\mathbf{H}^- & \mathbf{D}^+\end{bmatrix}_m\begin{bmatrix}\mathbf{A}_m^-\\ \mathbf{A}_m^+\end{bmatrix}=\mathbf{W}_m^+\begin{bmatrix}\mathbf{A}_m^-\\ \mathbf{A}_m^+\end{bmatrix}. \tag{12}$$

Substituting into Eq. (12) the amplitude vector $\mathbf{A}_m$ from Eq. (9), we obtain the transfer matrix $\mathbf{B}_m$,

$$\begin{bmatrix}\mathbf{u}_m\\ \boldsymbol{\sigma}_m\end{bmatrix}_m=\mathbf{W}_m^{+}(\mathbf{W}_m^{-})^{-1}\begin{bmatrix}\mathbf{u}_{m-1}\\ \boldsymbol{\sigma}_{m-1}\end{bmatrix}_m=\mathbf{B}_m\begin{bmatrix}\mathbf{u}_{m-1}\\ \boldsymbol{\sigma}_{m-1}\end{bmatrix}_m. \tag{13}$$

The relation between transfer matrix $\mathbf{B}_m$ and stiffness matrix $\mathbf{K}_m$ is given by

$$\mathbf{B}_m(6\times 6)=\begin{bmatrix}-(\mathbf{K}_m^{12})^{-1}\mathbf{K}_m^{11} & (\mathbf{K}_m^{12})^{-1}\\ \mathbf{K}_m^{21}-\mathbf{K}_m^{22}(\mathbf{K}_m^{12})^{-1}\mathbf{K}_m^{11} & \mathbf{K}_m^{22}(\mathbf{K}_m^{12})^{-1}\end{bmatrix}. \tag{14}$$

Thus the layer transfer $\mathbf{B}_m$ and stiffness $\mathbf{K}_m$ matrices are formed from the same matrix elements [Eqs. (3) and (5)]. The stiffness matrix can also be represented through the transfer matrix elements

$$\mathbf{K}_m(6\times 6)=\begin{bmatrix}-(\mathbf{B}_m^{12})^{-1}\mathbf{B}_m^{11} & (\mathbf{B}_m^{12})^{-1}\\ \mathbf{B}_m^{21}-\mathbf{B}_m^{22}(\mathbf{B}_m^{12})^{-1}\mathbf{B}_m^{11} & \mathbf{B}_m^{22}(\mathbf{B}_m^{12})^{-1}\end{bmatrix}. \tag{15}$$

The layer stiffness matrix cannot be calculated directly using Eq. (15), because it leads to computational instabilities (see the following). To obtain the layer stiffness matrix, computations have to be based on Eq. (6). Conversely, $\mathbf{B}_m$ computation using Eq. (14) is also unstable.

Relations (14) and (15) between the transfer matrix ($\mathbf{B}_m$) and the stiffness matrix ($\mathbf{K}_m$) are similar to the relations between the wave propagator ($\mathbf{Q}_m$) and the reflection matrices ($\mathbf{R}_m$) given by Fryer and Frazer[24] for the invariant embedding method (reflection method). However the physical meaning of the matrices involved is very different: one pair ($\mathbf{B}$, $\mathbf{K}$) relates displacements and stresses at opposite surfaces of a layer, the other pair ($\mathbf{Q}$, $\mathbf{R}$) amplitudes on the opposite sides of an interface. The transfer matrix $\mathbf{B}_m$ is defined as a matrix relating the stresses and displacements at a layer top surface to its bottom surface, while the stiffness matrix ($\mathbf{K}_m$) relates the stresses at a layer top and bottom surfaces to the displacements. The propagator ($\mathbf{Q}_m$) is defined as the matrix relating the amplitudes of waves propagating upward from the interface to those propagating downward, while the reflection matrix ($\mathbf{R}_m$) is the matrix of reflection and transmission coefficients defined on the interface.

At oblique incidence above the critical angle, the propagating waves become nonhomogeneous with complex wave numbers. If the thickness of the layer or the frequency increases to infinity, the elements of the matrix H_{ij} become zero. (We assume small attenuation in the system.) Substituting $H_{ij}=0$ into Eqs. (3) and (5), we obtain $\mathbf{K}_m^{21},\mathbf{K}_m^{12}=0$ and the stiffness matrix $\mathbf{K}_m$ and compliance matrix $\mathbf{S}_m$ become:

$$\mathbf{K}_m=\begin{bmatrix}\mathbf{D}^{+}(\mathbf{P}^{+})^{-1} & 0\\ 0 & \mathbf{D}^{-}(\mathbf{P}^{-})^{-1}\end{bmatrix},\qquad \mathbf{S}_m=\begin{bmatrix}\mathbf{P}^{+}(\mathbf{D}^{+})^{-1} & 0\\ 0 & \mathbf{P}^{-}(\mathbf{D}^{-})^{-1}\end{bmatrix}. \tag{16}$$

As can be seen, the stiffness matrix $\mathbf{K}_m$ degrades to the stiffness matrix for two half spaces [Eq. (10)] and its determinant $|\mathbf{K}_m|$ does not approach zero, i.e., $\mathbf{K}_m$ is not a singular matrix as the thickness and frequency tend to infinity. Thus the solution for the thick layer asymptotically decomposes to the two semispace solutions for the two surfaces of the layer.

The situation is different for the transfer matrix, because $H_{ij}=0$, the matrix $\mathbf{W}_m^{-}(6\times 6)=\left[\begin{smallmatrix}\mathbf{P}^{-} & 0\\ \mathbf{D}^{-} & 0\end{smallmatrix}\right]$ becomes singular ($|\mathbf{W}_m^{-}|=0$, and the inverse matrix $(\mathbf{W}_m^{-})^{-1}$ does not exist). This explains why the computation of the transfer matrix $\mathbf{B}_m$ is unstable with increased layer thickness or frequency.

The recursive method for multilayered media described here is based on the stiffness (compliance) matrix and as we will show is computationally stable.

C. Imperfect boundary conditions

If the layer thickness approaches zero, one can obtain the first- and second-order asymptotic solutions for the layer stiffness matrix from the transfer matrix $\mathbf{B}_m$.[28,44] If one neglects the coupling terms and includes only spring and inertia terms, the transfer matrix $\mathbf{B}_m$ becomes[28,44]

$$\mathbf{B}_m^{\mathrm{imp}}=\begin{bmatrix}1 & 0 & 0 & 1/K_{t1} & 0 & 0\\ 0 & 1 & 0 & 0 & 1/K_{t2} & 0\\ 0 & 0 & 1 & 0 & 0 & 1/K_n\\ -\omega^2M_{p1} & 0 & 0 & 1 & 0 & 0\\ 0 & -\omega^2M_{p2} & 0 & 0 & 1 & 0\\ 0 & 0 & -\omega^2M_n & 0 & 0 & 1\end{bmatrix}. \tag{17}$$

From Eq. (17), we obtain the stiffness matrix to describe imperfect conditions at the interface:

$$\mathbf{K}_m^{\mathrm{imp}}=\begin{bmatrix}K_{t1} & 0 & 0 & -\omega^2M_{p1}-K_{t1} & 0 & 0\\ 0 & K_{t2} & 0 & 0 & -\omega^2M_{p2}-K_{t2} & 0\\ 0 & 0 & K_n & 0 & 0 & -\omega^2M_n-K_n\\ K_{t1} & 0 & 0 & -K_{t1} & 0 & 0\\ 0 & K_{t2} & 0 & 0 & -K_{t2} & 0\\ 0 & 0 & K_n & 0 & 0 & -K_n\end{bmatrix}, \tag{18}$$

where the equations for K_i and M_i are given in terms of effective elastic constants of a thin imperfect interface layer.[44] If we neglect the inertial terms (M_i), we obtain a pure spring model for a thin layer. The stiffness matrix in this form is useful in our framework to describe imperfect boundary conditions between anisotropic layers and is naturally included in the recursive algorithm described in the following.

D. Recursive algorithm for computation of total stiffness (compliance) of multilayered media

To obtain the global stiffness matrix for a multilayered structure, we have developed a recursive algorithm based on the stiffness matrix for a single layer. We assume the interface between each layer in the multilayered structure is perfect (continuity of displacement and stress) or we will utilize the stiffness matrix [Eq. (18)] for imperfect contact. We can utilize the stiffness matrix for each layer to calculate the global stiffness matrix. First let us consider two neighboring bonded layers. The first layer stiffness matrix $\mathbf{K}^1=\mathbf{K}^A$ is given by

$$\begin{bmatrix}\boldsymbol{\sigma}_1\\ \boldsymbol{\sigma}_2\end{bmatrix}=\begin{bmatrix}\mathbf{K}_{11}^A & \mathbf{K}_{12}^A\\ \mathbf{K}_{21}^A & \mathbf{K}_{22}^A\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{bmatrix}, \tag{19}$$

and the second cell stiffness matrix $\mathbf{K}^2=\mathbf{K}^B$ is given by

$$\begin{bmatrix}\boldsymbol{\sigma}_2\\ \boldsymbol{\sigma}_3\end{bmatrix}=\begin{bmatrix}\mathbf{K}_{11}^B & \mathbf{K}_{12}^B\\ \mathbf{K}_{21}^B & \mathbf{K}_{22}^B\end{bmatrix}\begin{bmatrix}\mathbf{u}_2\\ \mathbf{u}_3\end{bmatrix}. \tag{20}$$

Now we relate $\boldsymbol{\sigma}_1$ and $\mathbf{u}_1$ to $\boldsymbol{\sigma}_3$ and $\mathbf{u}_3$ by excluding $\boldsymbol{\sigma}_2$ and $\mathbf{u}_2$ from Eqs. (19) and (20) and combining the stiffness matrices for these two bonded layers:

$$\begin{bmatrix}\boldsymbol{\sigma}_1\\ \boldsymbol{\sigma}_3\end{bmatrix}=\begin{bmatrix}\mathbf{K}_{11}^A+\mathbf{K}_{12}^A(\mathbf{K}_{11}^B-\mathbf{K}_{22}^A)^{-1}\mathbf{K}_{21}^A & -\mathbf{K}_{12}^A(\mathbf{K}_{11}^B-\mathbf{K}_{22}^A)^{-1}\mathbf{K}_{12}^B\\ \mathbf{K}_{21}^B(\mathbf{K}_{11}^B-\mathbf{K}_{22}^A)^{-1}\mathbf{K}_{21}^A & \mathbf{K}_{22}^B-\mathbf{K}_{21}^B(\mathbf{K}_{11}^B-\mathbf{K}_{22}^A)^{-1}\mathbf{K}_{12}^B\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_3\end{bmatrix}. \tag{21}$$

Now calling the matrix obtained $\mathbf{K}^A$ and the stiffness matrix for the next layer $\mathbf{K}^3=\mathbf{K}^B$, we can recursively use Eq. (21) to obtain the global stiffness matrix which relates the stresses to the displacements for the top and bottom surfaces of the whole structure. The algorithm can be described as

$$\mathbf{K}^M=\begin{bmatrix}\mathbf{K}_{11}^{M-1}+\mathbf{K}_{12}^{M-1}(\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1})^{-1}\mathbf{K}_{21}^{M-1} & -\mathbf{K}_{12}^{M-1}(\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1})^{-1}\mathbf{K}_{12}^m\\ \mathbf{K}_{21}^m(\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1})^{-1}\mathbf{K}_{21}^{M-1} & \mathbf{K}_{22}^m-\mathbf{K}_{21}^m(\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1})^{-1}\mathbf{K}_{12}^m\end{bmatrix}, \tag{22}$$

where $\mathbf{K}^M$ is the total stiffness matrix for the top m layers, $\mathbf{K}_{ij}^{M-1}$ is the total stiffness matrix for the top $m-1$ layers, $\mathbf{K}_{ij}^m$ are stiffness matrix elements for the mth layer. The submatrix difference $\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1}$ is a regular matrix [it is not null even for identical layers A and B in Eq. (21)]. The submatrices $\mathbf{K}_{11}^m$ and $\mathbf{K}_{22}^m$ relate stress and displacement on the opposite interfaces and their diagonal elements have opposite sign, thus when the stiffness matrices m and $M-1$ have similar properties, the diagonal elements in the term $\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1}$ are doubled in value. To clarify the physical meaning of this difference it is useful to consider the asymptotic properties for thin layers given in Eq. (18). For example the (1,1) element of the difference $(K_{t1})_A+(K_{t1})_B$ represents the effective longitudinal stiffness of the coupled A and B layers (the matrix inverse is the matrix of effective compliances). Thus the diagonal elements of the matrix $\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1}$ are dominated by the total stiffness of the layered system and the matrix is regular.

When thickness or frequency increases, the off-diagonal elements of the stiffness matrix tend to zero while the diagonal elements remain finite and the recursive algorithm (22) remains stable since it involves only the inverse matrix of the diagonal elements in the term $(\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1})^{-1}$. When the total layered system thickness tends to infinity for complex wave numbers (nonhomogeneous waves and propagating waves in media with attenuation), we can obtain

$$\begin{bmatrix}\boldsymbol{\sigma}_1\\ \boldsymbol{\sigma}_{N+1}\end{bmatrix}=\begin{bmatrix}\mathbf{K}_{11}^1 & 0\\ 0 & \mathbf{K}_{22}^N\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_{N+1}\end{bmatrix}. \tag{23}$$

Thus, the combined stiffness matrix decomposes into stiffness matrices for the top "1" and bottom "N" half spaces, maintaining its regularity.

Note that in Eq. (22), the submatrix $\mathbf{K}_{22}^M$ of the total stiffness matrix $\mathbf{K}^M$ for the top m layers depends only on submatrix $\mathbf{K}_{22}^{M-1}$ of the total stiffness $\mathbf{K}^{M-1}$ for the top $m-1$ layers. Therefore $\mathbf{K}_{22}^M$ can be obtained recursively without calculating all other submatrices of the total stiffness matrix. This is useful when one needs to calculate the total stiffness matrix for a layered half space $\mathbf{K}_S^M(\mathbf{K}_S^M=\mathbf{K}_{22}^M)$.

For a single layer, the computational intensity of the transfer matrix $\mathbf{B}_m$ [Eq. (13)] and stiffness ($\mathbf{K}_m$) or compliance ($\mathbf{S}_m$) matrices [Eq. (7) or Eq. (9)] is equivalent. For a multilayered structure, the global transfer matrix $\mathbf{B}$ for the multilayered structure is

$$\mathbf{B}=\mathbf{B}_1\mathbf{B}_2\mathbf{B}_3\cdots\mathbf{B}_N, \tag{24}$$

and as we will show in Sec. III, the computation time for both methods is equivalent and proportional to the number of layers N in the system. However, as we have discussed previously, the transfer matrix method is computationally unstable even for moderate layer thickness and frequency. Its stability will not be improved if one uses Eq. (14) for computation of the transfer matrix due to instability of the inversion of the off-diagonal submatrices of the stiffness matrix.

E. Computation of reflection and transmission coefficients

Using the recursive algorithm $N-1$ times, we obtain the total stiffness matrix for the N-layered structure. We denote the total stiffness matrix for the whole layered structure $\mathbf{K}$,

$$\begin{bmatrix} \boldsymbol{\sigma}_0 \\ \boldsymbol{\sigma}_N \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_N \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_N \end{bmatrix}, \tag{25}$$

and its inverse, the global compliance matrix

$$\mathbf{S} = \mathbf{K}^{-1} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}. \tag{26}$$

Let us consider an acoustic wave incident from fluid onto an immersed layered structure. The surface stresses and displacements corresponding to waves in the fluid are given in Eqs. (B3) and (B4). Combining the global compliance matrix with the reflected and transmitted waves, one may write (Appendix B) the transmission coefficient for laminates immersed in fluid as

$$T = 2\Lambda S_{21}^{33} / [(S_{11}^{33} + \Lambda)(S_{22}^{33} - \Lambda) - S_{21}^{33} S_{12}^{33}], \tag{27}$$

and the reflection coefficient as

$$R = [(S_{11}^{33} - \Lambda)(S_{22}^{33} - \Lambda) - S_{21}^{33} S_{12}^{33}] / [(S_{11}^{33} + \Lambda)(S_{22}^{33} - \Lambda) - S_{21}^{33} S_{12}^{33}], \tag{28}$$

where S_{ij}^{33} are the (3,3) elements in the matrix $\mathbf{S}_{ij}(3\times 3)$ of the total matrix (26) and $\Lambda = \cos\theta/(i\omega\rho_f V_f)$, ρ_f is the fluid density, and V_f is the acoustic velocity in the fluid, θ is the incident angle.

For the case when the layered anisotropic structure is bounded by two generally anisotropic solids, we can use Eqs. (3) and (5) to calculate the stresses and displacements produced by incident, reflected, and transmitted waves at the top and bottom structure surfaces and combine them with the global stiffness matrix to obtain the reflection and transmission coefficients. Let us consider an incident field as a combination of plane waves in the top half space. Then from Eqs. (3) and (5) at the top surface we have

$$\begin{bmatrix} \mathbf{u}_0 \\ \boldsymbol{\sigma}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{P}_0^- \\ \mathbf{D}_0^- \end{bmatrix} \mathbf{A}_{\text{in}} + \begin{bmatrix} \mathbf{P}_0^+ \\ \mathbf{D}_0^+ \end{bmatrix} \mathbf{A}_r \tag{29}$$

and at the bottom surface, we have

$$\begin{bmatrix} \mathbf{u}_N \\ \boldsymbol{\sigma}_N \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{N+1}^- \\ \mathbf{D}_{N+1}^- \end{bmatrix} \mathbf{A}_t, \tag{30}$$

where the vectors $\mathbf{A}_{\text{in}}$, $\mathbf{A}_r$, and $\mathbf{A}_t$ contain the amplitudes for three bulk waves in the incident, reflected, and transmitted wave fields, respectively. Combining Eqs. (29) and (30) with the global stiffness matrix $\mathbf{K}$, we obtain the linear system to find the unknown reflection and transmission amplitudes:

$$\begin{bmatrix} -\mathbf{K}^{11}\mathbf{P}_0^+ + \mathbf{D}_0^+ & -\mathbf{K}^{12}\mathbf{P}_{N+1}^- \\ -\mathbf{K}^{21}\mathbf{P}_0^+ & -\mathbf{K}^{22}\mathbf{P}_{N+1}^- + \mathbf{D}_{N+1}^- \end{bmatrix} \begin{bmatrix} \mathbf{A}_r \\ \mathbf{A}_t \end{bmatrix} = \begin{bmatrix} \mathbf{K}^{11}\mathbf{P}_0^- - \mathbf{D}_0^- \\ \mathbf{K}^{21}\mathbf{P}_0^- \end{bmatrix} \mathbf{A}_{\text{in}}. \tag{31}$$

F. Wave amplitudes inside the structure

In Sec. II E, the reflection and transmission coefficients have been obtained. From Eqs. (29), (30) or (B3), (B4), one has the displacement and stress at the top and bottom surfaces of the layered structure considering contributions from incident, reflected, and transmitted waves. In this section we discuss a backrecursive computation of displacements and stresses inside the structure from the surface displacements and stresses.

One may be tempted to calculate the wave amplitudes for each layer directly from Eqs. (3) and (5) computing the amplitude for the top layer:

$$\begin{bmatrix} \mathbf{A}_1^- \\ \mathbf{A}_1^+ \end{bmatrix} = \begin{bmatrix} \mathbf{P}^- & \mathbf{P}^+\mathbf{H}^+ \\ \mathbf{D}^- & \mathbf{D}^+\mathbf{H}^+ \end{bmatrix}_1^{-1} \begin{bmatrix} \mathbf{u}_0 \\ \boldsymbol{\sigma}_0 \end{bmatrix} \tag{32}$$

and next using these to obtain the displacements and stresses at the top surface of the second layer and calculating the wave amplitudes in transfer matrix style in the following layers. However, such a computation will not be stable for large thickness and high frequency.

For this reason, we utilize the ''backpropagation'' recursive algorithm to calculate the internal field, similar to the recursive algorithm described in Sec. II D to calculate the total stiffness matrix. Let us assume that we have calculated the total stiffness matrix for the top $m-1$ layers ($\mathbf{K}^{M-1}$) [Eq. (22)] and also have the stiffness matrix $\mathbf{K}^m$ for the mth layer [Eq. (7)]. These stiffness matrices provide the relationship

$$\begin{bmatrix} \boldsymbol{\sigma}_0 \\ \boldsymbol{\sigma}_m \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{11}^{M-1} & \mathbf{K}_{12}^{M-1} \\ \mathbf{K}_{21}^{M-1} & \mathbf{K}_{22}^{M-1} \end{bmatrix} \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_m \end{bmatrix} \tag{33}$$

and

$$\begin{bmatrix} \boldsymbol{\sigma}_m \\ \boldsymbol{\sigma}_{m+1} \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{11}^{m} & \mathbf{K}_{12}^{m} \\ \mathbf{K}_{21}^{m} & \mathbf{K}_{22}^{m} \end{bmatrix} \begin{bmatrix} \mathbf{u}_m \\ \mathbf{u}_{m+1} \end{bmatrix}. \tag{34}$$

Combining Eqs. (33) and (34), one finds

$$\mathbf{u}_m = (\mathbf{K}_{11}^m - \mathbf{K}_{22}^{M-1})^{-1}\mathbf{K}_{21}^{M-1}\mathbf{u}_0 - (\mathbf{K}_{11}^m - \mathbf{K}_{22}^{M-1})^{-1}\mathbf{K}_{12}^m\mathbf{u}_{m+1} \tag{35}$$

and

$$\boldsymbol{\sigma}_m = \mathbf{K}_{21}^{M-1}\mathbf{u}_0 - \mathbf{K}_{22}^{M-1}\mathbf{u}_m. \tag{36}$$

Recursively using Eqs. (33)–(36) one can obtain the displacement and stress at all interfaces. First we calculate the displacements at the top $z_0(\mathbf{u}_0)$ and the bottom $z_{N+1}(\mathbf{u}_{N+1})$ surfaces of the whole structure utilizing incident, reflected, and transmitted wave amplitudes. Calculating the reflection and transmission coefficients we have already obtained the total stiffness matrices for the top $N-1$ layers and the stiffness matrix for the Nth bottom layer; then, using Eq. (35), the displacement $\mathbf{u}_{N-1}$ at the bottom of $N-1$ layer can be obtained. Next we consider the top $N-1$ layers, and using the same procedure, we obtain $\mathbf{u}_{N-2}$. Therefore, recursively, we can obtain the displacements at all interfaces. Because this algorithm is started from the system bottom and propagates the solution to the top while the algorithm for computation of the total stiffness matrix is from the top layer to the

bottom, we call it the ''backpropagation'' recursive algorithm.

The wave amplitudes in each layer are determined by

$$\mathbf{A}_m^- = (\mathbf{P}^-)^{-1}\mathbf{u}_{m-1}, \quad \mathbf{A}_m^+ = (\mathbf{P}^+)^{-1}\mathbf{u}_m. \tag{37}$$

If one stores the stiffness matrices for each layer, they need not be calculated again and the amplitude calculation only involves matrix multiplication and inversion recursively using Eq. (35).

G. Lamb and surface waves

The characteristic equation for Lamb modes can be determined easily from the total stiffness matrix of the structure (25). If the top and bottom surfaces are free, $\left[\begin{smallmatrix}\boldsymbol{\sigma}_0\\ \boldsymbol{\sigma}_N\end{smallmatrix}\right]=0$. Thus we have $\mathbf{K}\left[\begin{smallmatrix}\mathbf{u}_0\\ \mathbf{u}_N\end{smallmatrix}\right]=0$ and the Lamb wave dispersion equation is determined as

$$\det(\mathbf{K})=0. \tag{38}$$

For a fluid-loaded layered anisotropic structure, the dispersion equation for leaky Lamb modes can be determined from the reflection and transmission coefficients [Eqs. (27) and (28)], by equating the denominator to zero:

$$(S_{11}^{33}+\Lambda)(S_{22}^{33}-\Lambda)-S_{21}^{33}S_{12}^{33}=0. \tag{39}$$

Because $\mathbf{S}=\mathbf{K}^{-1}$, each element of $\mathbf{S}$ can be written as

$$S_{ij}=(-1)^{i+j}\frac{\det(\mathbf{K}_{ji})}{\det(\mathbf{K})},$$

where $\mathbf{K}_{ji}$ is the matrix obtained by deleting row j and column i in $\mathbf{K}$. Upon substituting this equation into Eq. (37), the dispersion equation for leaky Lamb modes may be written as

$$\{\det(\mathbf{K}_{33})/\Lambda+\det(\mathbf{K})\}\{\det(\mathbf{K}_{66})/\Lambda-\det(\mathbf{K})\} - \det(\mathbf{K}_{63})\det(\mathbf{K}_{36})/\Lambda=0. \tag{40}$$

When the fluid density approaches zero, $1/\Lambda\to 0$, thus Eq. (40) becomes Eq. (38).

The characteristic equation for surface waves in a semi-infinite layered structure is obtained from Eq. (23):

$$\det(\mathbf{K}_{22}^N)=0, \tag{41}$$

where $\mathbf{K}_{22}^N$ can be either obtained from the total stiffness matrix $\mathbf{K}^N$ or recursively using the submatrix (2,2) in Eq. (22): $\mathbf{K}_S^M=\mathbf{K}_{22}^M=\mathbf{K}_{22}^m-\mathbf{K}_{21}^m(\mathbf{K}_{11}^m-\mathbf{K}_{22}^{M-1})^{-1}\mathbf{K}_{12}^m$ $(M,m=0,\ldots N)$ and taking $\mathbf{K}_{22}^0=\mathbf{D}_0^+(\mathbf{P}_0^+)^{-1}$ as the stiffness matrix of the top semispace; $\mathbf{K}_S^N$ is the layered semi-space stiffness matrix.

H. Application to periodic media: Stiffness matrix for a cell

For materials with structural periodicity, which is the case for composites, the recursive method allows very efficient computation of the total stiffness matrix. Let us introduce a repetitive cell p consisting of n different anisotropic layers; we define it as $[\theta_1/\theta_2/\cdots\theta_{n-1}/\theta_n]_{ps}$. Here θ_i indicates a layer in the cell p with rotation angle θ_i. For example, $[0°/45°/90°/-45°]$ represents a four layered cell with a combination of 0°, 45°, 90°, and −45° laminas. The index p indicates the number of cells in the layered structure and the index s indicates the structure symmetry. The traction and displacements at the top and bottom surfaces of the cell $[\theta_1/\theta_2/\cdots\theta_{n-1}/\theta_n]$ are denoted by $\boldsymbol{\sigma}^+$, $\boldsymbol{\sigma}^-$, $\mathbf{u}^+$, and $\mathbf{u}^-$. Let us define the total stiffness matrix for the cell as the matrix $\mathbf{K}_c$ which relates $\boldsymbol{\sigma}^+$, $\boldsymbol{\sigma}^-$ to $\mathbf{u}^+$, $\mathbf{u}^-$. Using the recursive algorithm (22), we can obtain the cell stiffness matrix $\mathbf{K}_c$.

In the case of two repetitive cells, the two cells are identical with identical stiffness matrices ($\mathbf{K}_{ij}^A=\mathbf{K}_{ij}^B$), so Eq. (21) becomes

$$\begin{bmatrix}\boldsymbol{\sigma}_1\\ \boldsymbol{\sigma}_3\end{bmatrix}=\begin{bmatrix}\mathbf{K}_{11}^A+\mathbf{K}_{12}^A(\mathbf{K}_{11}^A-\mathbf{K}_{22}^A)^{-1}\mathbf{K}_{21}^A & -\mathbf{K}_{12}^A(\mathbf{K}_{11}^A-\mathbf{K}_{22}^A)^{-1}\mathbf{K}_{12}^A\\ \mathbf{K}_{21}^A(\mathbf{K}_{11}^A-\mathbf{K}_{22}^A)^{-1}\mathbf{K}_{21}^A & \mathbf{K}_{22}^A-\mathbf{K}_{21}^A(\mathbf{K}_{11}^A-\mathbf{K}_{22}^A)^{-1}\mathbf{K}_{12}^A\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_3\end{bmatrix}. \tag{42}$$

To incorporate the next cell $\mathbf{K}_c$, one uses Eq. (22) to combine the matrix obtained by Eq. (42) with the additional already-known cell matrix $\mathbf{K}_c$. If the two cells are symmetrical about their bond line, then the two cell stiffness matrices have the relations

$$\mathbf{K}_{12}^A=-\mathbf{I}_2\mathbf{K}_{21}^B\mathbf{I}_2, \quad \mathbf{K}_{11}^A=-\mathbf{I}_2\mathbf{K}_{22}^B\mathbf{I}_2,$$

where

$$\mathbf{I}_2=\begin{bmatrix}1&0&0\\0&1&0\\0&0&-1\end{bmatrix}.$$

Therefore in the case of symmetry, Eq. (21) may be rewritten as

$$\begin{bmatrix}\boldsymbol{\sigma}_1\\ \boldsymbol{\sigma}_3\end{bmatrix}=\begin{bmatrix}\mathbf{K}_{11}^A-\mathbf{K}_{12}^A(\mathbf{I}_2\mathbf{K}_{22}^A\mathbf{I}_2+\mathbf{K}_{22}^A)^{-1}\mathbf{K}_{21}^A & \mathbf{K}_{12}^A(\mathbf{I}_2\mathbf{K}_{22}^A\mathbf{I}_2+\mathbf{K}_{22}^A)^{-1}\mathbf{I}_2\mathbf{K}_{21}^A\mathbf{I}_2\\ \mathbf{I}_2\mathbf{K}_{12}^A\mathbf{I}_2(\mathbf{I}_2\mathbf{K}_{22}^A\mathbf{I}_2+\mathbf{K}_{22}^A)^{-1}\mathbf{K}_{21}^A & \mathbf{I}_2(\mathbf{K}_{11}^A+\mathbf{K}_{12}^A(\mathbf{I}_2\mathbf{K}_{22}^A\mathbf{I}_2-\mathbf{K}_{22}^A)^{-1}\mathbf{K}_{21}^A)\mathbf{I}_2\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_3\end{bmatrix}. \tag{43}$$

For a complicated composite with multiple cell structures, one can calculate the global stiffness matrix using Eq. (22) where instead of a layer stiffness matrix one can use the cell stiffness matrix. Each different cell stiffness matrix can be obtained independently. For a symmetric composite defined as $[\theta_1/\theta_2/\cdots\theta_{n-1}/\theta_n]_{ps}$, for an arbitrary number of repetitions p, we can recursively apply p times the operation using Eqs. (22) and (43) to obtain the stiffness matrix for the entire laminate (total stiffness matrix). If p can be written as 2^i, then we can use Eq. (42) and only i operations are needed. The recursive computation is simple matrix multiplication and one inversion of the stiffness matrix $\mathbf{K}_{ij}$ and these are just 3 by 3 matrices. Therefore the algorithm is very efficient.

I. Infinite periodic media: Floquet waves

To facilitate further discussion let us consider an infinite periodic medium. The free waves propagating in a such medium are called Floquet waves. Due to periodicity the Floquet wave theory requires[12,40]

$$\begin{bmatrix} \mathbf{u}^+ \\ \boldsymbol{\sigma}^+ \end{bmatrix} = e^{i\zeta h} \begin{bmatrix} \mathbf{u}^- \\ \boldsymbol{\sigma}^- \end{bmatrix}, \tag{44}$$

where ζ represents the Floquet wave number and h is the thickness of the periodic unit cell. Consider a periodic medium consisting of the unit cell whose stiffness matrix is given by $\mathbf{K}_c$. Using the definition of the stiffness matrix, Eq. (44) may be written as

$$(e^{i\zeta h}\mathbf{K}_c^{21} - e^{-i\zeta h}\mathbf{K}_c^{12} + \mathbf{K}_c^{22} - \mathbf{K}_c^{11})\mathbf{u}^- = 0. \tag{45}$$

The determinant of the system of Eq. (42) gives the characteristic equation for the Floquet wave

$$\det(e^{i\zeta h}\mathbf{K}_c^{21} - e^{-i\zeta h}\mathbf{K}_c^{12} + \mathbf{K}_c^{22} - \mathbf{K}_c^{11}) = 0. \tag{46}$$

If the z axis is a material symmetry axis, Eq. (46) may be expanded as

$$A_3\cos(3\zeta h) + A_2\cos(2\zeta h) + A_1\cos(\zeta h) + A_0 = 0, \tag{47}$$

where $A_3 = |\mathbf{K}_c^{21}|$, $A_2 = (|\mathbf{E}+\mathbf{K}_c^{21}| + |\mathbf{E}-\mathbf{K}_c^{21}|)/2 - |\mathbf{E}|$,

$$A_1 = (|\mathbf{K}_c^{12}+\mathbf{K}_c^{21}| + |\mathbf{K}_c^{12}-\mathbf{K}_c^{21}| - |\mathbf{K}_c^{12}+\mathbf{E}| - |\mathbf{K}_c^{12}-\mathbf{E}|)/2 + |\mathbf{K}_c^{12}| - |\mathbf{K}_c^{21}|,$$

$$A_0 = 2|\mathbf{E}| + (|\mathbf{E}+\mathbf{K}_c^{12}-\mathbf{K}_c^{21}| + |\mathbf{E}+\mathbf{K}_c^{12}-\mathbf{K}_c^{21}| - |\mathbf{K}_c^{12} + \mathbf{E}| - |\mathbf{K}_c^{12}-\mathbf{E}| + |\mathbf{K}_c^{21}+\mathbf{E}| - |\mathbf{K}_c^{21}-\mathbf{E}|))/2$$

$\mathbf{E} = \mathbf{K}_c^{22} - \mathbf{K}_c^{11}$, $|\mathbf{E}|$ represents the determinant of matrix $\mathbf{E}$.

Equation (47) has six solutions for ζ. These solutions satisfy $\zeta_1 = -\zeta_2$, $\zeta_4 = -\zeta_3$, $\zeta_5 = -\zeta_6$. Using this Floquet wave theory, one can obtain reflection coefficients from periodic media, which is especially useful for periodic semispaces. Detailed discussion will be presented elsewhere.

III. COMPUTATIONAL EXAMPLES AND COMPARISON WITH EXPERIMENT

A. Computational efficiency

In this section, we will compare the computational efficiency of the stiffness matrix recursive algorithm and the transfer matrix method. In both methods, the total time for performing calculation for the layered system is the sum of two parts: (1) calculation of a single layer stiffness (transfer) matrix and (2) calculation of the total stiffness (transfer) matrix by combining the layer stiffness (transfer) matrices. The layer stiffness matrix is calculated by Eq. (6) and the layer transfer matrix is computed by Eq. (13). Due to the similarity of the equations the CPU time required to obtain the layer stiffness matrix and the transfer matrix are identical. Equation (24) is used to obtain the total transfer matrix, which involves multiplications of 6×6 matrices. A 6×6 matrix multiplication requires 6^3 multiplications. Therefore the number of multiplications required to obtain the total transfer matrix for N layers will be $(N-1)\times 6^3 = 216\times(N-1)$. The total stiffness matrix is obtained by the recursive algorithm (22). Each recursive operation requires six 3×3 matrix multiplications (the number of multiplications is $6\times 3^3 = 162$) and one inversion of a 3×3 matrix (the number of multiplications is 40). Thus the total number of multiplications in calculating the total stiffness matrix is $202\times(N-1)$. Therefore, both methods have the same computational efficiency with computation time proportional to the number of layers.

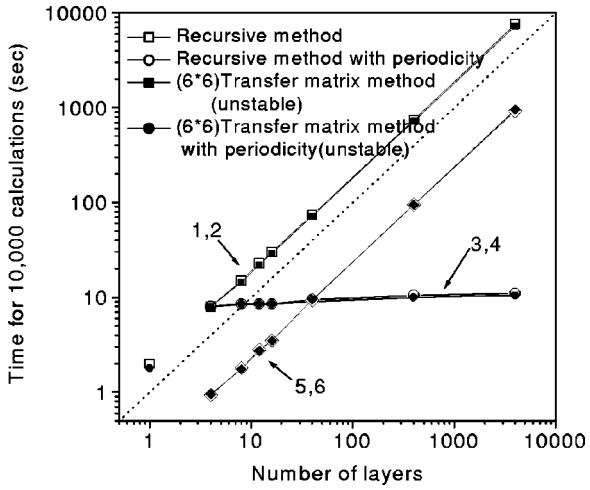


FIG. 2. Comparison of the computational efficiency of transfer matrix and stiffness matrix methods. It shows the CPU time required to perform 10 000 calculations using PIII600 computer vs number of layers in the system. The multilayered structure is a quasi-isotropic composite $[0°/45°/90°/-45°]_p$. Curves 1 and 2 without accounting for repetition; curves 3 and 4 with repetition; curves 5 and 6 indicate the time just for combining layer stiffness or transfer matrices using Eqs. (22) and (24).

Figure 2 shows the CPU time of the reflection coefficient calculation for 10 000 points on a Pentium III 600 CPU for $[0°/45°/90°/-45°]_p$ composites indicating equivalent computational time. Curves 1 and 2 in Fig. 2 are without considering the repetition in the layered structure. Curves 3

TABLE I. Properties of one lamina.

Elastic constants of one lamina (GPa)	
C_{11}	$143.2-0.32i$
C_{22}	$15.8-0.11i$
C_{12}	$7.5-0.031i$
C_{23}	$8.2-0.002i$
C_{55}	$7.0-0.07i$
Density (g/cm^3)	1.6
Thickness (mm)	0.194

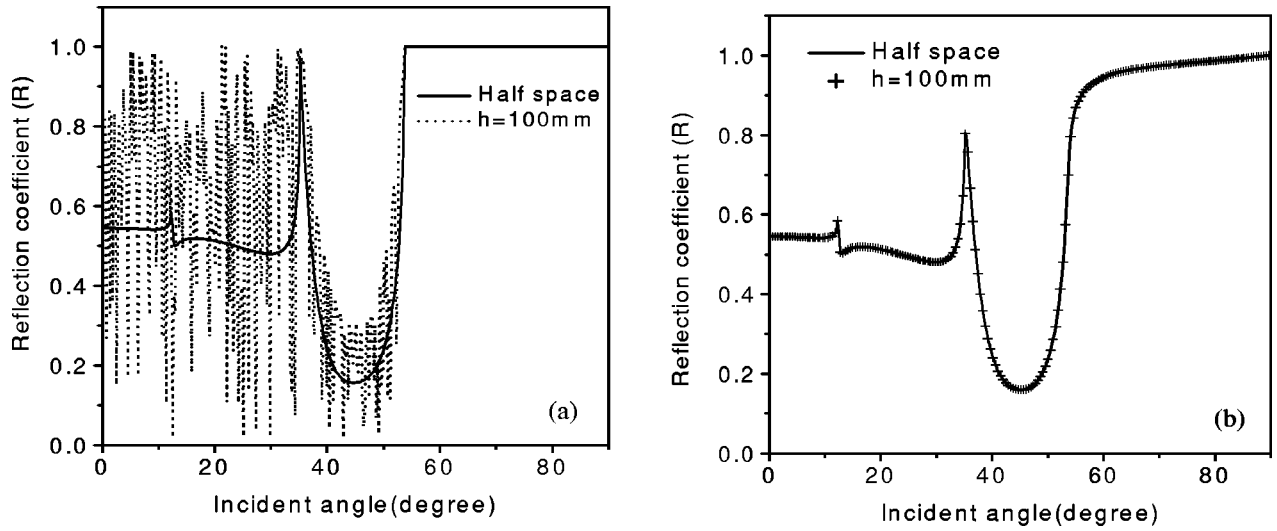


FIG. 3. Reflection coefficient for a unidirectional composite semispace (solid line) and for a unidirectional 100-mm-thick layer (··· or +). Fibers are oriented 45° to the incident plane. Wave is incident from water at frequency 10 MHz. (a) Without attenuation, (b) with attenuation.

and 4 are with considering the repetition. Curves 5 and 6 show the time just for the second part of the calculation, i.e, not considering the time required to calculate the single layer transfer or the stiffness matrices. By comparing curves 1, 2 with curves 5,6, one can see that the time used in the second part is only about 12% of the total time (this is due to anisotropy and the necessity to calculate wave numbers and polarization vectors for each layer).

B. Computational stability

In this section we will demonstrate numerically that the recursive stiffness matrix method is unconditionally stable. As examples we have utilized for computation several composite layouts. The properties of the lamina are given in Table I. Without limitation of generality, we consider a transverse isotropic lamina. Our computations of the reflection and transmission coefficients by different methods (transfer matrix, global matrix, delta matrix, and recursive stiffness matrix methods) for $[0°/90°]_{8s}$ laminate show that for the incident plane oriented at 0° and incident angle 50° the transfer matrix method becomes unstable at frequencies above 0.62 MHz. When the computation becomes unstable, the energy balance is not satisfied and the reflection and transmission coefficients vary randomly. For the same case the delta matrix method is stable with overflow error occurring at frequencies at about 14 MHz. The global matrix and recursive stiffness matrix methods are both stable. In the global matrix method, we use Gaussian elimination to solve the $6N\times 6N$ matrix equation as described in Ref. 38. We have never encountered computational instabilities for the recursive stiffness matrix method while using it intensively in our laboratory for several years.

In Sec. II, we theoretically demonstrate that the stiffness matrix solution approaches that of the semispace when the system thickness increases (one should consider media with attenuation). Here we will give numerical examples. Figures 3(a) and (b) show the reflection coefficient at frequency 10 MHz from a 100-mm-thick single unidirectional layer ori-

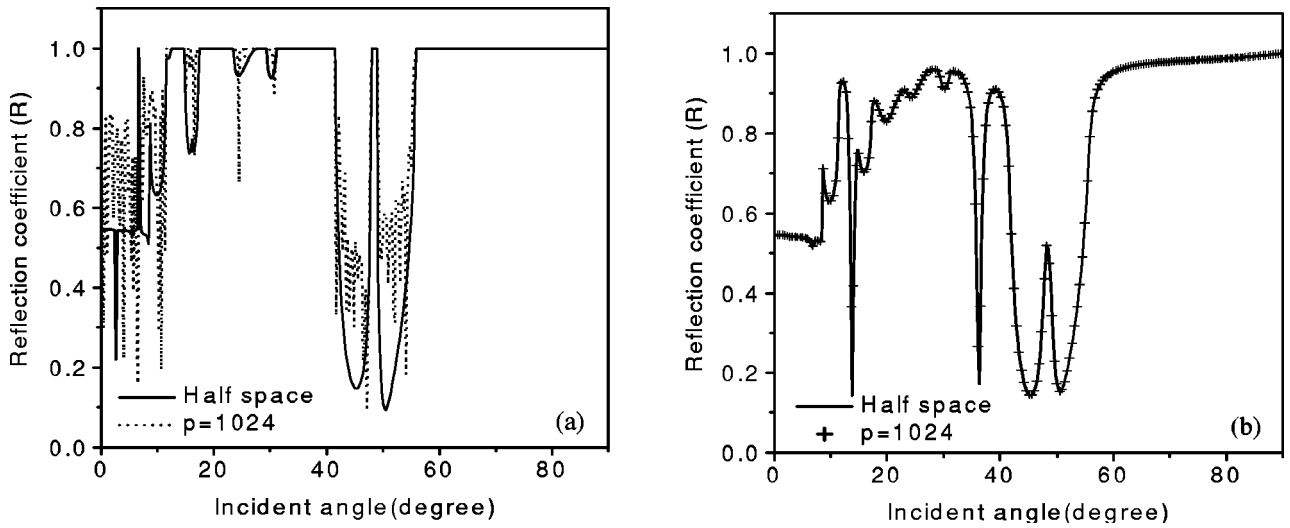


FIG. 4. Reflection coefficient for a periodic [0°/45°/90°/−45°] composite semispace (—) and multilayered 745-mm-thick structure (··· or +). Incident plane is orientated 0° to the top lamina. Wave is incident from water at frequency 3 MHz. (a) Without attenuation, (b) with attenuation.

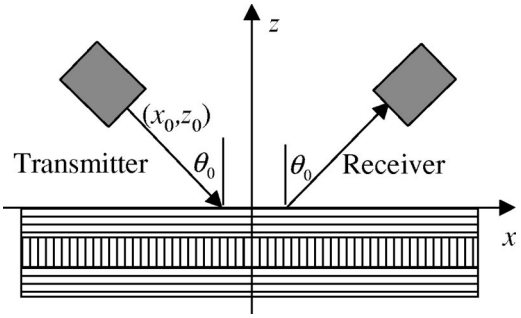


FIG. 5. Angle beam measurement setup ($x_0=35.2$ mm, $z_0=27.5$ mm, $\theta_0=52°$).

ented 45° to the plane of incidence. The solid line indicates the exact semispace solution. In Fig. 3(a), we assume no attenuation (the imaginary part in C_{ij} is zero in Table I). At incident angles below the slow shear wave critical angle, at least one of the wave numbers will be pure real, leading to resonances due to the existence of propagating modes. As soon as the incident angle is above the slow shear wave critical angle all wave numbers are complex, and the layer reflection at this thickness and frequency becomes unity as it is for the semispace. In Fig. 3(b), we use a material with attenuation (accounted for by imaginary parts in Table I). In this case, propagating waves are damped and the layer solution coincides with that for the semispace. Figures 4(a) and (b) show the results for a multi-ply composite with 4096 layers ($[0°/45°/90°/-45°]_{1024}$ layout with a total thickness $=795$ mm). The incident plane is oriented at 0° relative to the top lamina and frequency is 3 MHz. The semispace reflection coefficient is calculated using Floquet wave theory. Figure 4(a) shows the reflection coefficient when the attenuation is not included; the layer and semispace solutions are very similar in the stop bands where all Floquet wave numbers are complex. When attenuation is taken into account, the two solutions again coincide as shown in Fig. 4(b).

C. Time domain signal calculation and comparison with the experiment

The stability and computational efficiency of the recursive stiffness matrix method is essential for ultrasonic response calculations in time and angle domains. Here, as an example, we provide simulations of time domain response of angle beam reflection from multidirectional composites and compare them with experiments. Figure 5 shows the configuration of the angle beam experimental setup consisting of two transducers with a preselected angle of incidence. The transducer and sample are immersed in water. The output voltage for this dual transducer setup can be written as

$$V_{\text{out}}(z,t)=\int_{-\infty}^{+\infty} f_T(\omega) f_R(\omega) e^{i\omega t} d\omega \int_0^{\pi/2-i\infty} P_T(\theta,\omega) \times P_R(\theta,\omega) R(\theta,\omega) \exp(2i\omega/V_f(z_0\cos\theta + x_0\sin\theta)) d\theta. \tag{48}$$

$f_T(\omega)$ and $f_R(\omega)$ are the frequency responses of the transmitter and receiver which are experimentally determined when the transducers are focused on the surface of a homogeneous semispace solid such as aluminum. $P_T(\theta,\omega)$ and $P_R(\theta,\omega)$ are the angular responses of the transducers; V_f is the acoustic velocity in the fluid; x_0, z_0 is the center position of the transmitter. $R(\theta,\omega)$ is the plane wave reflection coefficient (31) from the immersed layered structure. To obtain the time domain response one needs to integrate over both the time and space domains which requires a significant number of computations of the kernel function $R(\theta,\omega)$ at different frequencies and incident angles. For the calculations we use Gaussian functions to approximate the angular responses [$P_T(\theta,\omega)$ and $P_R(\theta,\omega)$] of the transducers

$$P_{T,R}(\theta,\omega)=\frac{\delta_{T,R}}{\sqrt{2\pi}}\exp(k^2\delta_{T,R}^2/2), \tag{49}$$

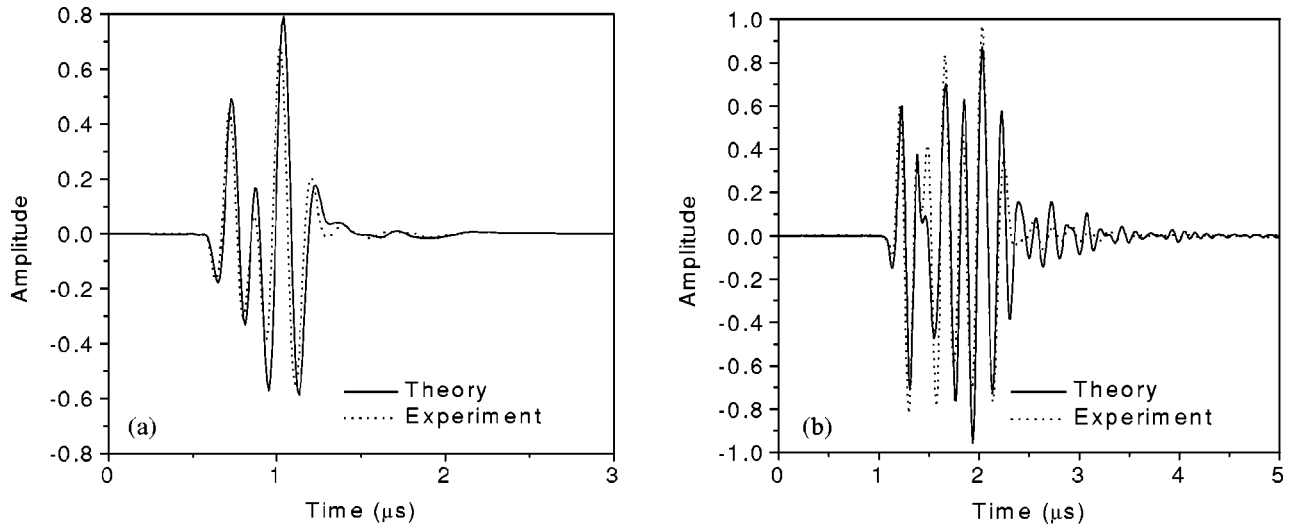


FIG. 6. Time-domain response for angle-beam measurement at 52° incidence. Dotted line represents experimental results and solid line simulated results. (a) The layered structure is a quasi-isotropic composite $[0/45/90/-45]_{2s}$. The orientation of the incident plane is 60° from the fiber direction of the top lamina. (b) The layered structure is a cross ply composite $[0/90]_{4s}$. The orientation of incident plane is 0°.

where $\delta_{T,R}$ is the beam width of the transmitter or receiver; $k=(\omega/V_f)\sin(\theta-\theta_0)$, θ_0 is the incident and reflection angle for the transducers. The beam widths $\delta_{T,R}$ are determined from experiment using the least-squares optimization fitting of reflected signals from a homogeneous (aluminum) semispace. The maximum integration angle and integration step are determined by trial and error. In most cases integration up to $\pi/2$ and integration step $\pi/1000$ are sufficient. Frequency domain integration is done by fast Fourier transform.

Figure 6(a) shows the time domain signal from a quasi-isotropic composite $[0/45/90/45]_{2s}$ for the transducer arrangement (Fig. 5) with incident angle 52° in water. The thickness of the composite is 3.1 mm and the incident plane orientation is 60°. The center frequency of the transducers used in this measurement is around 5 MHz. Figure 6(b) shows the same for a cross-ply composite $[0/90]_{4s}$ which has total thickness 2.5 mm (lamina thickness is 0.156 mm). The incident plane is oriented at 0°, the incident angle is also 52°. The experimental and computational results have reasonably good agreement considering that no fitting of parameters was done and the elastic properties of the lamina were obtained from independent experiments.

IV. CONCLUSION

It is well known that the transfer matrix (TM) algorithm, which is widely used for wave propagation analysis in layered media, is numerically unstable with increase of frequency or layer thickness. In this paper we have reformulated it in the form of a recursive stiffness matrix (RSM) algorithm, which has the same computational efficiency as the transfer matrix method, but is unconditionally stable: we have demonstrated the stability of the recursive matrix algorithm both theoretically and numerically. We find that the computational times for both methods (TM and RSM) are equal and depend linearly on the number of layers in the structure. The formulation of dispersion equations for Lamb and surface waves in the structure has also been addressed. The RSM algorithm is developed for a generally anisotropic medium when the stiffness matrix has dimensions (6×6); it is reduced to 4×4 for a plane of symmetry of orthotropic materials (and for higher symmetries, including isotropy). For application to laminated composites, we have adapted the algorithm to accommodate the repetition and symmetry in a layered structure, including derivation of the Floquet wave characteristic equation in terms of the stiffness matrix of the periodic structure unit cell. We applied this method to simulation of angle beam pulse reflection from multidirectional composites and obtained good agreement with experiment.

ACKNOWLEDGMENT

This work was partially sponsored by Adler Consultants Inc. through Navy Small Business Administration Program.

APPENDIX A: CALCULATIONAL PROCEDURE

The wave numbers and polarization vectors for the six waves in an anisotropic layer can be found by solving the Christoffel equation

$$(c_{ijln}k_jk_n-\rho\omega^2\delta_{il})p_l=0. \tag{A1}$$

This system of equations has a nontrivial solution for the unit displacement vector **p** only if the determinant of the system equals zero.

The algorithm for elastic wave propagation through an interface between two generally anisotropic solids is discussed in Ref. 6. Since all the wave vectors of the incident, reflected, and transmitted waves lie in one plane, it is possible to select the coordinate system in such a way that the interface plane coincides with the (x,y) coordinate plane and the incident plane coincides with the (x,z) coordinate plane (we also use indices 1,2,3 to denote coordinates x, y, z subsequently). The tensor of elastic constants for each layer is transformed into this coordinate system. In the selected principal coordinate system, all wave vectors will have $k_y=0$. From Snell's law, it follows that all the k_x are equal; therefore they are known because k_x is known for the incident wave. The unknown remaining component of the wave vector, k_z, can be found from

$$|c_{ijln}k_jk_n-\rho\omega^2\delta_{il}|=\det(\mathbf{G}(k_z))=0. \tag{A2}$$

Equation (A2) is a sixth-order polynomial in k_z.[6] For a given k_x, there will be six separate solutions for k_z in each medium which can be easily obtained numerically; k_z can be purely real or complex. A complex k_z corresponds to an evanescent wave. If one considers attenuation in the medium, real k_z becomes complex with a small imaginary part. A unit displacement **p** can be determined for each k_z from Eq. (A1). We use the general representation of Ref. 6 since it is valid both out of and in planes of symmetry,

$$\begin{aligned}
&W_{11}=G_{22}G_{33}-G_{23}^2, \quad W_{12}=G_{13}G_{23}-G_{12}G_{33},\\
&W_{22}=G_{33}G_{11}-G_{13}^2, \quad W_{13}=G_{12}G_{23}-G_{13}G_{22},\\
&W_{33}=G_{11}G_{22}-G_{12}^2, \quad W_{23}=G_{13}G_{12}-G_{23}G_{11},\\
&\Delta=W_{11}+W_{22}+W_{33}.
\end{aligned} \tag{A3}$$

Then the three components of the polarization vector are given by

$$p_x^2=W_{11}/\Delta, \quad p_y^2=W_{22}/\Delta, \quad p_z^2=W_{33}/\Delta. \tag{A4}$$

One can select the one with maximum absolute value, p_i, whose sign is arbitrary. If we use the positive sign, then p_i is given by

$$p_i=\sqrt{W_{ii}/\Delta}. \tag{A5}$$

The other component p_j $(j\neq i)$ can be calculated using the relationship

$$p_ip_j=W_{ij}/\Delta. \tag{A6}$$

According to the direction of energy flow (the ray velocity direction), the six plane waves are separated into three propagating in the $+z$ direction and three in the $-z$ direction. The components of the ray velocity V_g^i can be found if the displacement vector **p** is known:

$$V_g^i=\frac{1}{\rho\omega}c_{ijkl}k_lp_jp_k. \tag{A7}$$

For computing wave propagation in cross-ply composites, the layer in the layered system is represented by arbitrary rotation of an orthotropic unidirectional lamina around the symmetry axis z, which is equivalent to a layer with monoclinic symmetry. The constitutive relation in this case may be represented as

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{13} \\ \sigma_{12} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & C_{16} \\ C_{12} & C_{22} & C_{23} & 0 & 0 & C_{26} \\ C_{13} & C_{23} & C_{33} & 0 & 0 & C_{36} \\ 0 & 0 & 0 & C_{44} & C_{45} & 0 \\ 0 & 0 & 0 & C_{45} & C_{55} & 0 \\ C_{16} & C_{26} & C_{36} & 0 & 0 & C_{66} \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{22} \\ e_{33} \\ \gamma_{23} \\ \gamma_{13} \\ \gamma_{12} \end{bmatrix}, \tag{A8}$$

where the C_{ij} represent the elastic constants. We used the contracted subscript notations 1→11, 2→22, 3→33, 4→23, 5→13, 6→12 to relate C_{mn} to C_{ijkl}. In this case, the matrix elements of **G** in Eq. (A2) are

$$\begin{aligned} &G_{11}=C_{11}k_x^2-\rho\omega^2+C_{55}k_z^2, \quad G_{12}=C_{16}k_x^2+C_{45}k_z^2, \\ &G_{13}=(C_{13}+C_{55})k_xk_z, \\ &G_{22}=C_{66}k_x^2-\rho\omega^2+C_{44}k_z^2, \quad G_{23}=(C_{36}+C_{45})k_xk_z, \\ &G_{33}=C_{55}k_x^2-\rho\omega^2+C_{33}k_z^2. \end{aligned} \tag{A9}$$

Equation (A2) is a sixth-order polynomial in k_z with the six roots satisfying the conditions

$$k_z^{-1}=-k_z^{+1}, \quad k_z^{-2}=-k_z^{+2}, \quad k_z^{-3}=-k_z^{+3}. \tag{A10}$$

The stress coefficient $\mathbf{d}_j^{\pm}$ used in Eq. (5) in the $x-y$ plane is obtained from the constitutive relation

$$\mathbf{d}_j^{\pm}=\begin{bmatrix} d_1 \\ d_2 \\ \pm d_3 \end{bmatrix}_j = \begin{bmatrix} iC_{55}(p_x^jk_z^j+p_z^jk_x)+iC_{45}p_y^jk_z^j \\ iC_{45}(p_x^jk_z^j+p_z^jk_x)+iC_{44}p_y^jk_z^j \\ iC_{13}p_x^jk_x+C_{33}p_z^jk_z^j+iC_{36}p_y^jk_x \end{bmatrix}. \tag{A11}$$

In planes of symmetry,[13] the matrix elements C_{16}, C_{26}, C_{36}, and C_{45} are zero. One can solve Eq. (A2) analytically to obtain k_z. If we only consider the sagittal plane motion, then only four solutions are needed:

$$k_z^{+1}=k_z^{+3}=\sqrt{[-B\pm(B^2-4AC)^{1/2}]/2A}, \tag{A12}$$

where

$$\begin{aligned} &A=C_{33}C_{55}, \\ &B=-(C_{11}k_x^2-\rho\omega^2)C_{33}+(C_{55}k_x^2-\rho\omega^2)C_{55} \\ &\quad -(C_{13}+C_{55})^2k_x^2, \\ &C=-(C_{11}k_x^2-\rho\omega^2)(C_{55}k_x^2-\rho\omega^2). \end{aligned} \tag{A13}$$

The polarization vector **p** and stress **d** are reduced to

$$\begin{aligned} &p_x^i=(C_{13}+C_{55})k_xk_z^i, \quad p_z^i=(\rho\omega^2-C_{11}k_x^2-C_{55}(k_z^i)^2), \\ &d_1^i=\sigma_{13}=i(C_{55}p_x^ik_z^i+C_{55}p_z^ik_x), \\ &d_3^i=\sigma_{33}=i(C_{13}p_x^ik_x+C_{33}p_z^ik_z^i). \end{aligned} \tag{A14}$$

In this case, the stiffness matrix reduces to a 4×4 matrix. Equations (12)–(14) are also applicable for an isotropic solid.

APPENDIX B: REFLECTION AND TRANSMISSION COEFFICIENTS

Let us assume that an ultrasonic wave is incident on a layered structure in a fluid from the top fluid semispace as shown in Fig. 1. The acoustic potential in the upper half space ($z>0$) can be written as

$$\begin{aligned} \Phi_1(k_x,z,\omega)=&(A_{\text{in}}\exp(-iz\omega/V_f\cos\theta) \\ &+R\exp(iz\omega/V_f\cos\theta))e^{i(k_xx-\omega t)}, \end{aligned} \tag{B1}$$

and for the transmitted wave in the lower semispace

$$\Phi_2(k_x,z,\omega)=T\exp(i(k_xx-z\omega/V_f\cos\theta-\omega t)), \tag{B2}$$

where A_{in} are the incident amplitude and R and T are reflection and transmission coefficients, respectively. The displacements and stresses on the top surface will be

$$\begin{aligned} \begin{bmatrix} u_1^0 \\ u_2^0 \\ u_3^0 \\ \sigma_{13}^0 \\ \sigma_{23}^0 \\ \sigma_{33}^0 \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \cos\theta/V_f \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -i\omega\rho_f \end{bmatrix} \begin{bmatrix} u_1^0 \\ u_2^0 \\ R \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -\cos\theta/V_f \\ 0 \\ 0 \\ -i\omega\rho_f \end{bmatrix} A_{\text{in}} \\ &=\mathbf{FR}+\mathbf{F}_{\text{in}}A_{\text{in}}. \end{aligned} \tag{B3}$$

On the bottom surface,

$$\begin{bmatrix} u_1^N \\ u_2^N \\ u_3^N \\ \sigma_{13}^N \\ \sigma_{23}^N \\ \sigma_{33}^N \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\cos\theta/V_f \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -i\omega\rho_f \end{bmatrix} \begin{bmatrix} u_1^N \\ u_2^N \\ T \end{bmatrix} = \mathbf{FT}, \tag{B4}$$

where ρ_f is the fluid density, and V_f is the acoustic velocity. θ is the incident angle.

The compliance matrix for the whole structure is given by Eq. (23),

$$\begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_N \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma}_0 \\ \boldsymbol{\sigma}_N \end{bmatrix}. \tag{B5}$$

At the top and bottom surfaces, we have the shear stress $(\sigma_{31}^0,\sigma_{32}^0,\sigma_{31}^N,\sigma_{32}^N)$ equal to zero. Therefore the two equations with normal displacements (u_3^0,u_3^N) can be written as

$$u_3^0=S_{11}^{33}\sigma_{33}^0+S_{12}^{33}\sigma_{33}^N, \tag{B6}$$

and

$$u_3^N=S_{21}^{33}\sigma_{33}^0+S_{22}^{33}\sigma_{33}^N. \tag{B7}$$

The displacements and stresses $(u_3^0,u_3^N,\sigma_{33}^0,\sigma_{33}^N)$ can also be expressed as incident, reflected and transmitted waves from Eqs. (B3) and (B4). Upon substituting these equations into Eqs. (B6) and (B7), the reflection and transmission coeffi-

cients can be expressed as Eqs. (27) and (28).

[1] W. T. Thomson, J. Appl. Phys. **21**, 89–93 (1950).
[2] N. A. Haskell, Bull. Seismol. Soc. Am. **43**, 17–34 (1953).
[3] L. M. Brekhovskikh, *Waves in Layered Media* (Academic, New York, 1960).
[4] J. W. Dunkin, Bull. Seismol. Soc. Am. **55**, 335–358 (1965).
[5] A. H. Fahmy and E. L. Adler, Appl. Phys. Lett. **20**, 495–497 (1973).
[6] S. I. Rokhlin, T. K. Bolland, and L. Adler, J. Acoust. Soc. Am. **79**, 906–918 (1986).
[7] D. Levesque and L. Piche, J. Acoust. Soc. Am. **92**, 452–467 (1992).
[8] M. J. S. Lowe, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525–542 (1995).
[9] E. L. Adler, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 485–490 (1990).
[10] *Wave Propagation in Structure Composites*, edited by A. K. Mal and T. C. T. Ting (ASME-AMD, ASME New York, NY, 1988), Vol. 90.
[11] A. H. Nayfeh and D. E. Chimenti, J. Appl. Mech. **55**, 863–870 (1988).
[12] A. H. Nayfeh, J. Acoust. Soc. Am. **89**, 1521–1531 (1991).
[13] A. H. Nayfeh, *Wave Propagation in Layered Anisotropic Media* (North-Holland, Amsterdam, 1995).
[14] F. I. Solyanik, Sov. Phys. Acoust. **23**, 533–536 (1977).
[15] S. I. Rokhlin and Y. J. Wang, J. Acoust. Soc. Am. **91**, 1875–1887 (1992).
[16] M. Castaings and B. Hosten, J. Acoust. Soc. Am. **94**, 1488–1495 (1993).
[17] B. Hosten and M. Castaings, J. Acoust. Soc. Am. **95**, 1931–1941 (1993).
[18] C. Potel and J. Belleval, J. Appl. Phys. **74**, 2208–2213 (1993).
[19] B. L. Kennett, Bull. Seismol. Soc. Am. **65**, 1685–1696 (1974).
[20] B. L. Kennett, *Seismic Wave Propagation in Stratified Media* (Cambridge University Press, New York, 1983).
[21] B. L. Kennett and N. J. Kerry, Geophys. J. R. Astron. Soc. **57**, 557–583 (1979).
[22] D. C. Booth and S. Crampin, Geophys. J. R. Astron. Soc. **72**, 755–766 (1983).
[23] D. C. Booth and S. Crampin, Geophys. J. R. Astron. Soc. **72**, 767–782 (1983).
[24] G. J. Fryer and L. N. Frazer, Geophys. J. R. Astron. Soc. **78**, 691–710 (1984).
[25] G. J. Fryer and L. N. Frazer, Geophys. J. R. Astron. Soc. **91**, 73–101 (1987).
[26] S. Mallick and L. N. Frazer, Geophys. J. Int. **105**, 241–252 (1991).
[27] B. Nolte, L. N. Frazer, and S. Mallick, Geophys. J. Int. **111**, 127–140 (1992).
[28] S. I. Rokhlin and W. Huang, J. Acoust. Soc. Am. **92**, 1729–1742 (1992).
[29] H. Schmidt and G. Tango, Geophys. J. R. Astron. Soc. **84**, 331–359 (1986).
[30] H. Schmidt and F. B. Jensen, J. Acoust. Soc. Am. **77**, 813–825 (1985).
[31] A. K. Mal, Wave Motion **9**, 231–238 (1988).
[32] E. Kausel and J. Roesset, Bull. Seismol. Soc. Am. **71**, 1743–1761 (1981).
[33] Y. Wang and R. K. N. D. Rajapakse, J. Appl. Mech. **61**, 339–347 (1994).
[34] L. Wang and S. I. Rokhlin, Rev. Prog. Quant. Nondestr. Eval. **20B**, 1015–1022 (2000).
[35] L. Wang and S. I. Rokhlin, Rev. Prog. Quant. Nondestr. Eval. **11**, 1321–1328 (1998).
[36] E. Kausel, Int. J. Numer. Methods Eng. **23**, 1567–1578 (1986).
[37] E. Kausel, Int. J. Numer. Methods Eng. **37**, 927–941 (1994).
[38] F. B. Jensen, W. A. Kuperman, M. P. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer, Berlin, 2000), Chap. 4.
[39] C. Potel and J. Belleval, J. Acoust. Soc. Am. **93**, 2669–2670 (1993).
[40] A. M. Braga and G. Herrmann, J. Acoust. Soc. Am. **91**, 1211–1227 (1992).
[41] P. J. Shull, D. E. Chimeti, and S. K. Datta, J. Acoust. Soc. Am. **95**, 99–103 (1994).
[42] C. Potel, J. Belleval, and Y. Gargouri, J. Acoust. Soc. Am. **97**, 2815–2827 (1995).
[43] A. Auld, D. E. Chimenti, and P. J. Shull, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 319–323 (1996).
[44] S. I. Rokhlin and W. Huang, J. Acoust. Soc. Am. **94**, 3405–3420 (1993).

Diffraction correction for precision surface acoustic wave velocity measurements

Alberto Ruiz M. and Peter B. Nagy[a)]
Department of Aerospace Engineering and Engineering Mechanics, University of Cincinnati, Cincinnati, Ohio 45221-0700

(Received 1 December 2001; revised 15 March 2002; accepted 24 May 2002)

Surface wave dispersion measurements can be used to nondestructively characterize shot-peened, laser shock-peened, burnished, and otherwise surface-treated specimens. In recent years, there have been numerous efforts to separate the contribution of surface roughness from those of near-surface material variations, such as residual stress, texture, and increased dislocation density. As the accuracy of the dispersion measurements was gradually increased using state-of-the-art laser-ultrasonic scanning and sophisticated digital signal processing methods, it was recognized that a perceivable dispersive effect, similar to the one found on rough shot-peened specimens, is exhibited by untreated smooth surfaces as well. This dispersion effect is on the order of 0.1%, that is significantly higher than the experimental error associated with the measurements and comparable to the expected velocity change produced by near-surface compressive residual stresses in metals below their yield point. This paper demonstrates that the cause of this apparent dispersion is the diffraction of the surface acoustic wave (SAW) as it travels over the surface of the specimen. The results suggest that a diffraction correction may be introduced to increase the accuracy of surface wave dispersion measurements. A simple diffraction correction model was developed for surface waves and this correction was subsequently validated by laser-interferometric velocity measurements on aluminum specimens.
[DOI: 10.1121/1.1497368]

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

I. INTRODUCTION

Surface acoustic waves (SAWs) produce elastic displacements and stresses that are confined to the surface within a shallow depth of approximately 1 wavelength. Therefore, they are particularly sensitive to the boundary conditions at the surface and to the material properties in the near-surface region. Since the penetration depth of surface waves is frequency dependent, high-frequency SAWs are more affected by the material properties close to the surface while the low-frequency components give information on the bulk material properties of the substrate. This frequency dependence can be exploited to control the depth of the near-surface region to be inspected. Another interesting feature of surface waves is that they are less affected by beam spreading than bulk waves, since they are confined to travel on the surface of the material; therefore, they diverge only in one dimension rather than in two like bulk modes.

The use of surface acoustic waves for the characterization of surface roughness, surface residual stress, and coating thickness measurements has been studied by several authors.[1–5] The effect of surface cracks on the attenuation and dispersion of Rayleigh waves was explored by Zhang and Achenbach,[6] Warren *et al.*,[7] and Pecorari.[8–11] These studies have found that the total change in the Rayleigh wave velocity is typically less than 0.3%. The challenge, of course, is to measure the Rayleigh wave velocity with sufficiently high accuracy. At this precision, the surface wave velocity seems to be affected by beam diffraction, and it may be necessary to use appropriate diffraction corrections in order to assess the true material-related dispersion. Of course, in many cases when the velocity exhibits a more dramatic change, e.g., in the case of coated substrates, this diffraction correction might be negligible.

The theory of diffraction effects in the ultrasonic field of a piston source is straightforward and has been treated by many authors.[12–18] The adverse effect of diffraction in ultrasonic attenuation measurements is well known, and appropriate corrections are routinely used to increase the accuracy.[19–25] The application of similar diffraction corrections in ultrasonic velocity measurements is usually not necessary except for the most stringent accuracy requirements.[19,25–30] For a piston transducer radiating into a fluid medium, the complex diffraction correction can be calculated by either analytical means or numerical integration. For example, Rogers and Van Buren found a closed-form expression for the so-called Lommel diffraction correction for a circular piston radiator in the Fresnel approximation.[16] From the phase of their complex diffraction correction, the ''apparent'' local velocity of the diverging wave can be readily calculated. One finds that the apparent velocity is strongly frequency dependent even in nondispersive media and the dispersion decreases as the frequency increases, i.e., as the field asymptotically approaches a true plane wave.[19] The apparent dispersion also depends on the size of the receiver, which is important since the receiver aperture is often different from, typically much smaller than, the transmitter aperture. Using numerical integration, it can be shown that

a)Electronic mail: pnagy@uceng.uc.edu

the apparent velocity dispersion is perceivably higher when a point receiver is moved along the axis of the same transmitter instead of a finite-aperture piston receiver.

To the best of our knowledge, similar results are not available in the literature for surface waves. The main goals of this paper are to present a simple diffraction correction model for surface waves and to experimentally validate this model using laser-interferometric surface wave velocity measurements on smooth aluminum specimens. It will be demonstrated that a readily measurable apparent dispersion on the order of 0.1% is exhibited by untreated smooth surfaces, which is significantly higher than the inherent experimental error associated with the measurements and comparable to the expected velocity change produced by near-surface compressive residual stresses in shot-peened metals. Therefore, high-precision surface wave dispersion measurements used in nondestructive materials characterization should be corrected for diffraction effects in order to achieve maximum accuracy.

II. THEORETICAL MODEL

Diffraction effects begin at the face of the transmitter and continue out to a point where the transducer is so far away that it might be considered as a point source, and the diffraction effects can be easily approximated as the spreading of a divergent beam of given directivity pattern. Generally, both the amplitude and phase of the generated acoustic field deviate from the ideal case of a plane wave. In this section we will consider Rayleigh wave propagation on the free surface of an elastic half-space. For different transmitter/receiver configurations of practical importance, we will analyze the phase of the complex diffraction correction and derive appropriate dispersion corrections for precision velocity measurements.

Let us assume that the acoustic field of a given radiator is scanned by a receiver moving along the axis of the radiator. Generally, the receiver is sensitive to the weighted average of the measured field parameter (displacement, velocity, pressure, etc.) over the receiver aperture. The size of the receiver aperture can be either negligibly small (point sensor) or finite. At a given axial distance z from the radiator, the measured field parameter m can be written as

$$m(z)=m_0e^{i(kz-\omega t)}D(z)=m_0e^{i(kz-\omega t)}F(z)e^{i\varphi(z)}. \tag{1}$$

Here, m_0 is a constant, t denotes time, ω is the angular frequency, $k=\omega/c$ is the wave number, where c is the (nondispersive) sound velocity, $D(z)$ is a complex diffraction coefficient that represents the deviation of the field distribution from that of an ideal propagating plane wave, $F(z)$ denotes the amplitude variation along the axis, and $\varphi(z)$ is a diffraction phase correction. The ''apparent'' local phase velocity c_a can be calculated from the measured field parameter using

$$\frac{\omega}{c_a}=\frac{\partial}{\partial z}\mathrm{Arg}\{m(z)\}=\frac{\omega}{c}+\frac{\partial\varphi}{\partial z}. \tag{2}$$

Let us introduce the normalized velocity diffraction correction $\eta=(c_a-c)/c$ as the relative change of the apparent velocity with respect to the nondispersive real sound velocity c. We will also use $\zeta=z/a$ for the normalized distance and $s=ak$ for the normalized frequency, where a is an appropriate reference dimension, e.g., the half-length of a line source. Then, the normalized velocity diffraction correction can be expressed from Eq. (2) as follows:

$$\eta=\frac{1}{1+(1/s)\ (\partial\varphi/\partial\zeta)}-1=-\frac{1}{s}\frac{\partial\varphi}{\partial\zeta}. \tag{3}$$

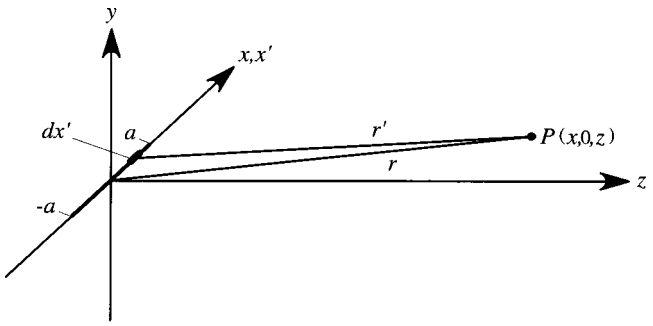


FIG. 1. Coordinate system used to calculate the acoustic field of a line source on the free surface ($y=0$) of an elastic half-space.

In ultrasonic nondestructive materials characterization, we often use Rayleigh-type surface waves for precision velocity measurements. The coordinate system used to calculate the acoustic field of a line source on the free surface ($y=0$) of an elastic half-space is shown in Fig. 1. The normal surface displacement of the Rayleigh wave produced by a line source of finite length on an isotropic half-space can be obtained from Lamb's classical solution for a point source[31]

$$u(x,z)=u_0e^{-i\omega t}\sqrt{\frac{k}{2\pi i}}\int_{-a}^{a}dx'\,\frac{e^{ikr'}}{\sqrt{r'}}, \tag{4}$$

where $r'^2=x'^2+z^2$, and u_0 is a constant determined by the load intensity and polarization of the source and the shear modulus and Poisson's ratio of the material. Equation (4) is the one-dimensional equivalent of the well-known two-dimensional Rayleigh integral,[32] which is often used to calculate the complex diffraction correction for bulk waves. It should be mentioned that the numerical prefactor in front of the integral could be easily incorporated into the constant u_0, but then its physical meaning would be changed. As it is, u_0 is the asymptotic limit of the normal surface displacement amplitude of the nondiverging Rayleigh wave produced by the line source when its length approaches infinity.

A. Line source and point receiver

Let us assume that we are using a point receiver to scan the normal surface displacement along the axis of the transducer ($x=0$). Then, we can write the displacement field as $u(z)=u_0e^{i(kz-\omega t)}D_{a0}(z)$, where the complex diffraction correction can be written in a normalized form as follows:

$$D_{a0}(\zeta)=\sqrt{\frac{s}{2\pi i}}\int_{-1}^{1}d\xi'\,\frac{e^{is(\sqrt{\xi'^2+\zeta^2}-\zeta)}}{\sqrt[4]{\xi'^2+\zeta^2}}, \tag{5}$$

where $\xi'=x'/a$, and $\zeta=z/a$ and $s=ak$ as before. Due to the above-described normalization, D_{a0} asymptotically approaches unit as $s\to\infty$. Equation (5) can be solved numeri-

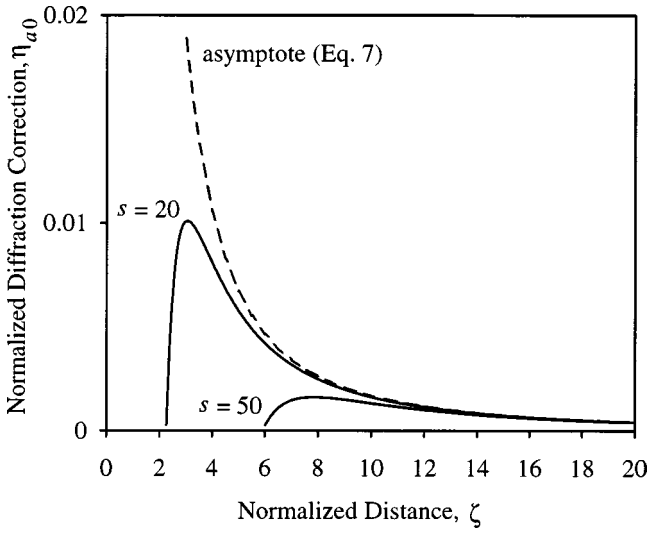


FIG. 2. Normalized diffraction correction for surface wave velocity measurements made by a point receiver along the axis of a line transmitter.

cally, and then the velocity diffraction correction can be calculated from Eq. (3). As an example, Fig. 2 shows the calculated normalized diffraction correction for surface wave velocity measurements made by a point receiver along the axis of a line transmitter. In the far field of the line source

$$\lim_{\zeta\to\infty} D_{a0}(\zeta) = \sqrt{\frac{2s}{\pi i\zeta}}\left(1+\frac{is}{6\zeta}\right). \tag{6}$$

From Eqs. (3) and (6), the far-field asymptote of the normalized velocity diffraction correction can be written as follows:

$$\lim_{\zeta\to\infty} \eta_{a0} = \frac{1}{6\zeta^2}. \tag{7}$$

It should be mentioned that the complex diffraction correction D could be more concisely written if the axial distance z were normalized to the near-field/far-field distance $N=a^2/\lambda$, as it is customary in the literature. For example, Eq. (6) could be rewritten as follows:

$$\lim_{S\to\infty} D_{a0}(S) = \frac{2}{\sqrt{iS}}\left(1+\frac{i\pi}{3S}\right), \tag{8}$$

where $S=z/N=2\pi\zeta/s$ with our previous notation. However, the normalized velocity diffraction correction requires differentiation with respect to the normalized frequency s according to Eq. (3); therefore, it is actually simpler when expressed in terms of ζ than in terms of S [see Eq. (7) and our results later].

B. Line source and line receiver

Surface waves are routinely detected by laser interferometers that are essentially point receivers. However, in order to increase the accuracy of the measurements the detection spot is often either extended laterally to a line of substantial length using a cylindrical lens or the focus spot is optically scanned in the lateral direction, while the received signal is digitized and averaged. This is often necessary to reduce the otherwise very high sensitivity of the point receiver to spatially incoherent structural noise in inherently inhomogeneous materials such as polycrystalline metals.[33] The question arises as to how much the velocity diffraction correction changes as a result of this averaging. Assuming that the line receiver has a total length of $2b$, the average normal surface displacement can be calculated from Eq. (4) as follows:

$$u_{\text{avg}}(z) = u_0 e^{-i\omega t}\sqrt{\frac{k}{2\pi i}}\frac{1}{2b}\int_{-b}^{b} dx \int_{-a}^{a} dx' \frac{e^{ikr'}}{\sqrt{r'}}$$
$$= u_0 e^{i(kz-\omega t)} D_{ab}(z), \tag{9}$$

where the complex diffraction correction can be written in the previously introduced normalized form as follows:

$$D_{ab}(\zeta) = \sqrt{\frac{s}{2\pi i}}\frac{1}{2\beta}\int_{-\beta}^{\beta} d\xi \int_{-1}^{1} d\xi' \frac{e^{is(\sqrt{(\xi-\xi')^2+\zeta^2}-\zeta)}}{\sqrt[4]{(\xi-\xi')^2+\zeta^2}}, \tag{10}$$

where $\beta=b/a$ denotes the length ratio between the receiver and the transmitter.

In general, the length of the receiver b can be either smaller or larger than the transmitter length a, and Eq. (10) can be solved only numerically. In the far field of both transducers, i.e., when $z\gg a^2/\lambda$ and $z\gg b^2/\lambda$, r' can be approximated by $r'=z+(x-x')^2/2z$ in the rapidly changing exponent and simply by $r'=z$ in the slowly changing denominator of Eq. (9). Furthermore, the exponential function can be approximated by the two leading terms in its Taylor expansion, so that we get

$$D_{ab}(\zeta) = \sqrt{\frac{s}{2\pi i\zeta}}\frac{1}{2\beta}\int_{-\beta}^{\beta} d\xi \int_{-1}^{1} d\xi' \left[1+is\frac{(\xi-\xi')^2}{2\zeta}\right], \tag{11}$$

that can be readily integrated to get the following approximation for the complex diffraction correction:

$$\lim_{\zeta\to\infty} D_{ab}(\zeta) = \sqrt{\frac{2s}{\pi i\zeta}}\left[1+\frac{is(1+\beta^2)}{6\zeta}\right]. \tag{12}$$

Finally, from Eqs. (3) and (12), the far-field asymptote of the normalized velocity diffraction correction can be written as follows:

$$\lim_{\zeta\to\infty} \eta_{ab} = \frac{1+\beta^2}{6\zeta^2}. \tag{13}$$

Clearly, for $b=0$ $\eta_{ab}(\zeta)=\eta_{a0}(\zeta)$, while for $b=a$ the apparent excess velocity is doubled, i.e., $\eta_{aa}(\zeta)=2\eta_{a0}(\zeta)$. This increase in the apparent phase velocity with respect to the case of a point receiver is due to the contribution of those oblique wave components that reach the receiver at larger angles, therefore at lower group velocity, but higher phase velocity along the axis.

C. Wedge transmitter and line receiver

Most angle-beam SAW transducers consist of a longitudinal transducer of circular cross section mounted on a polymer wedge. This configuration produces an apodized SAW

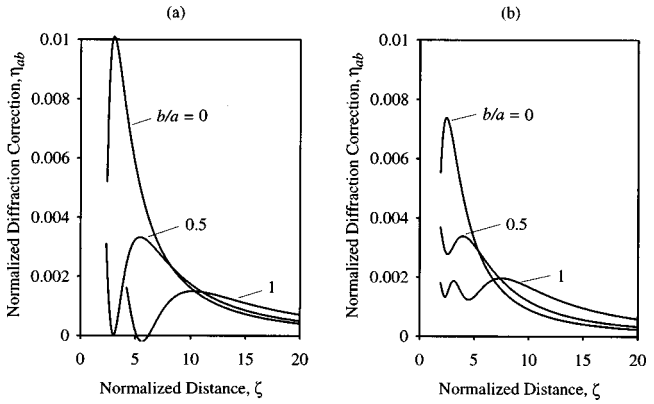


FIG. 3. Diffraction corrections for velocity measurements made by a finite-length line receiver along the axis of another line transmitter for $s=20$ without (a) and with (b) apodization.

displacement distribution because of the elliptical footprint of the circular piston on the surface. Although peripheral rays are further attenuated by leakage back to the wedge as the surface wave leaves the transmitter, a simple apodized model can be constructed by assuming a cosine SAW displacement distribution. Assuming that the receiver is not apodized, the complex diffraction correction given in Eq. (10) should be modified as follows:

$$D_{ab}(\zeta)=\sqrt{\frac{s}{2\pi i}}\,\frac{1}{2\beta}\int_{-\beta}^{\beta}d\xi\int_{-1}^{1}d\xi'\,\cos\left(\xi'\,\frac{\pi}{2}\right)\times\frac{e^{is(\sqrt{(\xi-\xi')^2+\zeta^2}-\zeta)}}{\sqrt[4]{(\xi-\xi')^2+\zeta^2}}. \tag{14}$$

Such apodization slightly reduces the velocity correction by weakening the oblique rays responsible for the apparent increase in phase velocity. Figure 3 shows the calculated normalized diffraction corrections for surface wave velocity measurements made by a finite-length line receiver along the axis of a line transmitter for $s=20$ (a) without and (b) with apodization. As before, most of the near-field oscillations were cut off for clarity. Close to the transmitter the correction decreases with increasing receiver length as the averaging over the transducer's aperture smoothes the phase variations, while further away it increases. The effect of apodization is twofold. First, it effectively reduces the transmitter length. Second, it smoothes the near-field oscillations.

III. EXPERIMENTAL METHOD AND RESULTS

Figure 4 shows a schematic diagram of the experimental arrangement used in our surface wave dispersion measurements. A wedge transducer was mounted on the specimen and exited by a Panametrics 5270 broadband pulser. Three 12.7-mm-diameter Accuscan-S Panametrics screw-in transducers of 2.25-, 3.5-, and 5-MHz nominal center frequency were used on an ABWML-5ST 90° wedge also made by Panametrics. The propagating surface wave was detected by a LUIS 35 Fabry–Perot interferometer. The laser beam was aimed at the surface of the specimen at a small angle of incidence, and an objective lens focused the diffuse reflection onto the tip of an optical fiber connected to the interferometer. The ultrasonic signal detected by the interferometer was digitized and averaged by a LeCroy 9310 oscilloscope and then sent to a computer for further processing. The specimen was mounted on a Velmex translation table and the relative position between the wedge transmitter and the laser spot was changed by a computer-accessible stepping motor controller. The scanning resolution of the translation table was 3.175 μm. Both the data acquisition and the scanner were controlled by a LABVIEW program. In order to assure the absolute accuracy of our velocity measurements, the temperature of the specimen was stabilized at 22 °C within ± 0.7 °C (temperature variations are not expected to lead to dispersion unless the surface temperature of the specimen is significantly different from that of the interior). The specimens were scanned in two directions. Spatial averaging was used during lateral scanning normal to the wave propagation direction to reduce the incoherent scattering by (i) surface roughness and (ii) the inhomogeneous microstructure as well as (iii) coherent diffraction effects, especially in the near field of the transmitter. A slower stepwise scanning was per-

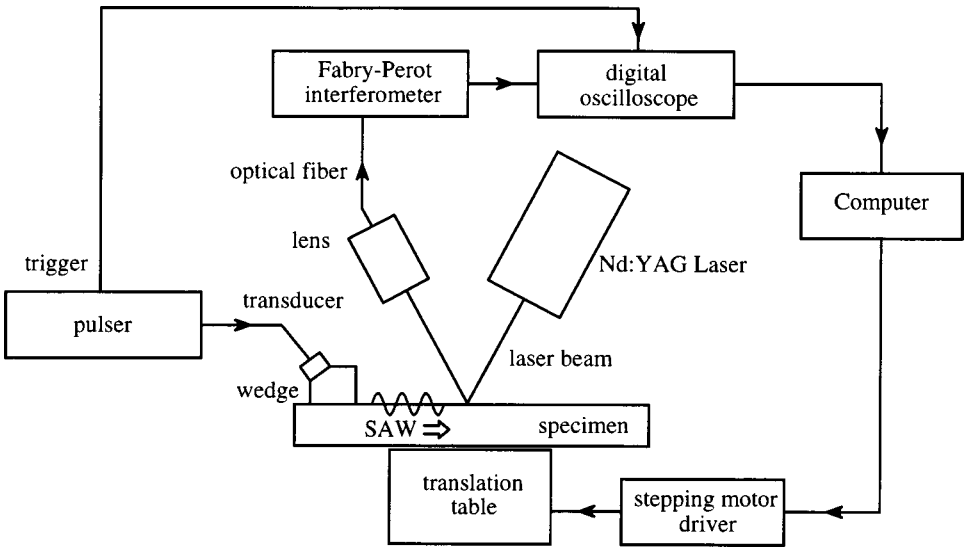


FIG. 4. Experimental setup for laser-ultrasonic SAW velocity measurement.

formed in the axial direction, i.e., parallel to the wave propagation, in order to map the phase of the surface vibration as a function of the propagation distance.

Six 2024-T351 aluminum bars of 50.8 mm (width)×203 mm (length)×12.7 mm (height) were carefully polished with 2000-grade sandpaper parallel to the length of the specimen to minimize scattering by surface irregularities. The specimens were shot-peened over a 50.8-mm×50.8-mm square area at their center using 1.4-mm-diameter steel shots and 100% coverage. Two identical series of specimens were prepared using three different Almen intensities (4A, 8A, and 10A).[34,35] The unpeened smooth side of one of the specimens was used as a reference (0A). The first series of specimens was evaluated directly after shot peening and the second set was heat treated for 1 h at 350 °C before the ultrasonic tests. The objective of the heat treatment was to remove near-surface material variations caused primarily by the shot peening, but, to a lesser degree, potentially present in the original cold-rolled bar stock material as well. These variations are associated with subtle near-surface effects such as the presence of residual stresses, increased hardness, dislocation density, and anisotropic texture. With the exception of surface roughness-induced scattering, annealing above the recrystallization temperature of the material effectively eliminates all other near-surface variations, which might otherwise contribute to the observed surface wave dispersion. To some degree, annealing also changes the microstructure and can lead to perceivable grain coarsening when it is overdone. As a result, the surface wave velocity might also change slightly, but, since the effect is essentially the same throughout the whole volume of the specimen, it does not necessarily lead to surface wave dispersion, though it might cause some microstructural dispersion via grain scattering in both bulk and surface wave propagation.

In our measurements we first placed the front of the wedge transducer at a 30-mm distance from the laser spot and then scanned the beam along the propagation direction over a total distance of 50 mm, i.e., up 80 mm from the wedge, in steps of 80 μm. At each of the 625 axial positions, we averaged 1000 pulses to increase the signal-to-noise ratio. In addition, during averaging we also used lateral scanning normal to the propagation direction to further improve the accuracy of the measurement. The lateral scanning dimension was ± 2.5 mm, i.e., significantly smaller than the beam width (the diameter of the piezoelectric transducer mounted on the wedge was 12.7 mm). In this way, both temporal and spatial averaging was achieved at the same time and the laser interferometer effectively acted as a 5-mm-long line receiver.

The LABVIEW program gradually changed the delay time of the LeCroy digital oscilloscope with respect to the trigger signal by using the programmable digital delay and highly accurate quartz master clock of the oscilloscope. At each step the new delay time was calculated from the scanning position using a nominal tracking velocity. In this way, the recorded signal moved only very slightly either forward or backward within the data window depending on whether the carefully chosen tracking velocity c_n was higher or lower, respectively, than the actual surface wave velocity. In an ideal case, the tracking velocity is very close to the actual one, and the broadband pulse does not move appreciably at all, though its shape slightly changes as a result of dispersion. The recorded data were spectrum analyzed by a discrete Fourier transform algorithm that determined the phase of the signal at 10–20 different frequencies depending on the bandwidth of the transmitter (the bandwidth of the interferometer is essentially flat from 2 to 100 MHz).

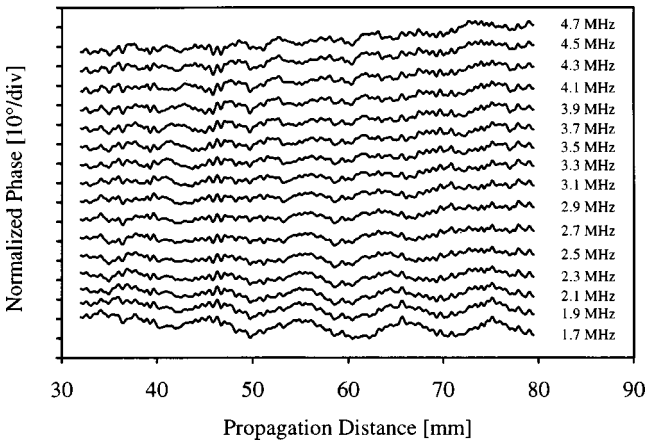


FIG. 5. Example of the measured normalized phase versus propagation distance curves on an Almen 4 shot-peened specimen at 16 different frequencies.

As an example, Fig. 5 shows the normalized phase as a function of the propagation distance obtained from a 4A-intensity shot-peened specimen using a 3.5-MHz transducer. In order to achieve the highest possible phase accuracy, we have to limit the overall range over which the phase might change. To a large degree, this is achieved by carefully selecting the nominal tracking velocity that is used to control the delay of the digitizer. However, some time shift will inevitably occur, which in turn will produce large changes in phase that are linearly proportional to frequency. These changes might hide smaller frequency-dependent changes caused by dispersive wave propagation, and therefore must be further suppressed. This was achieved by "normalizing" the actual phase φ measured at any given frequency f to the center frequency f_n of the transducer according to $\varphi_n = \varphi f_n / f$. This normalization of the phase allowed us to observe the change in the slope of the phase due to dispersion, as well as to avoid folding back of the phase at 360° intervals.

The velocity of the SAW was calculated from the slope of the normalized phase using the following formula:

$$c_m = \frac{c_n}{1+(c_n/\omega)\ (\partial\varphi/\partial z)} = \frac{c_n}{1+(c_n/\omega_n)\ (\partial\varphi_n/\partial z)}, \tag{15}$$

where c_m is the measured surface wave velocity, c_n is the tracking velocity chosen to minimize the observed total phase variation over the fairly long propagation distance. Equation (15) is the same as Eq. (2) with the measured surface wave velocity c_m taking the place of the apparent local phase velocity c_a, and the constant tracking velocity c_n taking the place of the nondispersive sound velocity c. On the left side of Eq. (15) we introduced the angular frequency

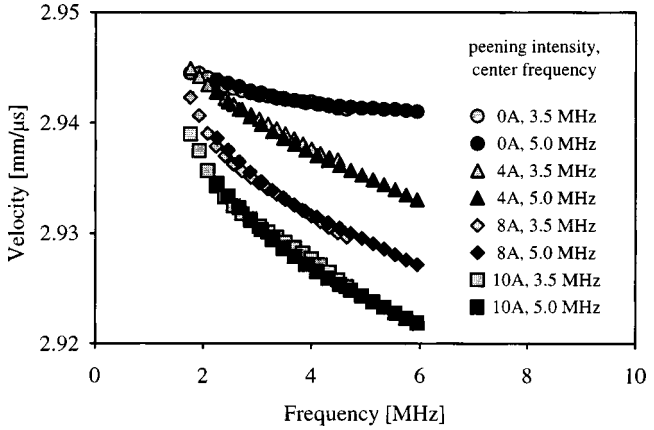


FIG. 6. Surface wave velocity measurements on four different aluminum specimens using two different transmitters of 3.5- and 5-MHz nominal center frequencies.

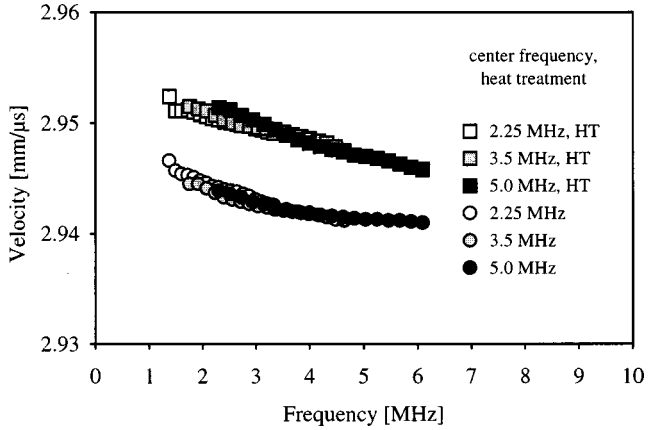


FIG. 7. Surface wave velocity measurements on unpeened smooth aluminum specimens before and after annealing (HT) using three transmitters of 2.25-, 3.5-, and 5-MHz nominal center frequencies.

used for normalization $\omega_n = 2\pi f_n$, that is typically chosen to be either the nominal or actual center frequency of the transmitter. The average slope of the normalized phase $\partial\varphi_n/\partial z$ was calculated at each frequency by finding the best-fitting linear approximation to all data points using the least-mean-square method. In Fig. 5, the low-frequency slopes are slightly negative, which, according to Eq. (15), means that the low-frequency components propagated at a velocity slightly higher than the tracking velocity. In comparison, the high-frequency slopes are positive, which indicates that they propagated slower than the tracking velocity, i.e., the velocity decreased with increasing frequency over the bandwidth of the transmitter.

Figure 6 shows the results of our surface wave velocity measurements on four different aluminum specimens using two different transmitters of 3.5- and 5-MHz nominal center frequencies. On the three shot-peened specimens we can observe that the dispersion increases with peening intensity. This effect is mainly caused by surface roughness-induced scattering, though the other previously mentioned effects of shot peening also contribute. For our immediate purposes, the most important observation is that, although the shot-peened specimens obviously exhibit much stronger dispersion, some dispersion, that is well above the experimental uncertainty of the measurement ($\approx \pm 0.02\%$), is also exhibited by the smooth, untreated specimen.

Figure 7 shows the results of our surface wave velocity measurements on unpeened smooth aluminum specimens before and after annealing using three transmitters of different nominal center frequencies. As was mentioned above, the observed dispersion even on an unpeened smooth surface suggests that the apparent velocity is affected not only by the surface roughness and other material variations (residual stress, dislocation density, local texture, and hardening) produced by shot peening, but also by the intrinsic diffraction of the surface acoustic wave. Because of the inevitable microstructural change caused by annealing, the absolute surface wave velocity is $\approx 0.2\%$ higher in the heat-treated specimen. Actually, the apparent dispersion is also somewhat higher in the heat-treated sample than in the original one, which indicates that some contribution from the microstructure is also present in the dispersion. It should be mentioned that at even higher frequencies the true material dispersion caused by grain scattering in the polycrystalline aluminum is expected to dominate, assuming that no imperfections are present.[36]

Figure 8 shows the comparison between the experimental dispersion curve measured on a smooth aluminum specimen and the theoretical predictions of our diffraction model. The normalized surface wave velocity was calculated from the apodized diffraction model using the experimental parameters (including a 22-mm added propagation length within the wedge from the transducer to the surface of the specimen) and the best-fitting $c = 2.9402$ mm/μs surface wave velocity. The measured velocity is more affected at low frequencies, while it becomes more frequency independent at high frequencies. When we plot the experimental normalized velocity against the theoretical prediction, there is a fairly good qualitative agreement between them. These results suggest that the measured velocity is perceivably affected by the previously discussed diffraction effects, and the apparent dispersion should be corrected for diffraction in order to realize the full potential of high-precision dispersion measurements in ultrasonic materials characterization.

These experimental results demonstrate that a measurable dispersive effect is exhibited by well-polished smooth

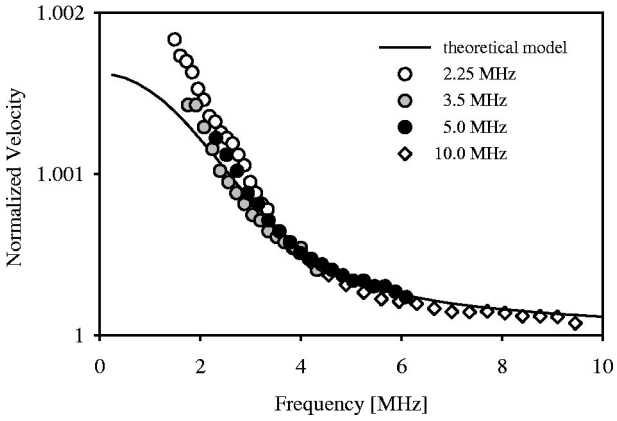


FIG. 8. Comparison between the experimental dispersion curve measured on a smooth aluminum specimen and theoretical predictions of the diffraction model ($c = 2.9402$ mm/μs).

surfaces, which cannot be attributed to surface roughness scattering. This behavior can be explained by our theoretical results that demonstrate that diffraction effects do cause an apparent dispersion that has very similar features (magnitude and frequency dependence) to those of the experimentally observed results (see Fig. 8). Of course, the measured dispersion could also be caused, at least in part, by the intrinsic attenuation in the solid, e.g., grain scattering. As we mentioned above, the perceptible increase of dispersion in annealed aluminum (see Fig. 7) could be attributed to excess material dispersion, especially at higher frequencies. In order to quantitatively estimate the degree of material dispersion caused by grain-scattering effects on SAW propagation, we could rely on existing, though extremely complicated, theoretical predictions.[36] Alternatively, much simpler order-of-magnitude type of predictions can be made based on the observed frequency-dependent attenuation using the well-known Kramers–Kronig relationship.

After Viktorov, we are going to estimate the dispersion of the SAW from the behavior of the shear wave, since it can be measured much easier and much more accurately.[37] Figure 9(a) shows the experimentally determined attenuation coefficient versus frequency curves for the shear wave in our aluminum specimen before and after annealing. In order to verify the similarity between the attenuation of the shear and Rayleigh waves, we also included in Fig. 9(a) the corresponding SAW results for the more attenuating annealed case. One of the difficulties in applying the global Kramers–Kronig relationship is the necessity of knowing the frequency dependence of the attenuation coefficient at every frequency to calculate the dispersion at a given frequency,[38,39] although approximate methods are also available in the literature to calculate the dispersion from the nearly local behavior of the attenuation.[40,41] Since in the case of grain scattering in polycrystalline materials the behavior of the attenuation coefficient is known over the whole frequency range, we used the global Kramers–Kronig relationship. First, we estimated the grain size by matching our experimental data with Stanke and Kino's unified model,[42] then used this model to estimate the grain-scattering-induced dispersion. It should be mentioned that similar results also can be obtained directly from Stanke and Kino's unified model for dispersion. Figure 9(a) shows the theoretically estimated material dispersion based on the global Kramers–Kronig relationship (the scale was kept the same as in Fig. 8 for easier comparison). These results indicate that the intrinsic dispersion of the specimen, as estimated from the measured frequency-dependent attenuation, is at least one order of magnitude lower than the measured dispersion and provides additional evidence that the measured apparent dispersion is indeed caused by the diffraction effect.

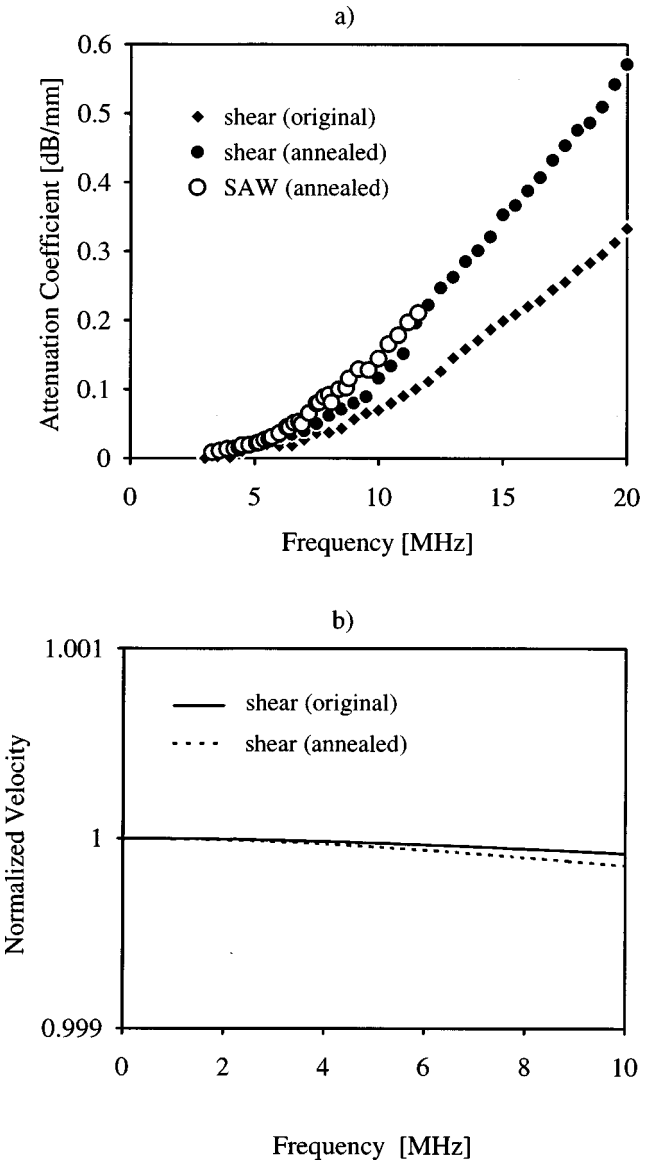


FIG. 9. Experimentally determined attenuation coefficient versus frequency curves (a) and the theoretically estimated material dispersion based on the global Kramers–Kronig relationship (b).

IV. CONCLUSIONS

Surface wave dispersion measurements are often used in ultrasonic, nondestructive testing to characterize near-surface material variations that might lead to significant dispersion. It was found, however, that a perceivable dispersion is exhibited by untreated smooth surfaces as well. Although this dispersion effect is only on the order of 0.1%, i.e., relatively small with respect to the dispersion caused by typical near-surface material variations, it is clearly higher than the experimental error associated with the measurements themselves. This paper demonstrated that the cause of this apparent dispersion, at least in part, is the diffraction of the SAW as it travels over the surface of the specimen. This conclusion suggests that a diffraction correction may be used to increase the accuracy of precision surface wave dispersion measurements. A simple diffraction correction model was developed for surface waves and this correction was subsequently validated by laser-interferometric velocity measurements on smooth aluminum specimens. Further work is needed to develop practical ways of separating diffraction-related dispersion from both surface roughness and microstructural scattering and from additional dispersive effects due to the presence of near-surface residual stresses and texture.

ACKNOWLEDGMENTS

The authors wish to thank the Universidad Michoacana de San Nicolás De Hidalgo and CONACYT-MEXICO for their support of Alberto Ruiz Marines during his Ph.D. studies. This work was supported by the Metals, Ceramics, and NDE Division of the Air Force Research Laboratory under Contract No. 01-S437-002-34-C1.

[1] A. I. Lavrentyev and W. A. Veronesi, ''Ultrasonic measurement of residual stress in shot peened aluminum alloy,'' in *Rev. Progr. Quant. Nondestr. Eval.* (AIP, Melville, 2001), Vol. 20, pp. 1472–1479.
[2] C. Glorieux and W. Gao, ''Surface acoustic wave depth profiling of elastically inhomogeneous materials,'' J. Appl. Phys. **88**, 4394–4400 (2000).
[3] M. Duquennoy, M. Ouaftouh, M. L. Qian, F. Jenot, and M. Ourak, ''Ultrasonic characterization of residual stresses in steel rods using a laser line source and piezoelectric transducers,'' NDT & E Int. **34**, 355–362 (2001).
[4] V. V. Krylov and Z. A. Smirnova, ''Experimental study of the dispersion of a Rayleigh wave on a rough surface,'' Sov. Phys. Acoust. **36**, 583–585 (1990).
[5] F. Lakestani, J. F. Coste, and R. Denis, ''Application of ultrasonic Rayleigh waves to thickness measurement of metallic coatings,'' NDT & E Int. **28**, 171–178 (1995).
[6] C. Zhang and J. D. Achenbach, ''Dispersion and attenuation of surface waves due to distributed surface-breaking cracks,'' J. Acoust. Soc. Am. **88**, 1986–1992 (1990).
[7] P. D. Warren, C. Pecorari, O. V. Kolosov, S. G. Roberts, and G. A. D. Briggs, ''Characterization of surface damage via surface acoustic waves,'' Nanotechnology **7**, 295–301 (1996).
[8] C. Pecorari, ''Modeling variations of Rayleigh wave velocity due to distribution of one-dimensional surface-breaking cracks,'' J. Acoust. Soc. Am. **100**, 1542–1550 (1996).
[9] C. Pecorari, ''Rayleigh wave dispersion due to a distribution of semi-elliptical surface-breaking cracks,'' J. Acoust. Soc. Am. **103**, 1383–1387 (1998).
[10] C. Pecorari, ''Attenuation and dispersion of Rayleigh waves propagating on a cracked surface: An effective field approach,'' Ultrasonics **38**, 754–760 (2000).
[11] C. Pecorari, ''Scattering of a Rayleigh wave by a surface-breaking crack with faces in partial contact,'' Wave Motion **33**, 259–270 (2001).
[12] A. O. Williams, Jr., ''The piston source at high frequencies,'' J. Acoust. Soc. Am. **23**, 1–6 (1951).
[13] H. Seki, A. Granato, and R. Truell, ''Diffraction effects in the ultrasonic field of a piston source and their importance in the accurate measurement of attenuation,'' J. Acoust. Soc. Am. **28**, 230–238 (1956).
[14] R. Bass, ''Diffraction effects in the ultrasonic field of a piston source,'' J. Acoust. Soc. Am. **30**, 602–605 (1958).
[15] E. P. Papadakis, ''Diffraction of ultrasound radiating into an elastically anisotropic medium,'' J. Acoust. Soc. Am. **36**, 414–422 (1964).
[16] P. H. Rogers and A. L. Van Buren, ''An exact expression for the Lommel diffraction correction integral,'' J. Acoust. Soc. Am. **55**, 724–728 (1974).
[17] K. Beissner, ''Exact integral expression for the diffraction loss of a circular piston source,'' Acustica **49**, 212–217 (1981).
[18] C. J. Daly and N. A. H. K. Rao, ''Time- and frequency-domain descriptions of spatially averaged one-way diffraction for an unfocused piston transducer,'' Ultrasonics **37**, 209–221 (1999).
[19] M. B. Gitis and A. S. Khimunin, ''Diffraction corrections for measurements of the absorption coefficient and velocity of sound,'' Sov. Phys. Acoust. **14**, 305–310 (1969).
[20] M. B. Gitis and A. S. Khimunin, ''Diffraction effects in ultrasonic measurements,'' Sov. Phys. Acoust. **14**, 413–431 (1969).
[21] A. S. Khimunin, ''Numerical calculation of the diffraction corrections for the precise measurement of ultrasound absorption,'' Acustica **27**, 173–181 (1972).
[22] E. P. Papadakis, K. A. Fowler, and L. C. Lynnworth, ''Ultrasonic attenuation by spectrum analysis of pulses in buffer rods: Method and diffraction correction,'' J. Acoust. Soc. Am. **53**, 1336–1343 (1973).
[23] W. Cobb, ''Frequency domain method for the prediction of ultrasonic field patterns of pulsed, focused radiators,'' J. Acoust. Soc. Am. **75**, 72–79 (1984).
[24] X. Chen, K. Q. Schwarz, and K. J. Parker, ''Acoustic coupling from a focused transducer to a flat plate and back to the transducer,'' J. Acoust. Soc. Am. **95**, 3049–3054 (1994).
[25] J. K. Kushibiki and M. Arakawa, ''Diffraction effects on bulk-wave ultrasonic velocity and attenuation measurements,'' J. Acoust. Soc. Am. **108**, 564–573 (2000).
[26] H. J. McSkimin, ''Empirical study of the effect of diffraction on velocity of propagation of high-frequency ultrasonic waves,'' J. Acoust. Soc. Am. **32**, 1401–1404 (1960).
[27] E. P. Papadakis, ''Ultrasonic diffraction loss and phase change in anisotropic materials,'' J. Acoust. Soc. Am. **40**, 863–876 (1966).
[28] E. P. Papadakis, ''Ultrasonic phase velocity by the pulse-echo-overlap method incorporating diffraction phase corrections,'' J. Acoust. Soc. Am. **42**, 1045–1051 (1967).
[29] A. S. Khimunin, ''Ultrasonic propagation parameter measurements incorporating exact diffraction corrections,'' Acustica **39**, 87–95 (1978).
[30] K. Negishi, K. Takagi, and H. Ozawa, ''Observation of phase advance due to diffraction of ultrasonic pulse,'' J. Acoust. Soc. Jpn. **1**, 11–15 (1980).
[31] H. Lamb, ''On the propagation of tremors over the surface of an elastic solid,'' Philos. Trans. R. Soc. London, Ser. A **203**, 1–41 (1904).
[32] J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Sec. 278.
[33] W. Hassan and P. B. Nagy, ''Experimental investigation of the grain noise in interferometric detection of ultrasonic waves,'' J. Nondestruct. Eval. **18**, 139–148 (1999).
[34] John O. Almen and P. H. Black, *Residual Stresses and Fatigue in Metals* (McGraw-Hill, New York, 1963).
[35] J. S. Eckersley and J. Champaigne, *Shot Peening: Theory and Application* (IITT-International, Gournay-Sur-Marne, 1991).
[36] I. M. Kaganova and A. A. Maradudin, ''Surface acoustic waves on a polycrystalline substrate,'' Phys. Scr. **T44**, 104–112 (1992).
[37] I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[38] R. L. Weaver and Y. H. Pao, ''Dispersion relations for linear wave propagation in homogeneous and inhomogeneous media,'' J. Math. Phys. **22**, 1909–1918 (1981).
[39] K. R. Waters, M. S. Hughes, J. Mobley, and J. G. Miller, ''On the applicability of Kramers–Kronig relations for ultrasonic attenuation obeying a frequency power law,'' J. Acoust. Soc. Am. **108**, 556–563 (2000).
[40] M. O. Donnell, E. T. Jaynes, and J. G. Miller, ''Kramers–Kronig relationship between ultrasonic attenuation and phase velocity,'' J. Acoust. Soc. Am. **69**, 696–701 (1981).
[41] P. He, ''Experimental verification of models for determining dispersion from attenuation,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 706–714 (1999).
[42] F. E. Stanke and G. S. Kino, ''A unified theory for elastic wave propagation in polycrystalline materials,'' J. Acoust. Soc. Am. **75**, 665–681 (1984).

Phenomenological model of propagation of the elastic waves in a fluid-saturated porous solid with nonzero boundary slip velocity

David Tsiklauri[a)]
Physics Department, University of Warwick, Coventry CV4 7AL, United Kingdom

(Received 30 January 2002; revised 26 March 2002; accepted 31 May 2002)

It is known that a boundary slip velocity starts to play an important role when the length scale over which the fluid velocity changes approaches the slip length, i.e., when the fluid is highly confined, for example, fluid flow through porous rock or blood vessel capillaries. Zhu and Granick [Phys. Rev. Lett. **87**, 096105 (2001)] have recently experimentally established the existence of a boundary slip in a Newtonian liquid. They reported typical values of the slip length of the order of few micrometers. In this light, the effect of introduction of the boundary slip into the theory of propagation of elastic waves in a fluid-saturated porous medium formulated by Biot [J. Acoust. Soc. Am. **28**, 179–191 (1956)] is investigated. Namely, the effect of introduction of boundary slip upon the function $F(\kappa)$ that measures the deviation from Poiseuille flow friction as a function of frequency parameter κ is studied. By postulating *phenomenological* dependence of the slip velocity upon frequency, notable deviations in the domain of intermediate frequencies in the behavior of $F(\kappa)$ are introduced with the incorporation of the boundary slip into the model. It is known that $F(\kappa)$ crucially enters Biot's equations, which describe dynamics of fluid-saturated porous solid. Thus, consequences of the nonzero boundary slip by calculating the phase velocities and attenuation coefficients of both rotational and dilatational waves with the variation of frequency are investigated. The new model should allow one to fit the experimental seismic data in circumstances when Biot's theory fails, as the introduction of phenomenological dependence of the slip velocity upon frequency, which is based on robust physical arguments, adds an additional degree of freedom to the model. In fact, it predicts *higher than the Biot's theory values of attenuation coefficients of the both rotational and dilatational waves* in the intermediate frequency domain, which is in qualitative agreement with the experimental data. Therefore, the introduction of the boundary slip yields threefold benefits: (1) better agreement of theory with experimental data since the parametric space of the model is larger (includes effects of boundary slip); (2) the possibility to identify types of porous medium and physical situations where boundary slip is important; and (3) constrain model parameters that are related to the boundary slip. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1499134]

PACS numbers: 43.20.Jr, 47.27.Lx, 47.55.Mh, 47.60.+i, 62.65.+k [JJM]

I. INTRODUCTION

It has been a common practice in fluid dynamics to assume that when a fluid flows over an interface with a solid, the fluid molecules adjacent to the solid have zero relative velocity with respect to the solid. So far, this widely used assumption, known as ''no-slip boundary condition,'' has been successfully applied to the theoretical modeling of almost all macroscopic experiments. As relevantly noticed by Craig *et al.*,[1] the success of this assumption does not reflect its accuracy, but rather insensitivity of the experiment to a partial-slip boundary condition. It is known that the boundary slip becomes important only when the length scale over which the fluid velocity changes approaches the slip length, that is the distance behind the interface at which the fluid velocity extrapolates to zero, i.e., when the fluid becomes highly confined, e.g., blood flow through capillaries or fluid flow through natural porous rock. Recently, the authors of Refs. 1 and 2 presented convincing experimental evidence of a boundary slip in a Newtonian liquid. They performed direct measurements of the hydrodynamic drainage forces, which show clear evidence of boundary slip. Also, they found that the boundary slip is a function of the fluid viscosity and the shear rate. These results have important implications for the blood dynamics in capillaries, the permeability of porous media, and lubrication of nanomachines. For example, the results of Craig *et al.* suggest that red blood cells squeeze through narrow capillary walls more easily and induce less shear stress on capillary walls due to the boundary slip. Also, in the oil production industry, the residual oil is difficult to produce due to its naturally low mobility. Thus, the enhanced oil recovery operations are used to increase production. It has been experimentally proven and theoretically validated that there is a substantial increase in the net fluid flow through a porous medium if the latter is treated with elastic waves.[3–6] We may conjecture that the elastic waves via the pore wall vibration cause boundary slip of the residual oil droplets, which like red blood cells, squeeze through pores with less resistance, effectively increasing permeability of the porous medium.

a)Electronic mail: tsikd@astro.warwick.ac.uk

A quantitative theory of propagation of elastic waves in a fluid-saturated porous solid has been formulated in the classic paper by Biot.[7] After its appearance, this theory has seen numerous modifications and generalizations. One of the major findings of Biot's work was that there was a breakdown in Poiseuille flow above a certain characteristic frequency specific to this fluid-saturated porous material. Biot theoretically studied this breakdown by considering the flow of a viscous fluid in a tube with longitudinally oscillating walls under an oscillatory pressure gradient. Biot's theory can be used to describe interaction of fluid-saturated porous solid with sound for a classic Newtonian fluid assuming no-slip boundary condition at the pore walls holds. However, in light of the recent experimental results of Ref. 2, revision of the classic theory is needed in order to investigate novelties brought about by the boundary slip.

Biot's theory has been a successful tool for interpreting the experimental data for decades, but there are circumstances when it fails. Gist[8] performed ultrasonic velocity and attenuation measurements in sandstones with a variety of saturating fluids and compared the data with the predictions of the Biot's theory. *Velocity data show systematic deviations from Biot theory as a function of pore fluid viscosity. Ultrasonic attenuation is much larger than the Biot prediction* and for several sandstones is nearly constant for a three decade variation in viscosity. This behavior contrasts with synthetic porous media such as sintered glass beads, where Biot theory provides accurate predictions of both velocity and attenuation. A similar claim has also been made by earlier experimental studies, e.g., Ref. 9. Even Biot himself acknowledged[10] that the model of a purely elastic solid matrix saturated with a viscous fluid is rather idealistic and there are rare occasions where it gives satisfactory agreement with the experiment. Thus, he generalized theory applying the correspondence principle, and incorporated the solid and fluid attenuation, and dissipation due to nonconnected pores. In this work Biot introduced the viscodynamic operator, which is a sum of the viscosity and inertial terms, and it accounts for the dynamic properties of the fluid motions in pores valid for the both low and high frequencies. The theory can model other, more complex, mechanisms such as thermoelastic dissipation, which produce a relaxation spectrum, or local flow mechanisms as the ''squirt'' flow.[11,12] A time domain formulation based on memory variables has been also successfully introduced.[13,14] In order to narrow down the source of discrepancy between Biot theory and experiment, Gist[8] tried to modify natural sandstone by curing a residual saturation of epoxy in the pore space, filling small pores and microcracks. This altered rock has a significantly reduced attenuation, demonstrating the dominance of small pores in controlling ultrasonic attenuation. Then he suggested that two non-Biot viscous attenuation mechanisms are needed: local flow in microcracks along grain contacts, and attenuation from pore-wall surface roughness. The model based on these two mechanisms seemed to succeed in improvement of fitting the data. However, this by no means excludes other possibilities, and all plausible effects that could explain the discrepancy should be explored. At low frequencies, the fluid flow is of Poiseuille type and the inertia effects are obviously negligible in comparison with the effects of the viscosity. At high frequencies, however, these effects are confined to a thin boundary layer in the vicinity of the pore walls and the inertia forces are dominant. Therefore, *we conjecture that in such situations, apart from other yet unknown effects, nonzero boundary slip effect may be responsible for the deviations between theory and experiment*. This justifies our aim to formulate a model that would account for the boundary slip.

In Sec. II we formulate the theoretical basis of our model and in Sec. III we present our numerical results. In Sec. IV we conclude with the discussion of the results.

II. THE MODEL

In our model we study a Newtonian fluid flowing in a cylindrical tube, which mimics a natural pore, whose walls are oscillating longitudinally and the fluid is subject to an oscillatory pressure gradient. We give analytical solutions of the problem in the frequency domain.

The governing equation of the problem is the linearized momentum equation of an incompressive fluid

$$\rho \frac{\partial \boldsymbol{v}}{\partial t} = -\nabla p + \mu \nabla^2 \boldsymbol{v}. \tag{1}$$

Here, $\boldsymbol{v}$, p, ρ denote velocity, pressure, and mass density of the fluid, while μ is the viscosity coefficient.

Now, let u be a velocity of the wall of the tube which oscillates in time as $e^{-i\omega t}$. The flow of fluid in a cylindrical tube with longitudinally oscillating walls can be described by a single component of the velocity, namely, its z component v_z (z axis is along the centerline of the tube). We use the cylindrical coordinate system (r,ϕ,z) in treatment of the problem. We introduce the relative velocity U_1 as $U_1 = v_z - u$. Thus, assuming that all physical quantities vary in time as $e^{-i\omega t}$, we arrive at the following master equation for U_1:

$$\nabla^2 U_1 + \frac{i\omega}{\nu} U_1 = -\frac{X}{\nu}. \tag{2}$$

Here, we have introduced the following notations:

$$\rho X = -\left(\nabla p + \rho \frac{\partial u}{\partial t} \right),$$

which is a sum of the applied pressure gradient and force exerted on the fluid from the oscillating wall of the tube and, ν, which is $\nu = \mu/\rho$.

The solution of Eq. (2) can be found to be[7]

$$U_1(r) = -\frac{X}{i\omega} + CJ_0(\beta r),$$

where J_0 is the Bessel function and $\beta = \sqrt{i\omega/\nu}$.

Assuming that the slip velocity is $U_1(a) = U_s$ at the wall of the tube, where a is its radius, we obtain

$$U_1(r) = -\frac{X}{i\omega}\left[1 - (1 + \overline{U_s}) \frac{J_0(\beta r)}{J_0(\beta a)} \right]. \tag{3}$$

Here,

$$\overline{U_s} \equiv U_s \frac{i\omega}{X} = U_s \frac{\nu}{a^2 X}(\beta a)^2.$$

Defining the cross-section averaged velocity as

$$\overline{U_1} = \frac{2}{a^2}\int_0^a U_1(r) r\, dr,$$

we obtain

$$\overline{U_1} = -\frac{Xa^2}{\nu}\frac{1}{(\beta a)^2}\left[1 - \frac{2(1+\overline{U_s})J_1(\beta a)}{(\beta a)J_0(\beta a)}\right]. \tag{4}$$

Following the work of Biot[7] we calculate the stress at the wall τ,

$$\tau = -\mu\left(\frac{\partial U_1(r)}{\partial r}\right)_{r=a} = \frac{\mu\beta X}{i\omega}(1+\overline{U_s})\frac{J_1(\beta a)}{J_0(\beta a)}. \tag{5}$$

The total friction force is $2\pi a\tau$. Following Biot we calculate the ratio of total friction force to the average velocity, i.e.,

$$\frac{2\pi a\mu}{\overline{U_1}} = -2\pi\mu(\beta a)(1+\overline{U_s})\frac{J_1(\beta a)}{J_0(\beta a)} \times\left[1 - \frac{2(1+\overline{U_s})J_1(\beta a)}{(\beta a)J_0(\beta a)}\right]^{-1}. \tag{6}$$

Simple analysis reveals that (assuming $\overline{U_s}\to 0$ as $\omega\to 0$, see the following discussion)

$$\lim_{\omega\to 0}\frac{2\pi a\tau}{\overline{U_1}} = 8\pi\mu,$$

which corresponds to the limiting case of Poiseuille flow. Following Biot,[7] we also introduce a function $F(\kappa)$ with κ being frequency parameter, $\kappa = a\sqrt{\omega/\nu}$, in the following manner:

$$\frac{2\pi a\tau}{\overline{U_1}} = 8\pi\mu F(\kappa),$$

thus,

$$F(\kappa) = -\frac{1}{4}\kappa\sqrt{i}(1+\overline{U_s})\frac{J_1(\kappa\sqrt{i})}{J_0(\kappa\sqrt{i})} \times\left[1 - (1+\overline{U_s})\frac{2J_1(\kappa\sqrt{i})}{\kappa\sqrt{i}J_0(\kappa\sqrt{i})}\right]^{-1}. \tag{7}$$

Note, that $F(\kappa)$ measures the deviation from Poiseuille flow friction as a function of frequency parameter κ. The Biot's expression for $F(\kappa)$ in the no boundary slip regime can be easily recovered from Eq. (7) by putting $\overline{U_s}\to 0$ for all κ's.

So far, we have not specified $\overline{U_s}$, however there are certain physical constraints it should satisfy.

(1) The authors of Ref. 1 demonstrated that the slip length (which, in fact, is proportional to the slip velocity) is a function of the approach rate and they showed why several previous careful measurements of confined liquids have not observed evidence for boundary slip. Under the low approach rates employed in previous measurements slip length is zero and no-slip boundary condition is applicable. Experiments reported in Ref. 1 were performed when the half-sphere approached a plane with liquid placed in-between at different approach rates. However, we should clearly realize what the term "approach rate" means in the context of Biot's theory: Biot investigated fluid flow in the cylindrical tube whose walls are harmonically oscillating in the longitudinal direction as $x(t) = Ae^{-i\omega t}$, therefore if similar to Ref. 1 the experiment would be done for the oscillation tube, and the "approach rate" would be the amplitude of $\dot{x}(t)$, i.e., $-i\omega A$. Thus, when $\omega\to 0$, $\overline{U_s}$ should also tend to zero.

(2) At high frequencies the viscous effects are negligible compared to the inertial effects. Thus, fluid at high frequencies behaves as an ideal (nonviscous) pore fluid, which allows for an arbitrary slip. Therefore, when $\omega\to\infty$, $\overline{U_s}$ should tend to zero. When we closely examine Fig. 6 from Ref. 1 and Fig. 3 (bottom panel) from Ref. 2 we should understand that the authors measured the ascending part where slip length increases with increase of the approach (flow) rate. Since, the viscous effects should be negligible compared to the inertial effects for large approach rates, there must also be a descending part—clearly slip length (or the same as slip velocity) cannot grow infinitely as approach rate increases. Viscous effects should give in to the inertial effects and fluid should behave as an ideal fluid allowing for an arbitrary slip.

Based upon the previous two physical arguments we conclude that $\overline{U_s}$ should be a smooth function which tends to zero when ω (or the same as κ) tends to both zero and ∞. Therefore, we postulate the following *phenomenological* expression for $\overline{U_s}$:

$$\mathrm{Re}[\overline{U_s}(\kappa)] = \mathrm{Im}[\overline{U_s}(\kappa)] = \xi\frac{B\kappa^4}{(A+\kappa^2)^5}.$$

This function has a maximum at $\kappa_* = \sqrt{2A/3}$. By fixing B at $5^5A^3/(4\times 3^3)$ we force it to achieve its maximum equal to unity at $\kappa = \kappa_*$ (when $\xi = 1$). We introduced ξ (usually $0 \leqslant \xi \leqslant 1$) as a sort of "weight" of the boundary slip effect into the solution. Putting $\xi = 0$ this way would easily allow us to recover the no-slip case, while an increase of ξ would allow us to trace the effect of nonzero slip onto the all physical quantities. We plot this function in Fig. 1 for $A = 25$ and $\xi = 1$.

It is worthwhile to mention that, at first glance, it seems that nonzero boundary slip, which appears at "high approach rates," should be attributed to the nonlinear effects. However, thorough interpretation of the experimental results of Ref. 1, in the context of the oscillatory tube [see points (1) and (2) previously noted] allows us to conclude that the nonzero boundary slip can also be incorporated into the *linear* Biot's theory.

Next, we plot both $F_r(\kappa) = \mathrm{Re}[F(\kappa)]$ and $F_i(\kappa) = -\mathrm{Im}[F(\kappa)]$ in Fig. 2 for the three cases: when there is no boundary slip ($\xi = 0$), and $\xi = 0.05, 0.1$. We gather from the plot that the $\xi = 0$ is identical to Fig. 4 of Ref. 7 as it should be. However, we also notice a noticeable difference from the classic case when ξ is nonzero in the *intermediate* frequencies domain. Of course, according to our definition of the

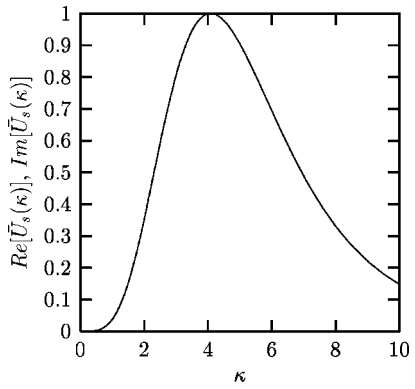


FIG. 1. Behavior of $\mathrm{Re}[\overline{U}_s(\kappa)]=\mathrm{Im}[\overline{U}_s(\kappa)]$ as function of frequency parameter, κ, for $\xi=1$, $A=25$.

phenomenological form of $\overline{U}_s(\kappa)$, even for nonzero ξ (nonzero boundary slip), when $\kappa\to 0$ and $\kappa\to\infty$, asymptotically $F(\kappa)$ behaves as classic Biot's solution,[7] i.e.,

$$\lim_{\kappa\to 0} F_r(\kappa)=1, \tag{8}$$

and

$$\lim_{\kappa\to 0} F_i(\kappa)=0, \tag{9}$$

$$\lim_{\kappa\to\infty} F(\kappa)=\frac{\kappa}{4}\sqrt{i}=\frac{\kappa}{4}\left(\frac{1+i}{\sqrt{2}}\right). \tag{10}$$

Since in our phenomenological model we allow for the deviations in the intermediate frequency domain, it is easy to foresee that these will have an impact on all of the predictions of Biot's theory precisely in that frequency range. Namely, all observable quantities predicted by Biot's theory, such as phase velocities and attenuation coefficients of both rotational and dilatational waves, will be affected by the introduction of boundary slip into the model in the intermediate frequency range.

Biot[7] showed that the general equations which govern propagation of rotational and dilatational high-frequency waves in a fluid-saturated porous medium are the same as in the low-frequency range provided the viscosity is replaced by its effective value as a function of frequency. In practice, it means replacing the resistance coefficient b by $bF(\kappa)$.

The equations describing dynamics of the rotational waves are[7]

$$\frac{\partial^2}{\partial t^2}(\rho_{11}\boldsymbol{\omega}+\rho_{12}\boldsymbol{\Omega})+bF(\kappa)\frac{\partial}{\partial t}(\boldsymbol{\omega}-\boldsymbol{\Omega})=N\nabla^2\boldsymbol{\omega}, \tag{11}$$

$$\frac{\partial^2}{\partial t^2}(\rho_{12}\boldsymbol{\omega}+\rho_{22}\boldsymbol{\Omega})-bF(\kappa)\frac{\partial}{\partial t}(\boldsymbol{\omega}-\boldsymbol{\Omega})=0, \tag{12}$$

where ρ_{11}, ρ_{12}, and ρ_{22} are mass density parameters for the solid and fluid and their inertia coupling; $\boldsymbol{\omega}=\mathrm{curl}\,\boldsymbol{u}$ and $\boldsymbol{\Omega}=\mathrm{curl}\,\boldsymbol{U}$ describe rotations of solid and fluid with $\boldsymbol{u}$ and $\boldsymbol{U}$ being their displacement vectors, while the rigidity of the solid is represented by the modulus N. Substitution of a plane rotational wave of the form

$$\omega=C_1e^{i(lx+\chi t)},\quad \Omega=C_2e^{i(lx+\chi t)}, \tag{13}$$

into Eqs. (11) and (12) allows us to obtain a characteristic equation

$$\frac{Nl^2}{\rho\chi^2}=E_r-iE_i, \tag{14}$$

where l is the wave number, $\chi=2\pi f$ is wave cyclic frequency, $\rho=\rho_{11}+2\rho_{12}+\rho_{22}$ is the mass density of the bulk material.

The real and imaginary parts of Eq. (14) can be written as

$$E_r=\frac{(\gamma_{11}\gamma_{22}-\gamma_{12}^2)(\gamma_{22}+\epsilon_2)+\gamma_{22}\epsilon_2+\epsilon_1^2+\epsilon_2^2}{(\gamma_{22}+\epsilon_2)^2+\epsilon_1^2}, \tag{15}$$

and

$$E_i=\frac{\epsilon_1(\gamma_{12}+\gamma_{22})^2}{(\gamma_{22}+\epsilon_2)^2+\epsilon_1^2}, \tag{16}$$

where $\gamma_{ij}=\rho_{ij}/\rho$, $\epsilon_1=(\gamma_{12}+\gamma_{22})(f_c/f)\,F_r(\kappa)=(\gamma_{12}+\gamma_{22})(f_c/f)\,F_r(\delta\sqrt{f/f_c})$, $\epsilon_2=(\gamma_{12}+\gamma_{22})(f_c/f)\,F_i(\kappa)=(\gamma_{12}+\gamma_{22})(f_c/f)\,F_i(\delta\sqrt{f/f_c})$. The function $F(\kappa)$ was written here more conveniently as a function of frequency f, i.e., $F(\kappa)=F(\delta\sqrt{f/f_c})$,[7] where δ is a factor dependent on pore geometry. For the hollow cylinder-like pores, $\delta=\sqrt{8}$ (Ref. 7) and we use this value throughout the paper. f_c is the critical frequency above which the Poiseuille flow breaks down, and it equals $b/(2\pi\rho_2)$. Here ρ_2 denotes the product of porosity and fluid mass density.

In order to obtain phase velocity and attenuation coefficient of the rotational waves, we put $l=\mathrm{Re}[l]+i\ \mathrm{Im}[l]$. Thus, the phase velocity is then $v_r=\chi/|\mathrm{Re}[l]|$. Introducing a reference velocity as $V_r=\sqrt{N/\rho}$, we obtain the dimensionless phase velocity as

$$\frac{v_r}{V_r}=\frac{\sqrt{2}}{[\sqrt{E_i^2+E_r^2}+E_r]^{1/2}}. \tag{17}$$

To obtain the attenuation coefficient of the rotational waves, we introduce a reference length, L_r, defined as $L_r=V_r/(2\pi f_c)$. The length x_a represents the distance over which the rotational wave amplitude is attenuated by a factor

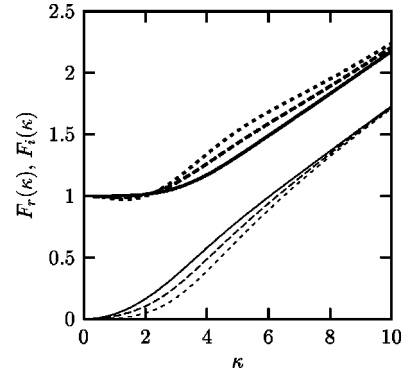


FIG. 2. Behavior of $F_r(\kappa)$ (thick curves) $F_i(\kappa)$ (thin curves) as function of frequency parameter, κ, according to Eq. (7). Solid curves correspond to the case when $\xi=0$ (no boundary slip), while long-dashed and short-dashed curves correspond to $\xi=0.05$ and $\xi=0.1$, respectively.

of $1/e$. Therefore we can construct the dimensionless attenuation coefficient as L_r/x_a,

$$\frac{L_r}{x_a}=\frac{f}{f_c}\frac{[\sqrt{E_i^2+E_r^2}-E_r]^{1/2}}{\sqrt{2}}. \tag{18}$$

The equations describing dynamics of the dilatational waves are[7]

$$\nabla^2(Pe+Q\epsilon)=\frac{\partial^2}{\partial t^2}(\rho_{11}e+\rho_{12}\epsilon)+bF(\kappa)\frac{\partial}{\partial t}(e-\epsilon), \tag{19}$$

$$\nabla^2(Qe+R\epsilon)=\frac{\partial^2}{\partial t^2}(\rho_{12}e+\rho_{22}\epsilon)-bF(\kappa)\frac{\partial}{\partial t}(e-\epsilon), \tag{20}$$

where P, Q, and R are the elastic coefficients, $e=\mathrm{div}\,\boldsymbol{u}$ and $\epsilon=\mathrm{div}\,\boldsymbol{U}$ are the divergence of solid and fluid displacements. Again, substitution of a plane dilatational wave of the form

$$e=C_1e^{i(lx+\chi t)},\quad \epsilon=C_2e^{i(lx+\chi t)}, \tag{21}$$

into Eqs. (19) and (20) allows us to obtain a characteristic equation

$$(z-z_1)(z-z_2)+iM(z-1)=0, \tag{22}$$

where $z=l^2V_c^2/\chi^2$, $V_c^2=(P+R+2Q)/\rho$ represents the velocity of a dilatational wave when the relative motion between fluid and solid is absent, $z_{1,2}=V_c^2/V_{1,2}^2$ with $V_{1,2}$ being the velocities of the purely elastic waves with subscripts 1,2 referring to the two roots of Eq. (22), and finally $M=(\epsilon_1+i\epsilon_2)/(\sigma_{11}\sigma_{22}-\sigma_{12}^2)$ with $\sigma_{11}=P/(P+R+2Q)$, $\sigma_{22}=R/(P+R+2Q)$, and $\sigma_{12}=Q/(P+R+2Q)$.

Equation (22) has two complex roots z_{I} and z_{II}. Phase velocities of the two kinds of dilatational waves can be defined as

$$\frac{v_{\mathrm{I}}}{V_c}=\frac{1}{\mathrm{Re}[\sqrt{z_{\mathrm{I}}}]}\quad \frac{v_{\mathrm{II}}}{V_c}=\frac{1}{\mathrm{Re}[\sqrt{z_{\mathrm{II}}}]}, \tag{23}$$

while the corresponding attenuation coefficients can be also introduced as

$$\frac{L_c}{x_{\mathrm{I}}}=\mathrm{Im}[\sqrt{z_{\mathrm{I}}}]\frac{f}{f_c},\quad \frac{L_c}{x_{\mathrm{II}}}=\mathrm{Im}[\sqrt{z_{\mathrm{II}}}]\frac{f}{f_c}. \tag{24}$$

III. NUMERICAL RESULTS

In order to investigate the novelties brought about into classical Biot's theory of propagation of elastic waves in porous medium[7] by the inclusion of boundary slip, we have studied the full parameter space of the problem.

In all forthcoming results, we calculate phase velocities and attenuation coefficients for case 1 from Table I taken from Ref. 7, which is $\sigma_{11}=0.610$, $\sigma_{22}=0.305$, $\sigma_{12}=0.043$, $\gamma_{11}=0.500$, $\gamma_{22}=0.500$, $\gamma_{12}=0$, $z_1=0.812$, and $z_2=1.674$.

We calculated normalized phase velocity of the plane rotational waves, v_r/V_r, and the attenuation coefficient L_r/x_a using our more general expression for $F(\kappa)$ (which takes into account nonzero boundary slip) given by Eq. (7).

In Fig. 3 we plot phase velocity v_r/V_r as a function of frequency for three cases: the solid curve corresponds to $\xi=0$ (no boundary slip), while long-dashed and short-dashed curves correspond to $\xi=1.0$ and $\xi=1.5$, respectively (we have used these large values in order to emphasize the effect of the boundary slip). Note that the $\xi=0$ case perfectly reproduces curve 1 in Fig. 5 from Ref. 7. For $\xi=1.0$ and 1.5 we notice a deviation from the classic behavior in the form of a decrease of phase velocity in the domain of intermediate frequencies.

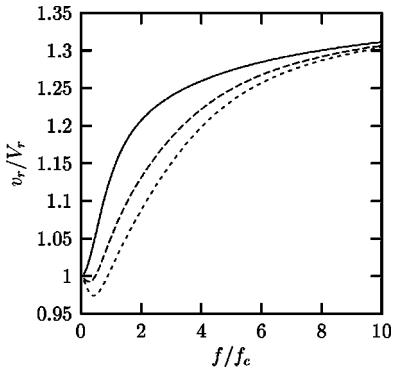


FIG. 3. Behavior of dimensionless, normalized phase velocity of the rotational wave, v_r/V_r, as a function of frequency. Solid curve corresponds to the case when $\xi=0$ (no boundary slip), while long-dashed and short-dashed curves correspond to $\xi=1.0$ and $\xi=1.5$, respectively.

Figure 4 shows the attenuation coefficient L_r/x_a of the rotational wave as a function of frequency for the three values of ξ: the solid curve corresponds to $\xi=0$ (no boundary slip), while long-dashed and short-dashed curves represent the cases $\xi=1.0$ and $\xi=1.5$, respectively. Note that the $\xi=0$ case coincides with curve 1 in Fig. 6 from Ref. 7. We observe that, in the intermediate frequency range, with increase of boundary slip our model yields substantially *higher* values of the attenuation coefficient than in the classic no-slip case, which is in qualitative agreement with experimental data.[8,9]

We also calculated normalized phase velocities of the plane dilatational waves, v_{I}/V_c and v_{II}/V_c, and the attenuation coefficients L_c/x_{I} and L_c/x_{II} using our more general expression for $F(\kappa)$ given by Eq. (7).

Figures 5 and 6 are similar to Fig. 3 except that now we plot phase velocities v_{I}/V_c and v_{II}/V_c as a function of frequency for $\xi=0, 0.05, 0.1$. Note that the solid curves on both

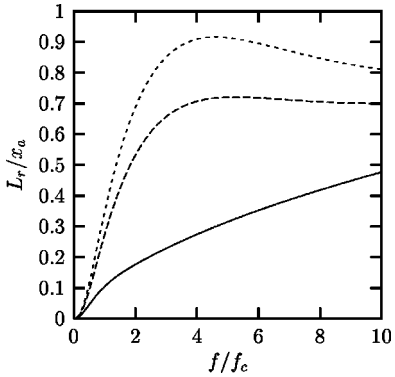


FIG. 4. Behavior of dimensionless, normalized attenuation coefficient of the rotational wave, L_r/x_a, as a function of frequency. Solid curve corresponds to the case when $\xi=0$ (no boundary slip), while long-dashed and short-dashed curves correspond to $\xi=1.0$ and $\xi=1.5$, respectively.

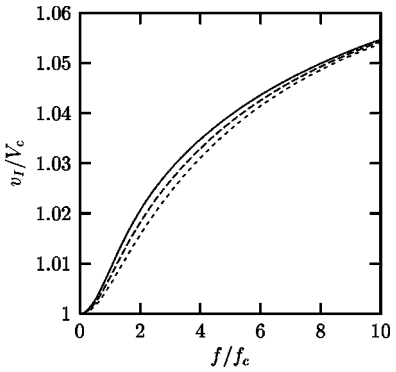


FIG. 5. Behavior of dimensionless, normalized phase velocity of the dilatational wave, v_1/V_c, as a function of frequency. Solid curve corresponds to the case when $\xi=0$ (no boundary slip), while long-dashed and short-dashed curves correspond to $\xi=0.05$ and $\xi=0.1$, respectively.

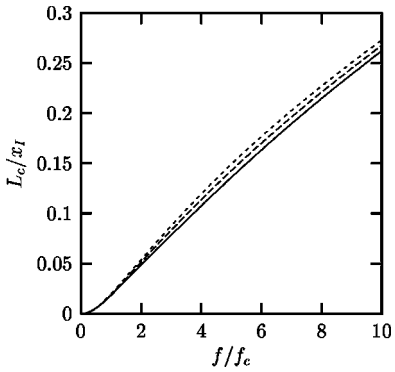


FIG. 7. Behavior of dimensionless, normalized attenuation coefficient of the dilatational wave, L_c/x_1, as a function of frequency. Solid curve corresponds to the case when $\xi=0$ (no boundary slip), while long-dashed and short-dashed curves correspond to $\xi=0.05$ and $\xi=0.1$, respectively.

graphs reproduce curves 1 in Figs. 11 and 12 of Ref. 7. We gather from the graph that introduction of the nonzero boundary slip leads to a decrease of v_I/V_c, while v_{II}/V_c increases in the domain of intermediate frequencies as boundary slip, ξ, increases.

In Figs. 7 and 8 we plot the attenuation coefficients L_c/x_I and L_c/x_{II} in a similar manner as in Fig. 4, but now for $\xi=0, 0.05, 0.1$. Again we observe that the $\xi=0$ case reproduces curve 1 in Figs. 13 and 14 of Ref. 7. We gather from Figs. 7 and 8 that introduction of the nonzero boundary slip yields *higher* than in the no boundary slip case values of the attenuation coefficients for both types of the dilatational waves in the domain of intermediate frequencies, which, again, is in qualitative agreement with experimental results.[8,9]

It is worthwhile to note that for nonzero ξ the asymptotic behavior of the elastic waves in the $f/f_c \to \infty$ as well as $f/f_c \to 0$ limit is identical to the classic behavior established by Biot.[7] This is a consequence of the assumptions of our phenomenological model, which is based on robust physical arguments.

IV. DISCUSSION

In this paper we have studied the effect of the introduction of boundary slip in the theory of propagation of elastic waves in a fluid-saturated porous solid originally formulated by Biot.[7] Biot's theory does not account for boundary slip effect, however, the boundary slip becomes important when the length scale over which the fluid velocity changes approaches the slip length, i.e., when the fluid is highly confined, for instance, fluid flow through porous rock or blood vessel capillaries. In light of recent convincing experimental evidence of a boundary slip in a Newtonian liquid,[2] it is necessary to take this effect into account in Biot's theory *where appropriate*. We have studied the effect of the introduction of boundary slip upon the function $F(\kappa)$ that measures the deviation from Poiseuille flow friction as a function of frequency parameter κ. This function crucially enters Biot's equations which describe dynamics of fluid-saturated porous solid. Therefore, a revision of Biot's theory was needed in order to incorporate boundary slip effect into all measurable predictions of this theory such as phase velocities and attenuation coefficients of both rotational and dilatational waves. We have performed such analysis, and in summary, we found that the introduction of the nonzero boundary slip into Biot's theory of propagation of the elastic waves in a fluid-saturated porous solid results in the following.

(1) A decrease, as compared to the no-slip limiting case, of the phase velocities of both rotational waves (v_r/V_r) and dilatational wave of the first kind (v_I/V_c) in the domain of intermediate frequencies. In contrast, the phase velocity of the dilatational wave of the second kind (v_{II}/V_c) experiences an increase as compared to the no-slip limiting case in the domain of intermediate frequencies.

(2) In the domain of intermediate frequencies the attenuation coefficients of both the rotational (L_r/x_a) and dilata-

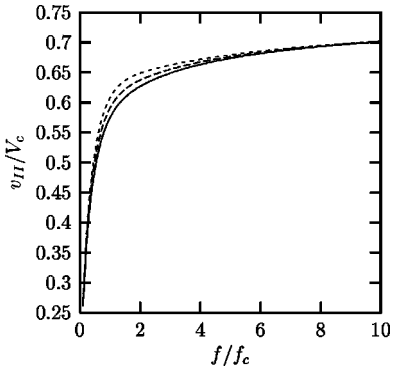


FIG. 6. The same as in Fig. 5 except that the curves are for v_{II}/V_c.

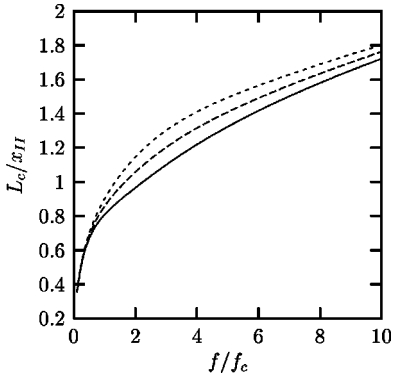


FIG. 8. The same as in Fig. 7 except that the curves are for L_c/x_{II}.

tional waves (L_c/x_{I} and L_c/x_{II}) are *increased* as compared to the no-slip limiting case as the boundary slip increases, which is in qualitative agreement with the experimental data.[8,9]

(3) The behavior of all physical quantities which describe the elastic waves in the asymptotic limits of both small and large frequencies is *not* affected by the introduction of the nonzero boundary slip. The deviation occurs only in the domain of intermediate frequencies, as prescribed by our phenomenological model.

The investigation of properties of elastic waves is important for a number of applications. The knowledge of phase velocities and attenuation coefficients of elastic waves is necessary, for example, to guide the oil-field exploration applications, acoustic stimulation of oil-producing fields (in order to increase the amount of recovered residual oil), and the acoustic clean-up of contaminated aquifers.[3–6] Therefore, our results would be useful for various applications in oil production as well as in ecology.

From the recent experimental results of Ref. 2 we gathered that there are physical situations where the no-slip boundary condition becomes invalid. We have formulated a *phenomenological* model of elastic waves in the fluid-saturated porous medium based on Biot's linear theory and certain physically justified assumptions on the variation of boundary slip velocity with frequency, $\overline{U_s}(\kappa)$. Since, so far there are no experimental measurements of $\overline{U_s}(\kappa)$, for a cylindrical tube, on which ''Biot-like'' theory relies, there is a certain freedom of choice, which could be used to obtain a better fit of experimental data with the theory *in cases where classic Biot's theory fails to do so*. In fact, our model predicts *higher than the Biot's theory values of attenuation coefficients of the both rotational and dilatational waves* in the intermediate frequency domain, which is in qualitative agreement with the experimental data.[8,9] Therefore, the introduction of the boundary slip yields threefold benefits:

(1) better agreement of theory with experimental data, since, the parametric space of the model is larger (includes effects of boundary slip),
(2) the possibility to identify types of porous medium and physical situations where boundary slip is important,
(3) constrain model parameters that are related to the boundary slip.

We would like to close this paper with the following two remarks: First, the reported slip length varies from tens of nanometers[1] to a few microns,[2] depending on the type of liquid and differences in the experimental setup. The latter experimental work clearly demonstrates the applicability of our model to the elastic wave phenomena in a usual porous media found in Nature. Second, in this work we *postulated* a phenomenological expression for the boundary slip velocity based on some physically sound arguments (in fact, a similar, successful, approach has been used by the authors of Ref. 15 in the context of solar physics). However, perhaps it may be possible to derive an expression based on more general ''first principles,'' which should be a subject of a future work.

[1] V. S. J. Craig, C. Neto, and D. R. M. Williams, ''Shear-dependent boundary slip in an aqueous Newtonian liquid,'' Phys. Rev. Lett. **87**, 054504 1-4 (2001).
[2] Y. Zhu and S. Granick, ''Rate-dependent slip of Newtonian liquid at smooth surfaces,'' Phys. Rev. Lett. **87**, 096105 (2001).
[3] I. A. Beresnev and P. A. Johnson, ''Elastic-wave stimulation of oil production: A review of methods and results,'' Geophysics **59**, 1000–1017 (1994).
[4] D. Tsiklauri and I. Beresnev, ''Non-Newtonian effects in the peristaltic flow of a Maxwell fluid,'' Phys. Rev. E **64**, 036303 (2001).
[5] D. Tsiklauri and I. Beresnev, ''Enhancement in the dynamic response of a viscoelastic fluid flowing through a longitudinally vibrating tube,'' Phys. Rev. E **63**, 046304 (2001).
[6] D. Tsiklauri and I. Beresnev, ''Properties of elastic waves in a non-Newtonian (Maxwell) fluid-saturated porous medium,'' Phys. Rev. E (submitted), preprint No. physics/0107078 (available at http://arxiv.org/abs/physics/0107078).
[7] 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).
[8] G. A. Gist, ''Fluid effects on velocity and attenuation in sandstones,'' J. Acoust. Soc. Am. **96**, 1158–1173 (1994).
[9] S. Mochizuki, ''Attenuation in partially saturated rocks,'' J. Geophys. Res. **87**, 8598–8604 (1982).
[10] M. A. Biot, ''Generalized theory of acoustic propagation in porous dissipative media,'' J. Acoust. Soc. Am. **34**, 1254–1264 (1962).
[11] W. F. Murphy, K. W. Winkler, and R. L. Kleinberg, ''Acoustic relaxation in sedimentary rocks: Dependence on grain contacts and fluid saturation,'' Geophysics **51**, 757–766 (1986).
[12] J. Dvorkin, R. Nolen-Hoeksema, and A. Nur, ''The squirt flow mechanism: Macroscopic description,'' Geophysics **59**, 428–438 (1994).
[13] J. M. Carcione, ''A 3-D time-domain wave equation for viscoacoustic saturated porous media,'' Eur. J. Mech. A/Solids **12**, 53–71 (1993).
[14] J. M. Carcione and G. Quiroga-Goode, ''Full frequency-range transient solution for compressional waves in a fluid-saturated viscoacoustic porous medium,'' Geophys. Prospect. **44**, 99–129 (1996).
[15] D. Tsiklauri and V. M. Nakariakov, ''Wide-spectrum slow magnetoacoustic waves in coronal loops,'' Astron. Astrophys. **379**, 1106–1112 (2001).

On the existence of subresonance generated in a one-dimensional chain of identical spheres

Anne-Christine Hladky-Hennion
IEMN, Department ISEN, 41 Boulevard Vauban, Lille Cedex 59046, France

Frédéric Cohen-Tenoudji
Laboratoire Environnement Développement, Université Paris 7, Case Courrier 7087, 2 Place Jussieu, 75251 Paris Cedex 05, France

Arnaud Devos
IEMN, Department ISEN, 41 Boulevard Vauban, Lille Cedex 59046, France

Michel de Billy
Laboratoire Environnement Développement, Université Paris 7, Case Courrier 7087, 2 Place Jussieu, 75251 Paris Cedex 05, France

(Received 15 September 2001; revised 5 May 2002; accepted 30 May 2002)

In this paper, the propagation of short compressional pulses through a one-dimensional chain of identical spherical beads is analyzed. First, a single sphere is studied. Then, an infinite chain of identical spheres is considered. Finally, a finite linear chain can be seen as a particular case of an infinite chain. It shows the existence of allowed and forbidden frequency bands. In the case of a finite chain of spheres, discrete resonance frequencies are excited and may be identified as ''subresonance'' of modes observed with a single sphere. We also analyze how the coupling between two adjacent spheres and the number of beads in the chain, modify the degeneracy of the stationary states. Finally, the theoretical results are qualitatively verified with experimental data obtained with the ''resonant sphere technique.'' © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1497369]

PACS numbers: 43.20.Ks [DEC]

I. INTRODUCTION

For many years the propagation of acoustic waves around a single sphere was investigated for applications in many domains: Geophysics, Medicine, Aerospace, etc.[1–7] These studies pointed out the existence of Rayleigh-type surface waves circumnavigating around the sphere. More recently, the excitation by short compressional pulses of a linear chain made up with identical spherical beads was considered[8] and experimentally confirms the existence of circumferential surface waves propagating from one sphere to the other through the contact point. The use of short ultrasonic pulses as excitation leads us to the generation of vibration modes (or resonances) which are experimentally revealed by the presence of peaks in the frequency spectra of the acoustic signals transmitted by the chain. In this study a modal analysis using the ATILA code[9] is carried out to identify the vibration modes and to determine their frequencies. The influence of the number of spheres in the finite chain as well as the coupling between the individual spheres on the ''frequency bands'' is studied. The results confirm the existence of allowed and forbidden frequency bands. This phenomenon is the occurrence of a general principle. It is well known in fact[10–13] that it is a behavior of coupled resonators regardless of the particular physical type of the resonators involved (i.e., mechanical, electrical, optical, etc.). It is demonstrated that, for weak coupling, there is a splitting of the modes which alter the frequency spectra and that the relative separation between two adjacent frequencies is proportional to the coupling stiffness.[14] In this study our first objective is to verify that a chain of spherical in contact or coupled beads may be treated by the theoretical formulation encountered with coupling oscillators in a linear one-dimensional periodic lattice or with independent particles moving in a one-dimensional periodic potential[15] and that the coupling spreads apart the natural frequencies of the individual sphere taken separately. The second objective is to evaluate how it is possible to characterize the quality and the intensity of the coupling in a finite chain made up of several beads; experimental results suggest that an effective contact radius may be used to describe the coupling factor between the beads. The next section deals with the physical model and numerical calculations. In Sec. III, experimental data obtained with the resonant scattering technique[16] are presented and compared with the numerical results.

II. PHYSICAL MODEL AND NUMERICAL CALCULATIONS

This section is devoted to the presentation of a physical model of three kinds of systems. First, we review the vibration properties of an elastic isolated sphere which are well known. Second, we examine the case of an infinite chain of identical spheres and discuss the collective modes which appear. Third, we discuss the finite size effects on the chain vibrations. In each case, the physical model helps us in understanding the numerical results which have been obtained with the ATILA finite element code.[9] This code has been specifically developed to aid in the design of sonar transduc-

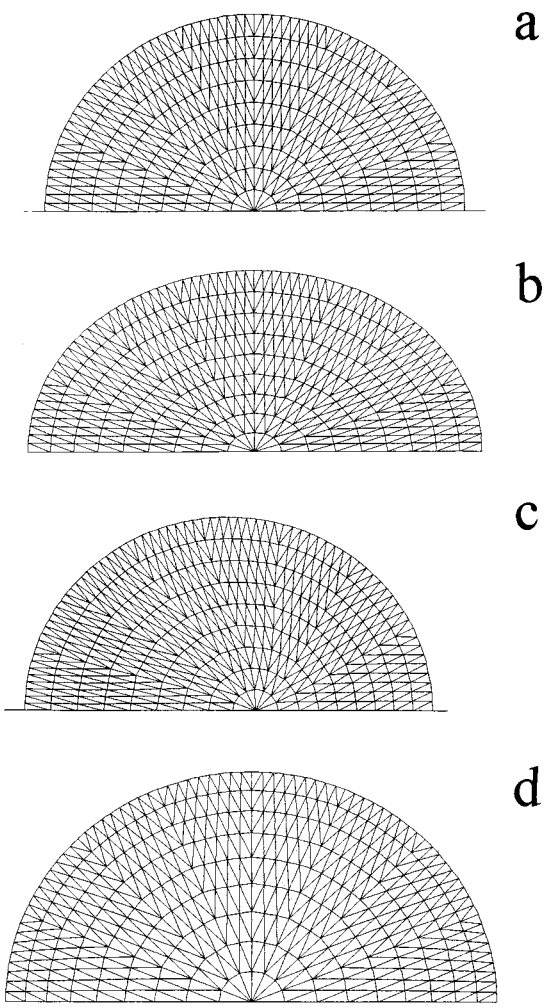


FIG. 1. Numerical displacement fields for a spherical bead of 8 mm in diameter: (a) Rest position; (b) mode $R(2,1)$; (c) mode WG(2,2); (d) breathing mode.

ers, but can also be used for all types of transducers or passive structures. In this study, the modal analysis is carried out to identify the vibration modes and to determine their resonance frequencies.

A. Case of a single sphere

The natural vibration modes (or resonance) of an *individual* vibrating sphere are of three types: first, the Rayleigh modes $R(n>1,1)$ where a wave is propagating around the sphere whereas the center of the sphere does not move; second, the whispering gallery modes $[\mathrm{WG}(n>0,l=2,3,4,...)]$ where the surface of the sphere does not move whereas a vibration is observed in the heart of the sphere; and finally, the well-known breathing modes, corresponding to a radial vibration of the sphere. We calculate numerically the vibrations of a steel sphere the diameter of which is 8 mm and characteristics are $\rho=7800$ kg/m^3, $C_L=5900$ m/s, and $C_T=3236$ m/s. The mesh used for the computation is presented in Fig. 1(a). The x axis is an axisymmetrical axis. The values of the resonance frequencies (f_0) obtained with a single sphere in air are given in the second column of Table I. With the help of the computed displacement field, each resonance frequency is classified in one of the three above types [Figs. 1(b), (c), and (d)]. In the third column of Table I, the experimental values of the identified modes are reported. It indicates that the Rayleigh series only is experimentally detected (see Sec. III). There is a good agreement between experimental and numerical values.

TABLE I. Comparison between the theoretical and experimental values of the resonance frequencies (in kHz) for a 8 mm diameter steel sphere in air.

MODE	$(f_R^{\mathrm{th}})^{\mathrm{a}}$	$(f_R^{\mathrm{exp}})^{\mathrm{b}}$
$R(2,1)$	341	340
WG(1,2)	458	
$R(3,1)$	508	502
WG(2,2)	649	
$R(4,1)$	652	644
Breathing mode	662	
$R(5,1)$	786	772
WG(3,2)	856	
$R(6,1)$	916	904
WG(1,3)	923	
$R(7,1)$	1044	1024
WG(1,4)	1057	
WG(4,2)	1064	
WG(2,3)	1105	
$R(8,1)$	1169	1144
WG(5,2)	1265	
WG(3,3)	1280	
$R(9,1)$	1295	1272
WG(2,4)	1360	

[a]Theory ($C_T=3236$ m/s and $C_L=5900$ m/s).
[b]Experiments.

B. Case of an infinite chain of identical spheres

Consider now an infinite chain of identical spheres of radius a, regularly spaced by a distance $2a$. The coupling between spheres is taken into account by introducing a contact area (r_c) described in Fig. 2. Due to the periodicity of the system, the wave number k is introduced for designating the vibration modes. Furthermore, one knows that k should belong to the first Brillouin zone which is given by $[-\pi/2a,\pi/2a]$. For each k value, a vibration frequency ω exists that leads to the notion of a dispersion curve $\omega(k)$.

Two kinds of collective vibrations would result from the association of spheres in a linear chain. First, each sphere can be seen as a rigid ''atom'' which can be translated in respect to its neighbors. One retrieves a system of mass coupled by springs and limiting the interaction between spheres to the first neighbors, one gets the well-known dispersion relationship

$$\omega(k)=(4K/m)^{1/2}\sin ka, \tag{1}$$

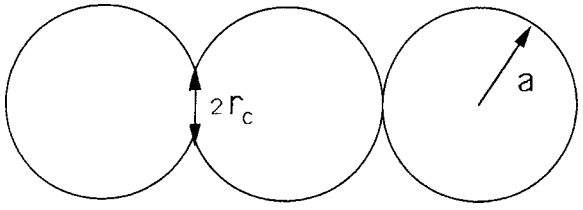


FIG. 2. Definition of the contact area (r_c) between two adjacent spheres.

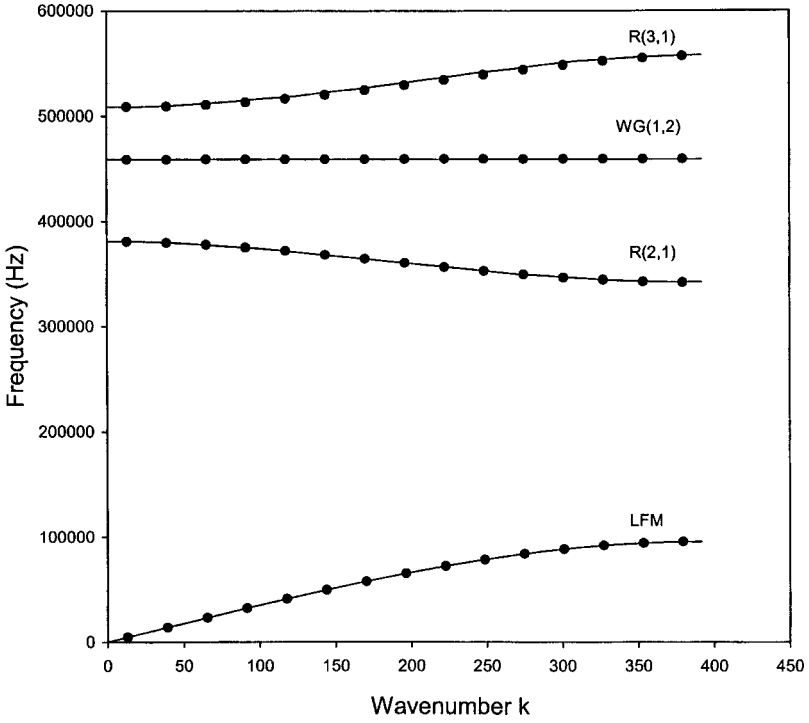


FIG. 3. Dispersion curve of a chain of steel spheres by FEM method. Full line, infinite chain; marks, finite chain of 15 spheres.

where K designates the coupling stiffness and m the mass of an individual sphere. These modes are low frequency modes due to the nature of the displacement of the whole sphere around its equilibrium position. If the sphere was perfectly stiff, these modes would be the only permitted modes.

A second kind of collective vibrations appears since each sphere is able to vibrate: due to the coupling between spheres, each sphere vibration generates a collective mode associated with a dispersion curve, that will be later designated by the mode of the sphere vibration (Rayleigh, whispering gallery, breathing). In the case of a weak coupling between spheres, the frequencies of collective vibration are expected to be close to the frequency of the corresponding mode of a single sphere.

The numerical propagation of surface waves in the periodic structure is studied using only the mesh of one unit cell, thanks to Bloch–Floquet relations. It provides dispersion curves from which results of physical interest can be easily extracted: identification of propagation modes, cutoff frequencies, passbands, and stopbands.[17] Figure 3 presents the dispersion curves obtained for the first four bands with r_c equal to 0.6 mm. The first band is a low frequency band (LF) and goes through the origin. It may be designated as the ''acoustic branch'' and is well described by Eq. (1) using $K/m = 9\times10^{10}\,\mathrm{N\,m^{-1}\,kg^{-1}}$. The other bands, which may be designated as ''optical branches,'' correspond to the second kind of collective modes which means they are related to a vibration of an isolated sphere. The corresponding vibrations are identified using the displacement field of the spheres and are, respectively, $R(2,1)$, WG(1,2), and $R(3,1)$. For each branch, the lowest frequency is very close to the value obtained for the single sphere (Table I).

The shape of each branch depends on the nature of the sphere vibration. It measures the coupling between spheres which vibrate in the same mode. In the case of WG modes, the displacement is localized inside the sphere so that weak coupling exists between spheres. The corresponding dispersion curve is a flat band. Concerning the Rayleigh modes, the coupling depends on the surface displacement. For a collective mode with $k=0$ all the spheres are vibrating in phase whereas with $k=\pi/2a$ two adjacent spheres are out of phase. As can be seen in Fig. 1(b), for the $R(2,1)$ vibration, the collective mode with $k=0$ needs more energy than the $k=\pi/2a$ mode in which the coupling is so small that one retrieves the isolated sphere frequency. It explains the decreasing of the dispersion curve. In the $R(3,1)$ case the situation is reversed due to the nature of the displacement of the sphere.

In Fig. 4 are plotted the dispersion curves obtained for two values of r_c: 0.6 and 0.14 mm. The curves are limited to the $R(2,1)$ mode. It is observed that the higher the coupling, the larger the width of the pass band.

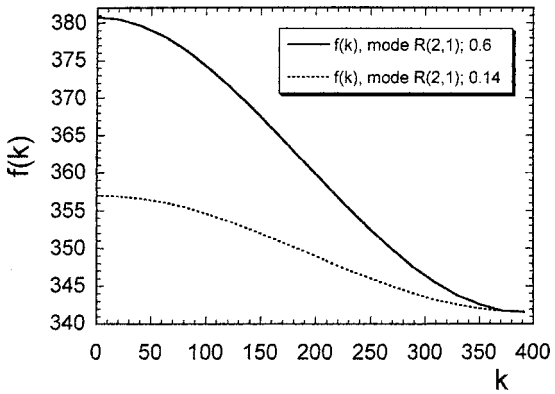


FIG. 4. Dispersion curves obtained for two values of r_c: case of the Rayleigh mode $R(2,1)$.

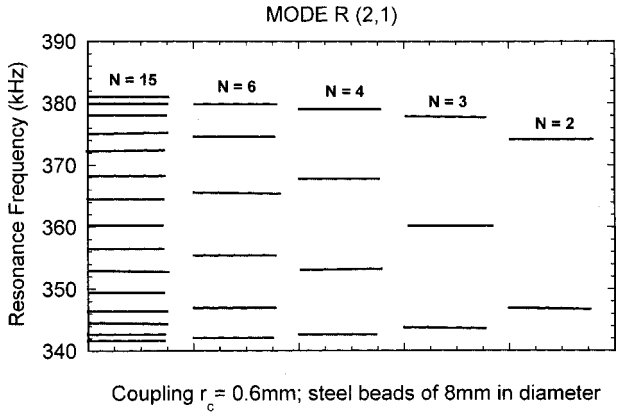


FIG. 5. Theoretical resonance frequency diagrams associated to the mode $R(2,1)$ for different values of $N(=2, 3, 4, 6,$ and $15)$. The coupling value is $r_c=0.6$ mm.

C. Case of a finite chain

A *finite linear chain* can be seen as a particular case of an infinite chain by applying the appropriate boundary conditions. Then, it implies a quantification of the wave number k (Ref. 15) and the corresponding values of the frequency of each dispersion curve transforms itself in a discrete number of vibrations of the whole system. Considering N particles with one end sphere fixed and the other one free—which simulate the experimental conditions—then the appropriate boundary conditions require

$$u(x=0)=0 \quad \text{and} \quad du(x=2Na)/dx=0. \tag{2}$$

If the displacement is written as

$$u(x)=A(x)\cos\omega t=(A\sin kx+B\cos kx)\cos\omega t, \tag{3}$$

relation (2) implies

$$B=0 \quad \text{and} \quad k=(2s+1)\pi/2N2a=(2s+1)\pi/2L, \tag{4}$$

where L represents the total length of the chain $(s=0,1,2,...,N-1)$.

For a finite chain of N spheres, we obtain N discrete frequencies for each branch (acoustical and optical branches).

The finite element method has been used to compute the vibration frequencies of a chain of N identical spheres ($N=2$, 3, 4, 6, and 15). As expected, N modes are obtained for each branch. In Fig. 5 are reported the numerical resonance frequency diagrams associated to the mode $R(2,1)$ for different values of $N(=2,3,4,6,$ and $15)-(r_c=0.6\,\text{mm})$. In Fig. 3, the discrete frequencies of a chain of 15 spheres are plotted, using the quantification of the wave number [Eq. (4)]. The discrete points come on the continuous branches, this validating our description. Figure 6 presents the bandwidth

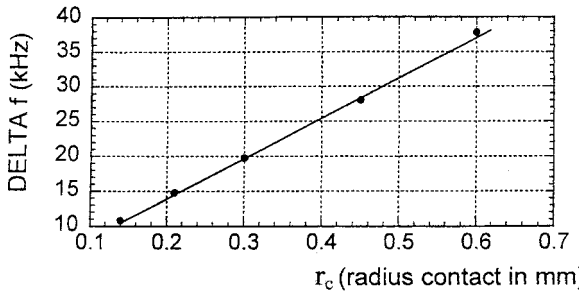


FIG. 6. Variations of the bandwidth of the $R(2,1)$ branch as a function of the radius contact $r_c(=0.6,\ 0.3,\ 0.21,$ and 0.14 mm). $N=6$.

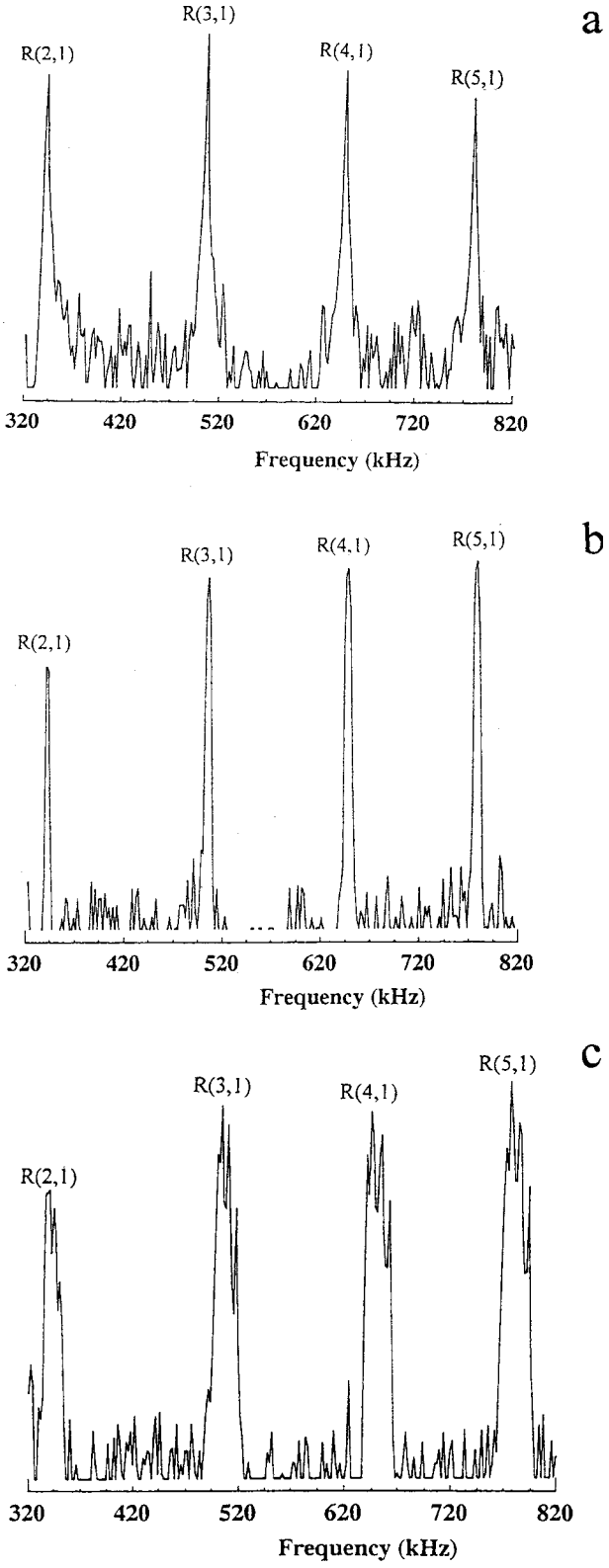


FIG. 7. Experimental frequency spectra obtained, respectively, with (a) a single steel sphere (diam=8 mm), (b) an array made up with four free steel spheres, (c) an array made up with four glued spheres.

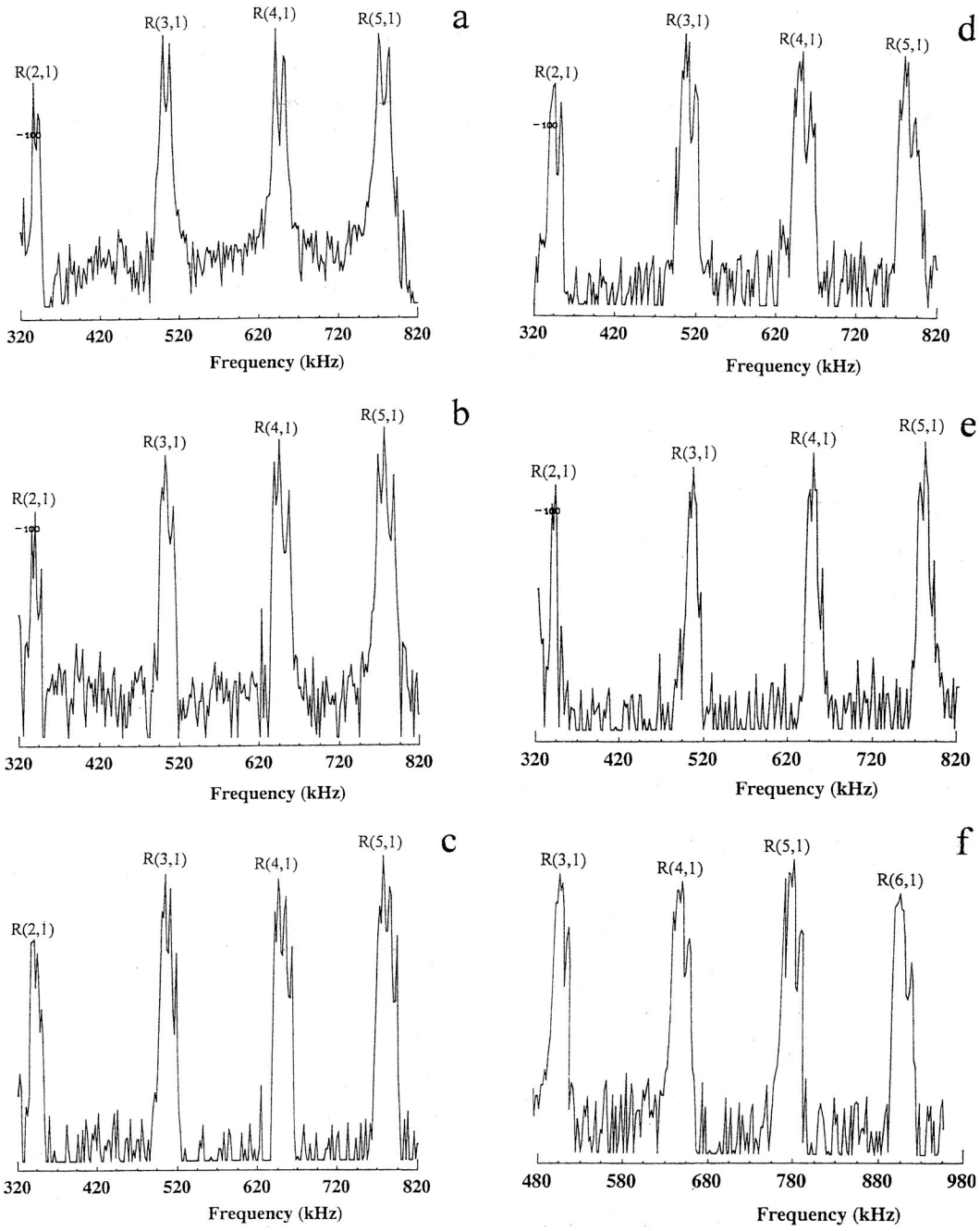


FIG. 8. Experimental frequency spectra obtained, respectively, with (a) two glued beads, (b) three glued beads, (c) four glued beads, (d) five glued beads, (e) six glued beads, and (f) 13 glued beads.

of the $R(2,1)$ branch as a function of the contact radius r_c. As the contact radius increases, coupling between the spheres also increases, and with it the bandwidth of nonflat branches.

III. EXPERIMENTAL OBSERVATIONS

Experiments have been performed on a single steel sphere and on finite chains of elastic spherical beads of 8 mm in diameter. The sphere contact technique[16] was used for this study. It includes a pair of broadband contact transmitters where the emitter is excited with short longitudinally polarized pulses. An axial force F_0 is applied to the transmitters and can be continuously varied from a few to about 60 Newton without damage to the transmitters. In the case of a chain of beads in contact, a cardboard cylinder is used to keep aligned the spheres which make up the chain. The acoustic signal detected by the receiver is displayed on the screen of a digital oscilloscope and its frequency spectrum is obtained via a discrete Fourier transform. The sampling frequency is 5 MHz and the frequency resolution is 2 kHz. The frequency position of the peaks observed in the spectra correspond to the resonance frequencies of the sample or to the excitation of a mode.

A. Case of a single steel sphere

The frequency spectrum recorded for a single steel sphere of 8 mm in diameter and submitted to a constant axial

pressure F_0 (<40 N) is given in Fig. 7(a). The frequency positions of the maxima in the frequency spectrum are given in the third column of the Table I and compared with the calculated resonance frequencies of the excited modes. We notice that only the modes of the Rayleigh series are detected. A good agreement is observed between the experimental and the numerical values. The eight Rayleigh modes are then identified from $R(2,1)$ up to $R(9,1)$.

B. Case of a finite chain

The purpose of this experimental analysis is to verify qualitatively the numerical results obtained in Sec. II C and more specifically to check the influence of the number of beads in the finite chain and of the coupling on the frequency response. The modification of the coupling conditions were realized by gluing the spheres over the contact points.

In Figs. 7(b) and (c) are given the frequency spectra observed with a finite chain ($N=4$) in different coupling conditions. Figure 7(b) is concerned with an array of free beads (no glue); Fig. 7(c) presents the case of a chain each bead of which is glued to its nearest neighbors. The following features characterize these experimental plots:

(1) The number of resonance frequencies observed with a single sphere is not affected by the number of balls which formed the linear chain.
(2) The more the number of glued spheres increases, the more the structure of the resonance is modified: when the beads are free [Fig. 7(b)], the resonance is ''smooth'' by comparison with the sharpness of the resonance which characterizes the peaks obtained for a single sphere [Fig. 7(a)] and it does not point out yet a structure. As soon as the spheres are artificially coupled (or glued), a succession of peaks appears in the resonance spectrum and the number of maximums in each resonance is equal to the number of beads in the chain [Fig. 7(c)] as it was predicted by the theory.
(3) An accurate analysis would confirm that the frequency of the lowest peak in each resonance coincides with the resonance frequency of the associated Rayleigh mode obtained for a single sphere as it was suggested by the numerical calculations.

Figures 8(a), (b), (c), (d), (e), and (f) represent, respectively, the experimental frequency spectra obtained with a finite chain made up of $N(=2, 3, 4, 5, 6,$ and 13) glued steel spheres (diameter$=8$ mm). These spectra are very interesting because they confirm that, with our experimental conditions, and up to $N=5$, the number of ''subresonance'' in each peak is equal to the number of beads. We notice also that for $N=6$ and 13 [Figs. 8(e) and (f)] the resonance does not reveal the existence of 6 and 13 peaks. This is due to the fact that the difference between two adjacent frequencies in the dispersion curve becomes small and the frequency resolution is not sufficient: then the peaks cannot be separated. A similar observation was mentioned in Ref. 18 dealing with two very close immersed cylindrical shells. The authors mention also a splitting of the peaks in the frequency spectra, but it appears to be due to interferential phenomenon which is different from the one studied in the present paper.

IV. CONCLUSION

This study demonstrates numerically that the propagation of short acoustic pulses in a chain made up of spherical particles points out in the frequency spectra the existence of ''allowed and forbidden frequency bands'' when the conditions for the excitation of Rayleigh waves are satisfied. Such bands are observed for each stationary state and their structure has been analyzed as a function of different parameters such as the number of the balls in the chain and the coupling between two adjacent spheres. Experiments conducted on short limited chains made up of steel spheres confirmed qualitatively the ''subresonance'' structure of the Rayleigh modes numerically predicted. It was shown also that from experimental measurements of the width of the frequency band it should be possible to get an estimation of the average coupling between the balls of the chain. This study must be continued:

(i) to check the validity of the results for higher Rayleigh modes;
(ii) to consider the results obtained with other materials;
(iii) to imagine how to get quantitative information on the coupling between the spheres regularly spaced in a chain.

[1] K. Sezawa, Bull. Earthquake Res. Inst., Univ. Tokyo **2**, 21 (1927).
[2] W. G. Neubauer, R. H. Vogt, and L. R. Dragonette, ''Acoustic reflection from elastic spheres. Steady state signals,'' J. Acoust. Soc. Am. **55**, 1123–1129 (1974).
[3] D. Royer, E. Dieulesaint, X. Jia, and Y. Shin, ''Optical generation and detection of surface acoustic waves on a sphere,'' Appl. Phys. Lett. **52**, 706–708 (1988).
[4] H. Überall, L. R. Dragonette, and L. Flax, ''Relation between creeping waves and normal modes of vibration of a curved body,'' J. Acoust. Soc. Am. **61**, 711–715 (1977).
[5] R. H. Vogt and W. G. Neubauer, ''Relationship between acoustic reflection and vibrational modes of elastic spheres,'' J. Acoust. Soc. Am. **60**, 15–22 (1976).
[6] Y. Sato and T. Usami, ''Basic study on the oscillation of a homogeneous elastic sphere. I. Frequency of the free oscillations,'' Geophys. Mag. **31**, 15 (1962).
[7] N. Gespa, La diffusion Acoustique, Cedocar Editor, 1987.
[8] M. de Billy, ''Experimental study of sound propagation in a chain of spherical beads,'' J. Acoust. Soc. Am. **108**, 1486–1495 (2000).
[9] ATILA Finite Element Code for Piezoelectric and Magnetostrictive Transducer Modeling, Version 5.1.1, User's Manual, ISEN, Acoustics Laboratory, Lille, France, 1997.
[10] C. Cohen-Tannoudji, B. Diu, and F. Laloë, Mécanique Quantique, Tome II, 1973.
[11] L. Brillouin and M. Parodi, Propagation dans les milieux périodiques, Masson et Cie, Dunod Editeurs, 1956.
[12] Y. S. Kivshar, ''Localized modes in a chain with non-linear on site potential,'' Phys. Lett. A **173**, 172–178 (1993).
[13] F. S. Crawford, Jr., *Waves*, Berkeley Physics Course, Vol. 3 (McGraw-Hill, New York, 1968).
[14] P. M. Morse, *Vibration and Sound* (McGraw-Hill, New York, 1948), Sec. 7, pp. 52–65.
[15] R. A. Smith, *Wave Mechanics of Crystalline Solids* (Chapman and Hall, London, 1963).
[16] N. Soga and O. L. Anderson, ''Elastic properties of Tektites measured by resonant sphere technique,'' J. Geophys. Res. **72**, 1733–1739 (1967).
[17] P. Langlet, A. C. Hladky-Hennion, and J. N. Decarpigny, ''Analysis of the propagation of plane acoustic waves in passive periodic materials using the finite element method,'' J. Acoust. Soc. Am. **98**, 2792–2800 (1995).
[18] S. Lethuillier, P. Pareige, J. M. Conoir, and J. L. Izbicki, ''Scattering by two very close immersed shells: numerical results,'' IEEE Int. Ultra. Symp. 1999 (Lake Tahoe, Nevada, 1999), pp. 731–734.

A measurement method of the flow rate in a pipe using a microphone array

Yong-Beum Kim[a)]
Mechanical and Material Department, Korea Institute of Nuclear Safety (KINS), P.O. Box 114, Yusong, Taejon 305-600, Republic of Korea

Yang-Hann Kim[b)]
Center for Noise and Vibration Control (NOVIC), Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Science Town, Taejon 305-701, Republic of Korea

(Received 11 December 2000; revised 3 January 2002; accepted 3 June 2002)

A method of measuring the flow rate in a pipe is proposed. The method utilizes one-dimensional acoustic pressure signals that are generated by a loud speaker. A microphone array mounted flush with the inner pipe wall is used to measure the signals. A formula for the flow rate, which is a function of the change of wave number, is derived from a simple mathematical model of sound field in the pipe conveying a viscous fluid. The change of the wave number, which is one of the results caused by flow, is estimated from the recursive relation among the measured microphone array signals. Since measurement errors, due to extraneous measurement noise and mismatch of response characteristics between microphones, exist in the estimated flow rate, a method of compensating the errors is proposed. By using this measurement method, the flow rate can be obtained more accurately than that of our previous method. To verify applicability of the measurement method, numerical simulation and experiments are performed. The estimated flow rates are within 5% error bound. [DOI: 10.1121/1.1496764]

PACS numbers: 43.20.Mv, 43.20.Ye, 43.60.Qv [ANN]

I. INTRODUCTION

In order to measure the flow rate in a pipe, various types of flowmeter such as Pitot tube, orifice meter, turbine meter, and thermal flowmeter are often used in industrial fields and laboratories.[1] However the flowmeters are inserted directly into the pipe, deforming the shape of flow path and causing pressure to drop because of flow disturbance.

There is a type of flowmeter that does not come in contact with the flow, not disturbing the flow. It utilizes an ultrasonic beam, which is bent and delayed in an uncontrollable and unpredictable way by the flow inhomogeneities, across the pipe's cross section, that always occur in practice.[2,3] The time delay is indistinguishable from that caused by the mean velocity to be measured because the flow measurement is based on the time delay; the inhomogeneities lead to inaccuracy. It is noteworthy that long acoustic waves, long compared with a pipe's diameter, are not sensitive to the flow inhomogeneities.[3] Accordingly, a method of measuring the flow rate by using the long waves can avoid the inaccuracy.

From our previous work, Kim and Kim[4] suggested a method of measuring the flow rate by using axial bending waves along the pipe measured by three accelerometers which were located in the axial direction of the pipe. However, incorrect measurement results are unavoidable from using only three sensors, for cases when the spatial distribution of the axial bending waves is symmetric with respect to a central sensor position.

Cheung *et al.*[5] recently suggested a long wavelength acoustic flowmeter, using microphone array installed in the pipe wall. Mach number-dependent rotation angle associated with the phase change was exploited to estimate the mean flow velocity in pipes. The measurement results show only the examples using the specific pure tone, so it is not clear whether the measurement method is valid for the acoustic plane waves of the other frequencies; it is required to propose the frequency range in which the measurement method is applicable. Also, when the signal-to-noise ratio is not large, errors from extraneous measurement noise are unavoidable.

Proposed in this paper is a method to measure the flow rate using acoustic pressures of much longer wavelength than a cutoff wave of the pipe. The acoustic pressures are measured by a microphone array, mounted flush with an inner pipe wall, which minimizes development of the flow disturbance. The microphone array consists of more than three sensors, so it can avoid the incorrect measurement results of Kim and Kim's method. Based on fundamental duct acoustics, the moving fluid in the pipe causes a change of wave number,[6] so the flow rate is obtained using a relationship between the flow rate and the change of the wave number. A recursive relation among the microphone array signals is utilized to measure the change of the wave number. A method is suggested to compensate possible errors associated with the measurement system, errors due to extraneous measurement noise and mismatch of response characteristics among array sensors. This measurement method enables us to obtain the

[a)]Electronic mail: ybkim@kins.re.kr
[b)]Electronic mail: yhkim@mail.kaist.ac.kr

flow rate in the frequency range excluding the singular frequency associated with sensor spacing. By using the measurement method, the flow rate can be measured with 5% error, which is more accurate than our previous method of 12% error.

II. FORMULATION OF FLOW RATE

The moving fluid in the pipe causes the change of sound wave number in the axial direction of the pipe. According to flow direction, the wave number increases or decreases in comparison with the case of the stationary fluid.[6] It is inferred from this physical behavior that a certain relationship between the flow rate and the change of the wave number exists.

In this study, it is assumed that the pipe has a uniform cross section with rigid-wall boundary condition and a subsonic flow with viscosity but without transverse and axial temperature gradients. The presence of viscosity brings into play a coupling between the axial and radial motions of the fluid particle in the pipe.[7] Even if axisymmetry is assumed, the wave propagation in the pipe would be two-dimensional. The acoustic pressure, however, can be regarded to be independent of the radius of the pipe for most of the common gases and liquids for which the following inequality is satisfied.[7]

$$\frac{\mu\omega}{\rho_0 c_0^2} \ll 1, \tag{1}$$

where μ is the viscosity of the medium, ω the angular frequency, ρ_0 the density of the medium, and c_0 the sound speed in the free medium. Equation (1) implies that viscous effect is small for the sound with comparatively long wavelength; the low frequency region satisfies $f \ll \rho_0 c_0^2/2\pi\mu$.

Also, in the range $k\delta_a \ll 1$, where sound wavelength is larger than the aerodynamic boundary-layer thickness, δ_a, refraction by flow gradient becomes quite negligible; thus, the flow velocity profile can be regarded to be uniform when estimating sound attenuation.[8,9] Here, k is the wave number and the value of δ_a can be considered to be equal to the radius of the pipe when the flow is a fully developed turbulent flow.[9]

Therefore, the only one-dimensional plane wave satisfying the above-mentioned conditions is dealt with; the acoustic pressure and the flow velocity profile may be regarded as uniform over the cross section of the pipe when estimating acoustic properties.

The sound field is considered in the presence of mean flow in a pipe as shown in Fig. 1. An acoustic source exists at one end of the pipe and a passive termination at the opposite end. The sound field for a plane wave propagating in the pipe can be expressed as[10,11]

$$P(x,f) = P^+(f)\exp(-j\gamma^+ x) + P^-(f)\exp(j\gamma^- x), \tag{2}$$

where $P(x,f)$ is the frequency spectrum of the measured acoustic pressure at position x, and $P^+(f)$ and $P^-(f)$ are the spectra of acoustic pressure in the downstream and upstream directions of the flow, respectively. Here γ^+ and γ^- denote the propagation constants for the acoustic wave in the downstream and upstream directions of the flow, respectively, defined as

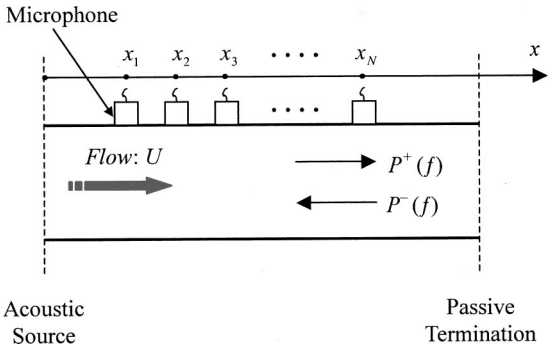


FIG. 1. Sound field for a plane wave propagating in the pipe with mean flow velocity U [x_N: position of the Nth microphone, $P^+(f)$ and $P^-(f)$: the spectra of acoustic pressure in the downstream and upstream directions of the flow, respectively].

$$\gamma^\pm = k^\pm - j\delta^\pm, \tag{3a}$$

where

$$k^\pm = k/(1 \pm M), \tag{3b}$$

$$k = k_0 + \delta, \tag{3c}$$

$$k_0 = \omega/c_0, \tag{3d}$$

$$\delta^\pm = \delta/(1 \pm M), \tag{3e}$$

$$\delta = \delta_0 + \xi M, \tag{3f}$$

$$\delta_0 = \frac{2}{c_0 d}\left(\frac{\omega\mu}{2\rho_0}\right)^{1/2}. \tag{3g}$$

Here k is the wave number, k_0 the free medium wave number, δ the attenuation constant, δ_0 the stationary medium viscothermal attenuation constant, and M the axial mean flow Mach number. ξ is equal to $F/2d$ where F is Froude's friction factor and d is the diameter of the pipe.

To obtain a relationship between the flow rate and the change of the wave number, the difference between the wave numbers in the downstream and upstream directions, Δk is defined as

$$\Delta k = \frac{k^- - k^+}{2} = \mathrm{Re}\left\{\frac{\gamma^- - \gamma^+}{2}\right\} = \frac{M}{1-M^2}k, \tag{4}$$

where $\mathrm{Re}\{\cdot\}$ represents the real part of a complex number.

Since Eq. (4) can be written as a second-order equation of M, the volume flow rate, Q, in the pipe can be expressed except for the case of stationary flow ($\Delta k = 0$) as

$$Q = \frac{-k + \sqrt{k^2 + 4(\Delta k)^2}}{2\Delta k}\frac{\pi d^2}{4}c_0. \tag{5}$$

It is worth noting from Eq. (5) that the flow rate in the pipe can be measured if information about the wave number, k, and the difference between the wave numbers in the downstream and upstream directions, Δk, are known.

III. METHOD FOR MEASURING FLOW RATE

The wave number, k, and the difference between the wave numbers, Δk, in Eq. (5) can be found by measuring the acoustic pressure in the pipe. In this study, the acoustic pressures are measured by using a microphone array located in the axial direction of the pipe as shown in Fig. 1. Use of the microphone array enables us to obtain the spatially averaged acoustic properties. Spacing of the microphone array is adjusted to be equidistant because the equidistant position of the array microphone yields the smallest error when estimating the acoustic properties.[12]

Since the propagation constants γ^+ and γ^- in Eq. (2) provide information about the wave number according to the flow rate, it is necessary to know the relationship between the propagation constants, γ^+, γ^-, and the measured acoustic pressures.

An expression for the acoustic pressure measured at one location has four unknown variables, i.e., γ^+, γ^-, $P^+(f)$, and $P^-(f)$ from Eq. (2). Accordingly, four expressions for the acoustic pressures at four locations are required to obtain solutions for the four unknowns. In order to find the relationship between the propagation constants, γ^+, γ^-, and the measured acoustic pressures, first, the terms of $P^+(f)$ and $P^-(f)$ are eliminated from the four expressions, leaving two equations with the γ^+, γ^- terms. The equations are, respectively, obtained by using three expressions for the acoustic pressures at any three of the four locations because the three expressions with four unknowns can be rearranged as one equation by eliminating $P^+(f)$ and $P^-(f)$; however, in this paper, each equation is obtained by using three expressions at three consecutive locations for the purpose of a systematic approach. An equation is obtained from the three expressions at three consecutive locations x_n, x_{n+1}, and x_{n+2} as

$$P_n(x,f)\exp\{j(\gamma^- - \gamma^+)\Delta x\} + P_{n+2}(x,f) = P_{n+1}(x,f)\{\exp(j\gamma^-\Delta x) + \exp(-j\gamma^+\Delta x)\}, \tag{6}$$

where $P_n(x,f)$ is the frequency spectrum of the measured acoustic pressure at position x_n and Δx is the microphone spacing. The detailed derivation of Eq. (6) is provided in Appendix A. Equation (6) represents a recursive relation between the propagation constants γ^+, γ^- and the acoustic pressures at three consecutive locations. Another recursive relation can be obtained systematically from Eq. (6) when three consecutive measurement points are shifted by one point. The two relations can be arranged in the form of linear equations as

$$\begin{bmatrix} S_{n+3,n}(f) & -S_{n+3,n+1}(f) \\ S_{n,n+1}(f) & -S_{n,n+2}(f) \end{bmatrix} \begin{Bmatrix} G_1(f) \\ G_2(f) \end{Bmatrix} = \begin{Bmatrix} -S_{n+3,n+2}(f) \\ -S_{n,n+3}(f) \end{Bmatrix}, \tag{7a}$$

where

$$G_1(f) = \exp[j(\gamma^- - \gamma^+)\Delta x], \tag{7b}$$

$$G_2(f) = \exp(j\gamma^-\Delta x) + \exp(-j\gamma^+\Delta x). \tag{7c}$$

Here, $S_{n,n+1}(f)$ denotes the cross-spectral density between $P_n(x,f)$ and $P_{n+1}(x,f)$. The detailed derivation of Eq. (7) is provided in Appendix A. Equation (7) represents the recursive relation between the propagation constants γ^+, γ^- and the acoustic pressures measured at four consecutive locations. It is noteworthy that use of the cross-spectral density in Eq. (7) enables us to exclude errors due to an uncorrelated measurement noise.[13] Appendix B shows how the effects of the measurement noise are excluded.

From the recursive relation of Eq. (7), if the acoustic pressures are measured at N locations, G_1 and G_2 can be expressed, in the form of the spatially averaged values, as

$$G_1 = \frac{1}{N-3}\sum_{n=1}^{N-3} \frac{S_{n+3,n+2}S_{n,n+2} - S_{n+3,n+1}S_{n,n+3}}{S_{n+3,n+1}S_{n,n+1} - S_{n+3,n}S_{n,n+2}}, \quad N \geqslant 4, \tag{8}$$

$$G_2 = \frac{1}{N-3}\sum_{n=1}^{N-3} \frac{S_{n+3,n+2}S_{n,n+1} - S_{n+3,n}S_{n,n+3}}{S_{n+3,n+1}S_{n,n+1} - S_{n+3,n}S_{n,n+2}}, \quad N \geqslant 4. \tag{9}$$

Here, frequency f in Eqs. (8) and (9) is omitted for brevity.

If the values of G_1 and G_2 are determined by using Eqs. (8) and (9), the wave propagation constants γ^+ and γ^- are obtained from Eqs. (7b) and (7c) as

$$\gamma^+ = \frac{-1}{j\Delta x}\ln\left\{\frac{1}{2}(G_2 + \sqrt{G_2^2 - 4G_1})\right\}, \tag{10}$$

$$\gamma^- = \frac{1}{j\Delta x}\ln\left\{\frac{G_1}{\frac{1}{2}(G_2 + \sqrt{G_2^2 - 4G_1})}\right\}. \tag{11}$$

Thus, Δk is expressed from Eqs. (10) and (11) as

$$\Delta k = \frac{k^- - k^+}{2} = \mathrm{Re}\left\{\frac{\gamma^- - \gamma^+}{2}\right\} = \frac{\angle G_1}{2\Delta x}, \tag{12}$$

where $\angle G_1$ denotes the phase angle of G_1. It is noteworthy that $\angle G_1$ is zero for stationary medium.

Also, considering $M < 1$ and $\delta^- - \delta^+ \ll 1$, δ is obtained using Eqs. (10) and (11) as

$$\delta = \frac{\delta^- + \delta^+}{2} = -\mathrm{Im}\left\{\frac{\gamma^- + \gamma^+}{2}\right\} = \frac{1}{2\Delta x}\ln\left|\frac{G_1}{\frac{1}{4}(G_2 + \sqrt{G_2^2 - 4G_1})^2}\right|, \tag{13}$$

where $\mathrm{Im}\{\cdot\}$ indicates the imaginary part of a complex number and $|\cdot|$ indicates magnitude of a complex number.

From Eqs. (5), (12), and (13), the volume flow rate, Q in the pipe can be measured by using

$$Q = \frac{-(k_0 + \delta) + \sqrt{(k_0 + \delta)^2 + 4(\angle G_1/2\Delta x)^2}}{\angle G_1/\Delta x}\frac{\pi d^2}{4}c_0. \tag{14}$$

IV. COMPENSATION OF MEASUREMENT ERRORS

The acoustic pressures measured by the microphone array are utilized to determine the flow rate with the attenuation constant, δ, and the phase angle of G_1, $\angle G_1$, in Eq. (14). We must, however, note that the measured acoustic pressures have extraneous measurement noise at the input

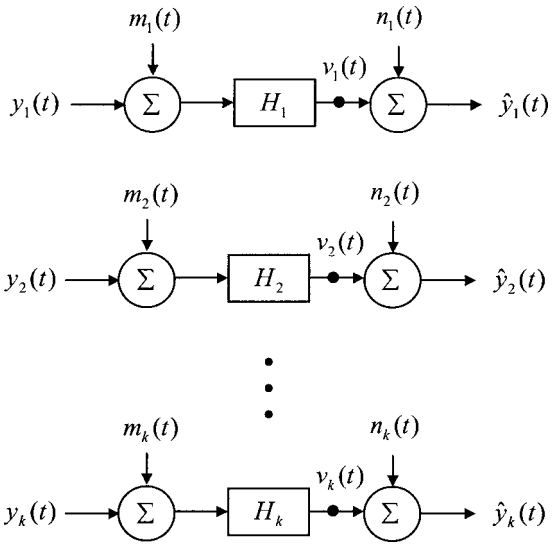


FIG. 2. Measurement system of a microphone array with extraneous noise (H_k: the transfer function of the kth microphone, m_k: the kth input noise, n_k: the kth output noise, $\hat{y}_k$: measured signal at the kth microphone, y_k: true signal at the kth microphone).

and output points of each microphone. In addition, the microphones can have different response characteristics from each other, and induce measurement errors. Therefore, the success of the proposed method greatly depends on how well we can minimize effects of the extraneous measurement noise and the mismatch of response characteristics between microphones.

The measurement system can be represented as a model where extraneous noise is measured at the input and output points to a microphone system as shown in Fig. 2. Each of the extraneous noises in the measurement system can be assumed to be mutually uncorrelated with each other, and with input and output signals. Therefore, in order to exclude the effect of the extraneous noise, it is desirable to use the cross-spectra between measurement signals of each channel when estimating acoustic properties.[13] Appendix B shows how the effect of the extraneous noise is excluded by using the cross-spectra.

If the acoustic pressures are measured by using microphones with different response characteristics from each other, inaccurate results are obtained when the flow rate is estimated using acoustic pressure spectra in Eq. (7). Thus, for obtaining more accurate flow rate, it is necessary to compensate the mismatch effects of the response characteristics between the microphones. Compensation of the mismatch effect between the microphones can be done by introducing a transfer function that expresses the response characteristic of the microphone in frequency domain. Appendix C shows the detailed derivation and associated results to compensate the effect of microphone mismatch. The results essentially show that we can minimize the errors associated with the microphone mismatch.

V. MEASUREMENT CONDITIONS

A. Propagation of one-dimensional plane waves

As discussed in Sec. II, the method proposed in this study is effective under the condition that only one-dimensional plane waves are propagated in the pipe. The upper limit of frequency range, in which only one-dimensional plane waves are propagated in the pipe, f_u, can be expressed as[14,15]

$$f_u = \frac{1.84}{\pi}\frac{c_0}{d}(1-M^2)^{1/2}. \tag{15}$$

In case of water and air, f_u is much smaller than the upper value in the frequency range, $f \ll \rho_0 c_0^2/2\pi\mu$, satisfying Eq. (1), since dynamic viscosity, μ/ρ_0, becomes about 0.11×10^{-4} and $0.16\times10^{-3}\,\mathrm{m^2/s}$ for water and air, respectively. This implies that the acoustic pressure in the frequency range below f_u can be regarded as independent of the radius of the pipe in case of the fluid with dynamic viscosity similar to water or air.

On the other hand, f_u is about 1.8 times the upper value in the frequency range, $f \ll c_0/2\pi\delta_a$, satisfying the inequality, $k\delta_a \ll 1$, considering that the aerodynamic boundary-layer thickness, δ_a, can be considered to be equal to the radius of the pipe for the case of a fully developed turbulent flow. This means that for the sound wave in the frequency range below 1/1.8 times the value of f_u, the flow velocity profile over the cross section of the pipe can be regarded to be uniform.

B. Avoidance of singular frequencies

Since Eq. (6) is utilized as the basic equation for measuring the flow rate, it is necessary to know the conditions under which Eq. (6) is valid. Equation (6) is established by using Eq. (A5). Equation (A5) has an inverse matrix, there exists a singular condition under which the matrix in Eq. (A5) becomes singular. Under the singular condition, the measured flow rate becomes inaccurate because Eq. (6) is not well established.

Assuming that the effect of sound attenuation is negligible, the singular condition, under which the determinant of the matrix in Eq. (A5) is equal to zero, is arranged as

$$\exp\{j(k^- + k^+)\Delta x\} = 1. \tag{16}$$

A singular frequency, f_s, can be expressed from Eq. (16) as

$$f_s = \left(\frac{1-M^2}{2}\right)\left(\frac{nc_0}{\Delta x}\right), \quad n = 0,1,2,..., \tag{17}$$

where Eq. (17) is obtained assuming that each microphone is positioned at one point as the ideal case.

In an actual case, the effect of microphone diameter d_m should be also considered. Thus, the range of singular frequencies in which the flow rate measured becomes inaccurate can be expressed, considering the microphone diameter and spacing, as

$$\left(\frac{1-M^2}{2}\right)\left(\frac{nc_0}{\Delta x + d_m}\right) \leqslant f_s \leqslant \left(\frac{1-M^2}{2}\right)\left(\frac{nc_0}{\Delta x - d_m}\right), \quad n = 0,1,2,.... \tag{18}$$

TABLE I. Random number distribution of error on each channel.

Error	Ch 1	Ch 2	Ch 3	Ch 4
ϵ_{mn}	0.0258	0.9210	0.7008	0.1901
ϵ_{pn}	0.5387	0.3815	0.0512	0.2851

VI. NUMERICAL SIMULATION

Numerical simulation was performed to confirm the adequacy of the proposed method. In order to get the sound field in the pipe, it was assumed that the fluid was air at 20 °C, the flow rate was 10 l/s, $d=0.04$ m, $\mu/\rho_0=0.15\times10^{-4}$ m^2/s, and $F=0.052$, and four microphones were arranged at intervals of 10 cm. With respect to the measurement conditions in Sec. V, the upper frequency for one-dimensional plane wave propagation was about 5 kHz from Eq. (15). Also, singular frequency was about 1700 Hz from Eq. (17) for $n=1$.

To study the effects of errors from microphone mismatch on the flow rate measurement, the transfer function, H_n, for the microphone of the nth channel is defined, in which errors are included in the magnitude and the phase of the transfer function as follows:

$$H_n=(1+a_m\epsilon_{mn})\exp(ja_p\epsilon_{pn}\pi/180), \tag{19}$$

where a is a coefficient which adjusts the magnitude of error and ϵ denotes the error which is determined as a random number distributed evenly between 0 and 1. Subscripts n, m, and p refer to the channel number, magnitude, and phase of the transfer function, respectively. The values of ϵ_{mn} and ϵ_{pn} are shown in Table I.

During our study, in the first case of the simulation, the microphones are differentiated only in the magnitude of the transfer function, i.e., $a_m=0.01$, 0.05, and 0.1, respectively, with $a_p=0$. Figure 3 shows the flow rate obtained by this method without compensating the resulting errors.

The microphones, in the second case, are differentiated only in the phase of the transfer function, i.e., $a_p=0.5$, 1, and 5, respectively, with $a_m=0$. Figure 4 shows the flow rate obtained without compensating the resulting errors.

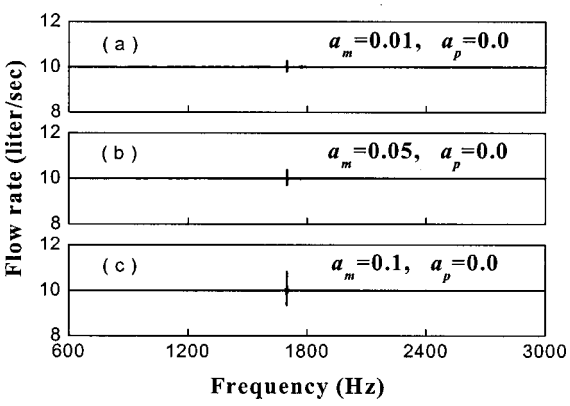


FIG. 3. Numerical simulation for flow rate of 10 l/s including errors induced by the only magnitude mismatch between microphones: (a) $a_m=0.01$, $a_p=0$; (b) $a_m=0.05$, $a_p=0$; (c) $a_m=0.1$, $a_p=0$.

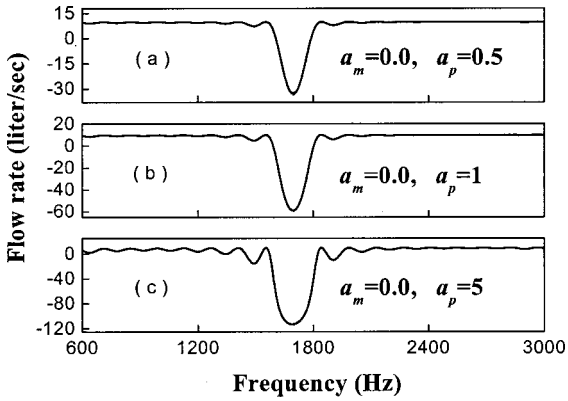


FIG. 4. Numerical simulation for flow rate of 10 l/s including errors induced by the only phase mismatch between microphones: (a) $a_m=0$, $a_p=0.5$; (b) $a_m=0$, $a_p=1$; (c) $a_m=0$, $a_p=5$.

Figures 3 and 4 show large errors in the flow rate near 1700 Hz. This frequency corresponds to the singular frequency calculated by Eq. (17) for $n=1$. It can be seen in Fig. 3 that, when the microphones are differentiated only in the magnitude of the transfer function, the flow rate is affected weakly in the frequency range excluding the singular frequency. In contrast, it can be seen in Fig. 4 that, when the microphones are differentiated only in the phase of the transfer function, the flow rate, greatly affected by the phase, fluctuates in a comparatively wide frequency region centering around the singular frequency.

In the third case, the microphones are differentiated in both the magnitude and the phase of the transfer function, i.e., $a_m=0.1$, and $a_p=5$. Figure 5(a) shows the flow rate obtained without compensating the resulting errors. The flow rate fluctuates in Fig. 5(a) much like that of Fig. 4. On the other hand, when the method of compensating errors presented in Sec. IV is applied to these results, the compensated flow rate as shown in Fig. 5(b) is obtained. It can be seen in Fig. 5(b) that some errors occur near the singular frequency but the errors at the other frequencies are compensated properly.

VII. EXPERIMENTAL RESULTS

To verify the proposed method and measurement configuration, experiments on an acryl pipe with air flow were

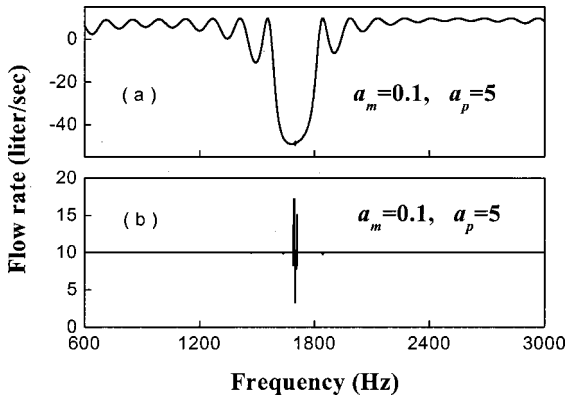


FIG. 5. Numerical simulation for flow rate of 10 l/s including errors and after compensation of errors from the phase and magnitude mismatch ($a_m=0.1$, $a_p=5$) between microphones are shown in (a) and (b), respectively.

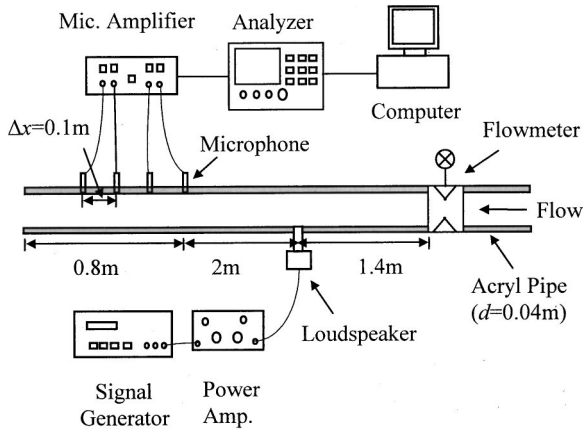


FIG. 6. Experimental setup for the measurement of flow rate in a pipe using a microphone array.

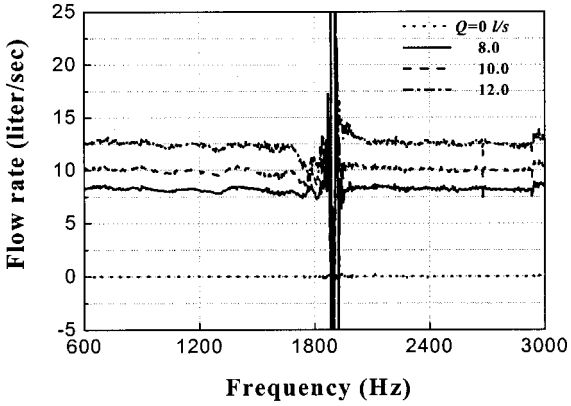


FIG. 8. Estimated air flow rates for $Q=0,8,10,12$ l/s in a pipe using four microphones with compensated errors from the phase and magnitude mismatch between microphones.

performed, as shown in Fig. 6. The diameter of the pipe was 0.04 m. Air was at atmospheric pressure and its temperature was about 20 °C. Four B&K-type 4938 microphones with the diameter of 1/4 in. were placed at the top of the pipe, and spaced longitudinally 10 cm apart from each other, and mounted flush with the inner pipe wall. White noise as the acoustic pressure field in the pipe was generated by a loudspeaker connected to the inlet pipe. Constant flow rate was maintained by using an air compressor, an air storage tank, and an automatic flow rate control valve. A conventional flowmeter, for which the calibrated uncertainty of the flow rate measurement was less than $\pm 1\%$ with a confidence level of 95%, was installed in the pipe to compare the estimated flow rate with the actual value. The flow rates became 0, 8, 10, and 12 l/s, respectively. The cross-spectra of the sound pressure signals were measured by a multichannel analyzer of a B&K-type 3560 PULSE. With respect to the measurement conditions, the upper frequency for one-dimensional plane wave propagation was about 5 kHz from Eq. (15). The range of singular frequencies, considering the microphone diameter and spacing, was about 1610–1830 Hz from Eq. (18) for $n=1$.

Figure 7 shows the estimated flow rates in the frequency range of 600–3000 Hz without compensating errors. Large errors occur near the singular frequency range of 1600–1900 Hz as expected. Also, the estimated flow rates fluctuate much like that of Fig. 5(a) centering around the singular frequency region.

Figure 8 shows the estimated flow rates with compensated errors. The flow rates approach the true values in the frequency range excluding the singular frequency region. It is noteworthy that the value of δ/k_0 is obtained to be less than 0.003 in the frequency range excluding the singular frequency region by using Eq. (13). The estimated flow rates are affected by less than 0.02% due to the effect of the attenuation constant. From this result, it can be deduced that the effect of the attenuation constant is negligible.

The estimated flow rates from Kim and Kim's method[4] were obtained to compare the present method with the one. Besides the singular frequency region, the flow rates of Kim and Kim in Fig. 9 show that there are incorrectly large values, for cases when the spatial distribution of acoustic pressure magnitude is symmetric with respect to a central sensor position because only three sensors are used. Comparison of the results in Figs. 8 and 9 shows that the proposed method is more accurate than the method of Kim and Kim.[4]

In order to get the mean flow rate from results in Fig. 8, if we use an averaging process which is commonly used to eliminate the random errors, the results could be biased because large singular values have strong effect on the average.

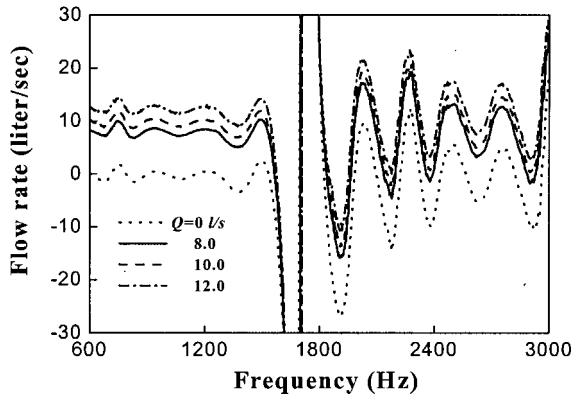


FIG. 7. Estimated air flow rates for $Q=0,8,10,12$ l/s in a pipe using four microphones without compensating errors from the phase and magnitude mismatch between microphones.

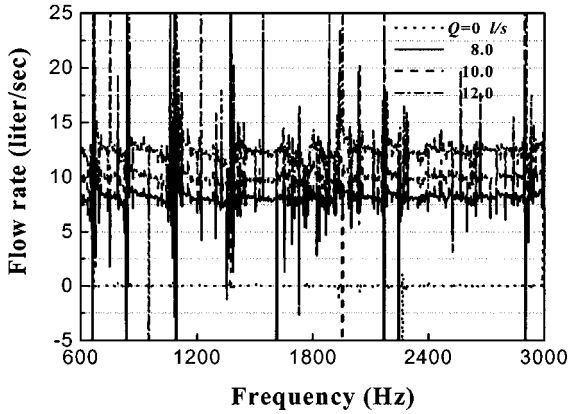


FIG. 9. Estimated air flow rates for $Q=0,8,10,12$ l/s in a pipe using the method of Kim and Kim.

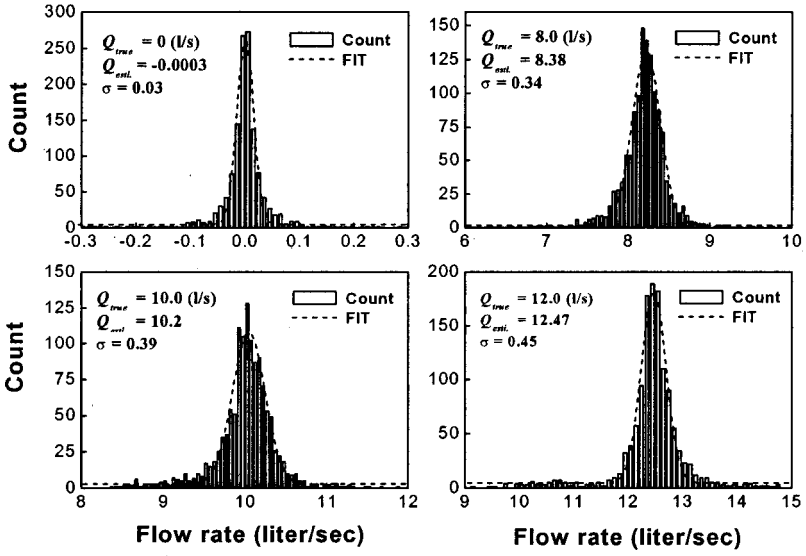


FIG. 10. Histograms of the estimated flow rate at each discrete frequency and Gaussian fit in Fig. 8 (Q_{true}: true flow rate, Q_{esti}: estimated flow rate, σ: standard deviation of the estimated flow rate).

To avoid this, histograms of the estimated flow rates and Gaussian fits, which are the curve fittings to Gaussian functions, are employed for this process. Figure 10 shows the histograms of the estimated flow rates in Fig. 8 at each discrete frequency and the Gaussian fits. As shown in Fig. 10, the mean values of Gaussian fits coincide well with the true flow rates. The percentage errors of these mean values are calculated with the true flow rates. We also calculated the standard deviations of the Gaussian fits, the rough error bounds of experimental results, as shown in Table II. The maximum errors and the maximum standard deviation are 4.8% and 0.45 l/s, respectively.

VIII. CONCLUSION

A method of measuring the flow rate in a pipe has been proposed and verified by various experiments. The method utilizes one-dimensional acoustic pressure signals measured by a microphone array. The flow rate is estimated by using the change of the acoustic wave number in the axial direction of the pipe. The change of the wave number is obtained by utilizing the recursive relation among the measured microphone array signals. A method was suggested to compensate the measurement errors of the acoustic pressures due to extraneous measurement noise and mismatch of response characteristics between microphones. By using the proposed method, the flow rates can be measured with 5% error, which is more accurate than our previous method of 12% error.

TABLE II. The flow rates and their errors estimated from the Gaussian fit.

True flow rate Q (l/s)	Estimated flow rate $\hat{Q}$ (l/s)	% error in flow rate $(\lvert\hat{Q}-Q\rvert/Q)\times 100$	Standard deviation of $\hat{Q}$ (l/s)
0	−0.0003	⋯	0.03
8.0	8.38	4.8	0.34
10.0	10.2	2	0.39
12.0	12.47	3.9	0.45

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge the support of the Korea Institute of Nuclear Safety (KINS).

APPENDIX A: DERIVATION OF RECURSIVE RELATION AMONG THE ARRAY SIGNALS

By using Eq. (2), three expressions for the acoustic pressures measured at three consecutive locations, x_n, x_{n+1}, and x_{n+2}, as shown in Fig. 1 are arranged as

$$\begin{Bmatrix} P_n(x,f) \\ P_{n+1}(x,f) \\ P_{n+2}(x,f) \end{Bmatrix} = \begin{bmatrix} E_n^+ & E_n^- \\ E_{n+1}^+ & E_{n+1}^- \\ E_{n+2}^+ & E_{n+2}^- \end{bmatrix} \begin{Bmatrix} P^+(f) \\ P^-(f) \end{Bmatrix}, \tag{A1}$$

where

$$E_n^+ = \exp(-j\gamma^+ x_n), \tag{A2}$$

$$E_n^- = \exp(j\gamma^- x_n), \tag{A3}$$

$$x_n = x_1 + (n-1)\Delta x, \quad n = 1,2,3,\ldots,N. \tag{A4}$$

Here Δx is the sensor spacing. Since Eq. (A1) represents a system of three equations in four unknowns, γ^+, γ^-, $P^+(f)$, and $P^-(f)$, the three equations can be rearranged as one equation by eliminating $P^+(f)$ and $P^-(f)$.

For example, $P^+(f)$ and $P^-(f)$ are given by taking the first and second rows in Eq. (A1) as

$$\begin{Bmatrix} P^+(f) \\ P^-(f) \end{Bmatrix} = \begin{bmatrix} E_n^+ & E_n^- \\ E_{n+1}^+ & E_{n+1}^- \end{bmatrix}^{-1} \begin{Bmatrix} P_n(x,f) \\ P_{n+1}(x,f) \end{Bmatrix}. \tag{A5}$$

A relation can be obtained by substituting Eq. (A5) into the remaining row of Eq. (A1) as

$$P_n(x,f)\exp\{j(\gamma^- - \gamma^+)\Delta x\} + P_{n+2}(x,f) = P_{n+1}(x,f)\{\exp(j\gamma^-\Delta x) + \exp(-j\gamma^+\Delta x)\}. \tag{A6}$$

Equation (A6) represents the recursive relation between the propagation constants γ^+, γ^- and the acoustic pressures measured at three consecutive locations. When three con-

secutive measurement points are shifted by one point, another recursive relation can be obtained systematically from Eq. (A6). The two recursive relations can be written as

$$\begin{bmatrix} P_n(x,f) & -P_{n+1}(x,f) \\ P_{n+1}(x,f) & -P_{n+2}(x,f) \end{bmatrix} \begin{Bmatrix} G_1(f) \\ G_2(f) \end{Bmatrix} = \begin{Bmatrix} -P_{n+2}(x,f) \\ -P_{n+3}(x,f) \end{Bmatrix}, \tag{A7}$$

where

$$G_1(f) = \exp[j(\gamma^- - \gamma^+)\Delta x], \tag{A8}$$

$$G_2(f) = \exp(j\gamma^- \Delta x) + \exp(-j\gamma^+ \Delta x). \tag{A9}$$

Equation (A7) can be rewritten in terms of cross-spectral density between the acoustic pressures to exclude errors due to an uncorrelated measurement noise. Multiplying the first row of Eq. (A7) by $P^*_{n+3}(x,f)$ and the second row by $P^*_n(x,f)$ yields[13]

$$\begin{bmatrix} S_{n+3,n}(f) & -S_{n+3,n+1}(f) \\ S_{n,n+1}(f) & -S_{n,n+2}(f) \end{bmatrix} \begin{Bmatrix} G_1(f) \\ G_2(f) \end{Bmatrix} = \begin{Bmatrix} -S_{n+3,n+2}(f) \\ -S_{n,n+3}(f) \end{Bmatrix}, \tag{A10}$$

where the superscript asterisk denotes complex conjugate and $S_{n,n+1}(f)$ denotes the cross-spectral density between $P_n(x,f)$ and $P_{n+1}(x,f)$.

APPENDIX B: EXCLUSION OF MEASUREMENT NOISE EFFECTS

The measurement system can be represented as a model where extraneous noises are measured at input and output points to the ith linear system H_i as shown in Fig. 2. Each of the extraneous noise terms $m_i(t)$, $n_i(t)$ is assumed to be mutually uncorrelated with each other and with input and output signals.

In Fig. 2, the Fourier transformed values $\hat{Y}_i(f)$ of the output signal $\hat{y}_i(t)$ can be written as

$$\hat{Y}_i(f) = V_i(f) + N_i(f), \tag{B1}$$

$$V_i(f) = (Y_i(f) + M_i(f))H_i(f), \tag{B2}$$

where $Y_i(f)$, $M_i(f)$, $N_i(f)$, and $H_i(f)$ are the input signal, extraneous noises at input and output, and the transfer function of the sensor, respectively.

From the correlation properties of $m_i(t)$, $n_i(t)$, the cross-spectra[13] between the extraneous noise and the signals become zero:

$$S_{y_i,m_j}(f) = 0, \tag{B3}$$

$$S_{y_i,n_j}(f) = 0, \tag{B4}$$

$$S_{m_i,n_j}(f) = 0, \tag{B5}$$

$$S_{m_i,m_j}(f) = 0, \tag{B6}$$

$$S_{n_i,n_j}(f) = 0. \tag{B7}$$

Therefore, the cross-spectra between measurement signals of each channel can be expressed as

$$S_{i,j}(f) = \frac{\hat{S}_{i,j}(f)}{H_i^*(f) \cdot H_j(f)}, \tag{B8}$$

where $S_{i,j}(f)$ denotes $S_{y_i,y_j}(f)$. It can be seen from Eq. (B8) that use of the cross-spectra enables us to exclude effects of the uncorrelated measurement noise at the input and output points.

APPENDIX C: DETERMINATION OF COMPENSATION FACTORS REGARDING THE MISMATCH BETWEEN THE MICROPHONES

The transfer function of the microphone is introduced to obtain compensation factors regarding the mismatch of the response characteristics between the microphones. The transfer function representing the response characteristic of the ith channel microphone, as shown in Fig. 2, is defined as

$$H_i(f) = \frac{1}{\alpha_i(f) + j\beta_i(f)}, \tag{C1}$$

where $\alpha_i(f)$ and $\beta_i(f)$ $(i = 1,2,3,...)$ are the real and imaginary part of $1/H_i(f)$.

In order to compensate the effect of microphone mismatch, Eq. (7) is rewritten by using Eqs. (C1) and (B8) as

$$\begin{bmatrix} \hat{S}_{n+3,n}(1+j\eta_1) & -\hat{S}_{n+3,n+1}(\eta_2 + j\eta_3) \\ \hat{S}_{n,n+1}(1+j\zeta_1) & -\hat{S}_{n,n+2}(\zeta_2 + j\zeta_3) \end{bmatrix} \begin{Bmatrix} G_1 \\ G_2 \end{Bmatrix} = \begin{Bmatrix} -\hat{S}_{n+3,n+2}(\eta_4 + j\eta_5) \\ -\hat{S}_{n,n+3}(\zeta_4 + j\zeta_5) \end{Bmatrix}, \tag{C2}$$

where

$$\eta_1 = \frac{\alpha_{n+3}\beta_n - \alpha_n\beta_{n+3}}{\Delta_1}, \quad \eta_2 = \frac{\alpha_{n+3}\alpha_{n+1} - \beta_{n+3}\beta_{n+1}}{\Delta_1}, \quad \eta_3 = \frac{\alpha_{n+3}\beta_{n+1} - \alpha_{n+1}\beta_{n+3}}{\Delta_1},$$

$$\eta_4 = \frac{\alpha_{n+3}\alpha_{n+2} - \beta_{n+3}\beta_{n+2}}{\Delta_1}, \quad \eta_5 = \frac{\alpha_{n+3}\beta_{n+2} - \alpha_{n+2}\beta_{n+3}}{\Delta_1}, \quad \Delta_1 = \alpha_{n+3}\alpha_n - \beta_{n+3}\beta_n,$$

$$\zeta_1 = \frac{\alpha_n\beta_{n+1} - \alpha_{n+1}\beta_n}{\Delta_2}, \quad \zeta_2 = \frac{\alpha_n\alpha_{n+2} - \beta_n\beta_{n+2}}{\Delta_2}, \quad \zeta_3 = \frac{\alpha_n\beta_{n+2} - \alpha_{n+2}\beta_n}{\Delta_2},$$

$$\zeta_4 = \frac{\alpha_n\alpha_{n+3} - \beta_{n+3}\beta_n}{\Delta_2}, \quad \zeta_5 = \frac{\alpha_n\beta_{n+3} - \alpha_{n+3}\beta_n}{\Delta_2}, \quad \Delta_2 = \alpha_n\alpha_{n+1} - \beta_n\beta_{n+1}.$$

Here, frequency f in Eq. (C2) is omitted for brevity and η_r and ζ_r $(r=1,2,3,4,5)$ denote compensation factors associated with the effect of microphone mismatch.

The unknown compensation factors, η_r and ζ_r in Eq. (C2) are real numbers, and if $\hat{S}_{i,j}$ $(i,j=1,2,3\ldots)$ and G_m $(m=1,2)$ in Eq. (C2) are divided into real and imaginary parts, two equations for η_r and ζ_r can be obtained, respectively. To obtain the equations, the real and imaginary parts of $\hat{S}_{i,j}$ and G_m are defined as follows:

$$\hat{C}_{i,j}=\mathrm{Re}(\hat{S}_{i,j}), \tag{C3}$$

$$\hat{Q}_{i,j}=-\mathrm{Im}(\hat{S}_{i,j}), \tag{C4}$$

$$G_{mr}=\mathrm{Re}(G_m), \tag{C5}$$

$$G_{mi}=\mathrm{Im}(G_m). \tag{C6}$$

Substituting Eqs. (C3)–(C6) into Eq. (C2) and separating the resultant equation into its real and imaginary parts result in Eqs. (C7) and (C8), respectively,

$$[A_\eta(f)]\{\eta(f)\}=\{B_\eta(f)\}, \tag{C7}$$

where

$$[A_\eta(f)]=\begin{bmatrix} -(\hat{C}_{n+3,n}G_{1i}-\hat{Q}_{n+3,n}G_{1r}) & -(\hat{C}_{n+3,n+1}G_{2r}+\hat{Q}_{n+3,n+1}G_{2i}) & (\hat{C}_{n+3,n+1}G_{2i}-\hat{Q}_{n+3,n+1}G_{2r}) & \hat{C}_{n+3,n+2} & \hat{Q}_{n+3,n+2} \\ (\hat{C}_{n+3,n}G_{1r}+\hat{Q}_{n+3,n}G_{1i}) & -(\hat{C}_{n+3,n+1}G_{2i}-\hat{Q}_{n+3,n+1}G_{2r}) & -(\hat{C}_{n+3,n+1}G_{2r}-\hat{Q}_{n+3,n+1}G_{2i}) & -\hat{Q}_{n+3,n+2} & \hat{C}_{n+3,n+2} \end{bmatrix},$$

$$\{\eta(f)\}=[\eta_1,\eta_2,\eta_3,\eta_4,\eta_5]^T,$$

$$\{B_\eta(f)\}=\begin{Bmatrix} -(\hat{C}_{n+3,n}G_{1r}+\hat{Q}_{n+3,n}G_{1i} \\ -(\hat{C}_{n+3,n}G_{1i}-\hat{Q}_{n+3,n}G_{1r}) \end{Bmatrix}.$$

Also,

$$[A_\zeta(f)]\{\zeta(f)\}=\{B_\zeta(f)\}, \tag{C8}$$

where

$$[A_\zeta(f)]=\begin{bmatrix} -(\hat{C}_{n,n+1}G_{1i}-\hat{Q}_{n,n+1}G_{1r}) & -(\hat{C}_{n,n+2}G_{2r}+\hat{Q}_{n,n+2}G_{2i}) & (\hat{C}_{n,n+2}G_{2i}-\hat{Q}_{n,n+2}G_{2r}) & \hat{C}_{n,n+3} & \hat{Q}_{n,n+3} \\ (\hat{C}_{n,n+1}G_{1r}+\hat{Q}_{n,n+1}G_{1i}) & -(\hat{C}_{n,n+2}G_{2i}-\hat{Q}_{n,n+2}G_{2r}) & -(\hat{C}_{n,n+2}G_{2r}-\hat{Q}_{n,n+2}G_{2i}) & -\hat{Q}_{n,n+3} & \hat{C}_{n,n+3} \end{bmatrix},$$

$$\{\zeta(f)\}=[\zeta_1,\zeta_2,\zeta_3,\zeta_4,\zeta_5]^T,$$

$$\{B_\zeta(f)\}=\begin{Bmatrix} -(\hat{C}_{n,n+1}G_{1r}+\hat{Q}_{n,n+1}G_{1i}) \\ -(\hat{C}_{n,n+1}G_{1i}-\hat{Q}_{n,n+1}G_{1r}) \end{Bmatrix}.$$

Equations (C7) and (C8) represent the recursive relations between measured spectrum and the unknown compensation factors. Equations (C7) and (C8) consist of two equations, respectively, with the five unknowns for η_r and ζ_r. To solve the equations, assuming that the compensation factors are the same in the narrow frequency band around the frequency f of interest, more than six equations can be formed in this frequency band as follows:

$$[A_\eta(f_k)]\{\eta(f_k)\}=\{B_\eta(f_k)\} \quad (f-\Delta f/2<f_k<f+\Delta f/2,\ k=1,2,\ldots,n,\quad n\geqslant 3), \tag{C9}$$

$$[A_\zeta(f_k)]\{\zeta(f_k)\}=\{B_\zeta(f_k)\} \quad (f-\Delta f/2<f_k<f+\Delta f/2,\ k=1,2,\ldots,n,\quad n\geqslant 3), \tag{C10}$$

where Δf is frequency bandwidth, and f_k is an arbitrary frequency within the frequency band from $f-\Delta f/2$ to $f+\Delta f/2$. Since Eqs. (C9) and (C10) become overdetermined linear matrix equations, the unknown compensation factors η_r and ζ_r are determined by applying the least-squares method as follows:[16]

$$\{\eta\}=([A_\eta]^H[A_\eta])^{-1}[A_\eta]^H\{B_\eta\}, \tag{C11}$$

$$\{\zeta\}=([A_\zeta]^H[A_\zeta])^{-1}[A_\zeta]^H\{B_\zeta\}, \tag{C12}$$

where $[A]^H$ denotes the Hermitian matrix of $[A]$.

Actually, the compensation factors, η_r and ζ_r are obtained using the measured spectrum $\hat{S}_{i,j}$ and the values of G_1 and G_2 estimated from Eqs. (7b) and (7c) under the stationary fluid condition.

From the fact that response characteristics of microphones are irrespective of the flow rate, the compensation factors associated with the transfer functions of the microphones are also irrespective of the flow rate, thus, the compensation factors can be applied to the moving fluid condition.

Under the moving fluid condition, the values of G_1 and G_2 are obtained from Eq. (C2), using the measured spectrum $\hat{S}_{i,j}$ and the compensation factors, η_r and ζ_r. The values G_1 and G_2 represent the values with the compensation of the

microphone mismatch errors. Therefore, by using the compensated values of G_1 and G_2, the microphone mismatch errors can be minimized and the flow rate can be measured accurately.

[1] R. J. Goldstein, *Fluid Mechanics Measurements* (Springer, New York, 1983), Chap. 6, pp. 245–306.
[2] R. N. Thurston and A. D. Pierce, *Ultrasonic Instruments and Devices I* (Academic, New York, 1999), pp. 353–358.
[3] B. Robertson, ''Flow and temperature profile independence of flow measurements using long acoustic waves,'' ASME J. Fluids Eng. **106**, 18–20 (1984).
[4] Y.-K. Kim and Y.-H. Kim, ''A three accelerometer method for the measurement of flow rate in pipe,'' J. Acoust. Soc. Am. **100**, 717–726 (1996).
[5] W.-S. Cheung, H.-S. Kwon, K.-A. Park, and J.-S. Paik, ''Acoustic flowmeter for the measurement of the mean flow velocity in pipes,'' J. Acoust. Soc. Am. **110**, 2308–2314 (2001).
[6] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), pp. 698–716.
[7] M. L. Munjal, *Acoustics of Ducts and Mufflers* (Wiley, New York, 1987), pp. 13–26.
[8] D. C. Pridmore-Brown, ''Sound Propagation in a fluid flowing through an attenuating duct,'' J. Fluid Mech. **4**, 393–406 (1958).
[9] D. H. Tack and R. F. Lambert, ''Influence of shear flow on sound attenuation in lined ducts,'' J. Acoust. Soc. Am. **38**, 655–666 (1965).
[10] M. L. Munjal and A. G. Doige, ''The two-microphone method incorporating the effects of mean flow and acoustic damping,'' J. Sound Vib. **137**, 135–138 (1990).
[11] U. Ingard and V. K. Singhal, ''Sound attenuation in turbulent pipe flow,'' J. Acoust. Soc. Am. **55**, 535–538 (1974).
[12] S. H. Jang and J.-G. Ih, ''On the multiple microphone method for measuring in-duct acoustic properties in the presence of mean flow,'' J. Acoust. Soc. Am. **103**, 1520–1526 (1998).
[13] J. S. Bendat and A. G. Piersol, *Random Data* (Wiely, New York, 1986), pp. 164–181.
[14] V. Mason, ''Some experiments on the propagation of sound along a cylindrical duct containing flowing air,'' J. Sound Vib. **10**, 208–226 (1969).
[15] F. Fahy, *Sound and Structural Vibration* (Academic, New York, 1985), pp. 205–210.
[16] Å. Björck, *Numerical Methods for Least Squares Problems* (Society for Industrial and Applied Mathematics, 1996), pp. 9–18.

Responses of partially immersed elastic structures using a symmetric formulation for coupled boundary element and finite element methods

Pei-Tai Chen[a)]
Department of System Engineering and Naval Architecture, National Taiwan Ocean University, Keelung, Taiwan, Republic of China

Chorng-Shyan Lin
Chung-Shan Institute of Science and Technology, Taoyuan, Taiwan, Republic of China

Tachung Yang
Department of Mechanical Engineering, Yuan-Ze University, Taoyuan, Taiwan, Republic of China

(Received 25 June 2001; accepted for publication 23 May 2002)

Using a coupled BEM/FEM, this work describes a numerical method to compute the response and acoustic radiation for structures partially immersed in fluid. The structures and their responses are assumed to be symmetric about a symmetric plane. A symmetric complex matrix derived from the BEM and a reciprocal principle for surface acoustics is also used to represent the acoustic loading against the structures. In addition, selecting a proper Green's function based on image source method satisfies the boundary conditions of pressure release on the fluid surface and null normal velocity on the symmetric plane. Moreover, a boundary integral equation emerges when the field point approaches the structural surface where the normal derivative of the Green's function over partial, infinitesimal spheres is evaluated. These limiting values depend on locations of the field point on the surface. Owing to the symmetry of the acoustic loading matrix, the matrix for the coupled BEM/FEM is a banded, symmetric one, thereby allowing us to employ a variable banded storage method and invert of the matrix. Doing so markedly increases computational efficiency. Furthermore, an analytical solution of a spherical thin shell with the lower semi-sphere immersed in water is carried out by characteristic function expansions for shell equation and acoustic loading. These analytical solutions compare with the results obtained from the proposed numerical method. A good correlation for low frequencies is obtained and minor discrepancies are observed with an increasing frequency. [DOI: 10.1121/1.1494448]

PACS numbers: 43.20.Rz, 43.20.Tb [JGM]

I. INTRODUCTION

Many numerical methods have been proposed in recent decades to predict vibrations of elastic structures vibrating in a fluid medium. For an arbitrarily elastic structure, the finite element method is a conventional means of modeling the response subject to external loading without fluid loading. The solution of an acoustic wave equation under mono-frequency oscillation in an infinite exterior domain is best described using the boundary element method. When a structure is immersed in a fluid, the surrounding fluid affects its vibration, subsequently leading to a coupled structural acoustic problem. Everstine and Henderson[1] discretized a Helmholtz boundary integral equation by using a constant element to couple with a NASTRAN program, which modeled the structural equation, in order to determine the response and acoustic radiation. The coupling of acoustic loading with the structural equation is achieved by eliminating structural variables in which the surface pressure is used as the unknown variable. The method was implemented on supercomputers for large scale problems. Jeans and Mathews[2] studied the coupling problem by using a variational boundary method in which the normal derivative Helmholtz integral equation is regularized and formulated onto a variational form. In doing so, the surface acoustics for thin shells on the both sides of the surfaces are represented by the integral equation and its normal derivative equation. Their investigations also included elastoacoustic analysis for fluid-filled thin shells.[3] The structural equation was coupled to the fluid by using *in vacuo* natural mode expansions. Seybert *et al.*[4] considered the structural displacement expanded by Ritz vectors to take the advantage of the fact that the displacement response depends on the distribution of applied external forces. That work also proposed a participation factor based on the loading condition so that the basis Ritz vectors could be selected properly before solving the problem. Other coupling methods and related applications can be found in Refs. 5–8.

Most matrix equations for fluid loading do not possess symmetric properties, making them numerically inefficient for memory storage and matrix inversion. However, the matrices derived from the finite element method are banded, symmetric features, implying that the coupled fluid/structure problems could be efficiently solved if symmetric acoustic loading could be formulated. Our recent work[9] proposed such a symmetric acoustic loading in which the variable band (skyline method) storage and matrix inversion are used

a)Electronic mail: ptchen@sena.ntou.edu.tw

to significantly increase numerical efficiency. Symmetry of the acoustic loading is achieved by identifying an acoustic reciprocity principle for a body immersed in an infinite extent medium.

Mikami and Yoshimura[10] studied free vibrations of shells of revolution with external or internal fluid loading. The fluid motion is treated analytically using eigenfunction expansion of the fluid domain. The fluid is incompressible and structures are either partially filled or partially submerged with the fluid. The equations of motion of the shells are derived as integro-differential equations and are solved by a collocation method. Amabili[11] analyzed free vibrations of partially fluid-loaded simply supported cylindrical shells. The fluid loading can be internally or externally, both compressible and incompressible fluids. He expands the fluid pressure and shell deformations as linear combinations of eigenfunction expansion. The model properties are investigated using a Rayleigh quotient in which the potential and kinetic energies are derived from the eigenfunction expansions. The mode shapes and natural frequencies are obtained for circular cylindrical shells with various fluid loading levels either internally or externally. Amabili[12] later studied circular cylindrical shells filled with a dense fluid and partially immersed in a different liquid having a free surface. The effect of free surface waves on the structures is studied for which the amplitude of the free surface waves is larger or smaller than the wall displacement. In the study, the external fluid is incompressible.

Despite the above investigations, solutions using discretized numerical structures partially immersed in fluid, such as a surface ship floating on a water surface, have seldom been addressed. Such structures are often symmetric about a symmetric plane. Therefore, this study develops a symmetric loading matrix for partially submerged structures with symmetric responses about the structure's symmetric planes. A closely related issue is a modified Green's function in a boundary integral equation making the boundary conditions satisfied automatically. Such investigations were first seen in Ref. 13, where an image source was added to a Green's function to satisfy the hard wall condition. The present analysis uses a boundary integral equation with a proper modified Green's function to formulate the acoustic loading effect as a symmetric complex matrix satisfying the associated boundary conditions. An analytical solution for a spherical thin shell with the lower semi-sphere immersed in water is also derived in order to compare the results obtained from the proposed numerical method with the analytical solution.

II. A BOUNDARY ELEMENT METHOD FOR A PARTIALLY IMMERSED BODY WHOSE RESPONSE IS SYMMETRIC ABOUT A PLANE

Figure 1 illustrates a floating body whose geometry and normal velocity are symmetric about a plane. The acoustic medium is located on the lower half of the space. Owing to its symmetry, the acoustic domain $\mathbf{\Omega}$ considered herein is one half of the fluid domain Ω shown in the figure. Since the normal velocity on the body is symmetric about the plane, the boundary condition of the symmetric plane is a null condition of the normal velocity. The fluid surface is a pressure release boundary condition. In order to formulate an integral representation of the acoustic domain, a modified Green's function is employed to meet the boundary conditions. Seyber and Wu[13] proposed such modifications using acoustic image points that a pair of positive images yields a hard surface while a pair of out of phase images leads to a soft surface. The present analysis combines the positive and negative image points to meet the boundary conditions on the water surface as well as the symmetric plane. The modified Green's function is

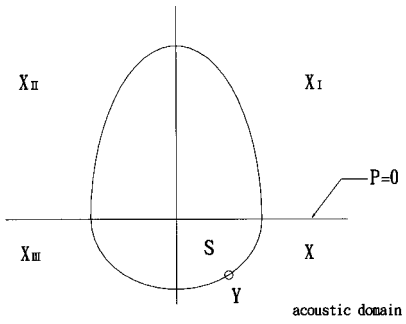


FIG. 1. Geometry of semi-immersed symmetric structures.

$$G(X,Y)=\frac{1}{4\pi}\left(\frac{e^{ik|X-Y|}}{|X-Y|}+\frac{e^{ik|X_{\mathrm{III}}-Y|}}{|X_{\mathrm{III}}-Y|}-\frac{e^{ik|X_{\mathrm{I}}-Y|}}{|X_{\mathrm{I}}-Y|}-\frac{e^{ik|X_{\mathrm{II}}-Y|}}{|X_{\mathrm{II}}-Y|}\right), \tag{1}$$

where X is a field point in the acoustic domain; X_{I}, X_{II}, X_{III} are the corresponding image points, X_{III} representing the positive in-phase image point, X_{I} and X_{II} representing the negative out-phase image points; Y is a source point on the wetted surface; k denotes a wave number, $k=\omega/c_0$; ω is the circular vibrating frequency; and c_0 is the sound speed of the fluid.

One uses the above Green's function leading to a boundary integral formulation, satisfying the zero normal velocity on the symmetric plane and null pressure on the water surface. The field pressure in the water $p(x)$ relates to the wetted surface pressure $p(Y)$ and its surface normal derive $\partial p(Y)/\partial n_y$ as

$$p(X)=\int_S\left[p(Y)\frac{\partial G(X,Y)}{\partial n_y}-\frac{\partial p(Y)}{\partial n_y}G(X,Y)\right]dS_y, \tag{2}$$

where the surface integral is only defined on the half-wetted surface of the symmetric body, and the variables $p(Y)$ and $\partial p(Y)/\partial n_y$ are defined on the half-wetted surface (see Fig. 1 where the S is defined on the right hand side). The symbol $(\partial/\partial n_y)$ denotes normal derivative on the surface S, and dS_y is the differential area integrating with respect to the source variable Y. Equation (2) satisfies the above stated boundary conditions. The Green's function is zero for the field point X on the water surface. The surface pressure is also zero on the water surface. These facts lead to the surface integral of Eq. (2) vanishing for the field point located on the water surface. Moreover, the Green's function in Eq. (1) is unchanged when one interchanges the field point X with the image point X_{III},

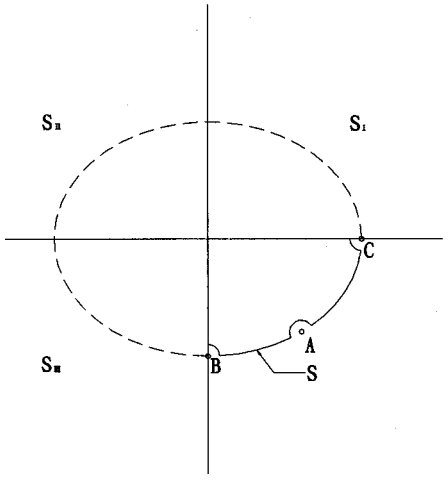


FIG. 2. The limiting field points X on the surface S.

as well as the point X_{I} with X_{II}, implying that the Green's function is symmetric about the symmetric plane. It can also be verified that the normal gradient $\partial G/\partial n_y$ is symmetric. These symmetries cause the field pressure $p(X)$ to be symmetric, indicating that the normal derivative of the pressure on the symmetric plane is zero, and thus satisfying the null normal velocity condition. The above observations reveal that Eq. (2) with the Green's function defined in Eq. (1) satisfies the boundary conditions on the fluid surface as well as the symmetric plane.

The surface integral equation relating surface pressure and its normal derivative on the surface S emerges when one allows the field point X approach to the surface S. The limiting value of the above equation for the field point X on the surface is obtained by considering a small fractional sphere subtracted from the vibrating surface as shown in Fig. 2. The contribution of the small sphere arises from the normal derivative of the Green's function, whereas the integration associated with the Green's function is limited to zero when the radius of the small sphere shrinks to zero. One of three situations occurs when evaluating the normal derivative of the surface integral of the small fractional sphere: (I) a point on the symmetric plane comes into contact with the body; (II) a point on the fluid surface touches with the body; or (III) a point is neither case I nor II. For case III, the image points X_{I}, X_{II}, X_{III} are finite distances from the point X. Thus, the singularity contributing to the surface integral only arises from the function normal derivative of $e^{ik|X-Y|}/|X-Y|$. Referring to Fig. 2, the limiting value when the small radius tends to infinitesimal becomes a solid angle subtending within the body at point X divided by 4π. The limiting value for X on location A is given by (see Fig. 2)

$$\Omega(X)=\frac{1}{4\pi}\int_{S+S_{\mathrm{I}}+S_{\mathrm{II}}+S_{\mathrm{III}}}\frac{\partial(1/|X-Y|)}{\partial n_y}dS_y, \tag{3}$$

where S_{I}, S_{II}, S_{III} denote reflecting surfaces of S either with the symmetric plane or the water surface as depicted in Fig. 2. In case I, the image point X_{III} is coincident with X and the points X_{I} and X_{II} have finite distances with the point X. Thus, the contribution of the small partial sphere integration for X on location B becomes

$$\Omega(X)=2\cdot\frac{1}{4\pi}\int_{S+S_{\mathrm{I}}}\frac{\partial(1/|X-Y|)}{\partial n_y}dS_y, \tag{4}$$

where the factor of 2 is due to the coincidence of the points X and X_{II} and the surface integral only refers to the surfaces S and S_{I}. (S_{I} is the image surface of S, see Fig. 2.) In the above equation, the surface integrals associated with the terms $e^{ik|X_{\mathrm{I}}-Y|}/|X_{\mathrm{I}}-Y|$ and $e^{ik|X_{\mathrm{II}}-Y|}/|X_{\mathrm{II}}-Y|$ vanish for those two regular functions. When the field point X is located on case II, the points X and X_{I}, X_{II} and X_{III} are identical, respectively, thus making the Green's function and its normal derivative become zero. Correspondingly, the limiting surface integral value of the small sphere for X on location C is zero:

$$\Omega(X)=0. \tag{5}$$

Thus, the surface integral equation relating the surface pressure and its normal derivative becomes

$$p(X)=p(X)\Omega(X)+\int_S\left[p(Y)\frac{\partial G(X,Y)}{\partial n_y}-\frac{\partial p(Y)}{\partial n_y}G(X,Y)\right]dS_y, \tag{6}$$

where the $\Omega(X)$ is defined in Eqs. (3)–(5), depending on the location of X.

The surface pressure and its normal derivative can be discretized by shape function approximations. The surface pressure and its normal derivative are conveniently interpolated by linear order shape functions. Doing so leads to

$$[H]\{P\}=[G]\left\{\frac{\partial P}{\partial n}\right\}, \tag{7}$$

where the two surface quantities p and $\partial p/\partial n$ in Eq. (6) relate to the nodal values $\{P\}$ and $\{\partial P/\partial n\}$ as

$$p=\sum_i P_i\psi_i, \tag{8}$$

$$\frac{\partial p}{\partial n}=\sum_i\left(\frac{\partial P}{\partial n}\right)_i\Phi_i, \tag{9}$$

in which ψ_i and Φ_i denote shape functions for surface pressure and normal velocity, respectively. In the following, a symmetric formulation for the acoustic loading is presented.

III. A COUPLED EQUATION FOR ACOUSTIC STRUCTURAL INTERACTIONS

The acoustic loading on vibrating structures is derived from the virtual work δW done on the structure due to the surface pressure

$$\delta W=\int_S p\,\delta w\,dS=\int_S\left(\sum_i P_i\psi_i\right)\left(\sum_j \delta x_{n,j}\Phi_j\right)dS$$
$$=\sum_i\sum_j p_i\delta x_{n,j}\int\psi_i\Phi_j\,dS$$
$$=\sum_i\sum_j p_i\delta x_{n,j}N_{ij}, \tag{10}$$

where P_i and $x_{n,j}$ are the surface and normal displacement nodal values, respectively, the symbol "δ" denotes virtual increments of the normal displacement, and w is the normal displacement. The coefficients N_{ij} are shape factors arising from the surface integral of the shape functions interpolating the nodal values. The structural vibration can be described by a finite element formulation. Since the surface normal displacement has been chosen as one variable, two perpendicular variables that are tangent to the surface displacement are naturally selected as the other two variables to define the linear displacement variables. For the nodal points of the structural equation not coming into contact with the fluid, the orientation of the three displacements can be arbitrarily selected, e.g., based on a global coordinate system. The coordinate orientation for rotation variables can also be chosen as the global coordinate system. Therefore, a typical element stiffness matrix oriented to the normal surface direction can be computed by

$$[K_e]=T^T[\hat{K}_e]\,T, \tag{11}$$

where the matrix T transforms the individual stiffness matrix $\lfloor\hat{k}_e\rfloor$ from the global coordinate system to the orientation described above. The matrix T is defined as

$$T=\begin{bmatrix} \Phi_i & & & & 0 \\ & I & & & \\ & & \Phi_j & & \\ & & & \cdots & \\ 0 & & & & I \end{bmatrix}, \tag{12}$$

where Φ_i and Φ_j are proper 3×3 directional cosine matrices for linear displacements on the wetted surface between the global coordinate system and the normal surface directions, and I is a unit 3×3 identity matrix corresponding to those nodal points not on the surface or the rotational displacements. Summing the elementary matrices lead to a total stiffness matrix $[K]$,

$$[K]=\sum_e\,[K_e]. \tag{13}$$

Numerical practitioners for modeling shell structures by finite element methods of shell structures frequently use linear order shape functions. It is useful to comment on the convergence of shell elements as discretized by shape functions. A curved shell could be approximated by a surface built up of small, flat elements.[14] This approximation is of a physical, rather than a mathematical, nature. A classical thin bending plate modeling requires not only the normal displacement continuity but its slope because the bending strain energy in a plate model involves twice derivative of the normal displacement. However, using elements to impose slope continuity between interface elements often raises many mathematical and computational difficulties. It was pointed out in Ref. 14 that a linear shape function still be able to make discretizations convergent since it satisfies the "constant strain" and the "patch test" criterions.

The total stiffness matrix can then be partitioned into the surface normal degree of freedom as denoted by a subscript symbol "n" and the remaining degrees of freedom with "i", which include the surface tangent and rotational displacements,

$$[K]=\begin{bmatrix} K_{nn} & K_{ni} \\ K_{in} & K_{ii} \end{bmatrix}. \tag{14}$$

Correspondingly, the submerged structural equation subject to the acoustic loading and external forces oscillating at a frequency ω can be obtained by applying Hamilton's principle, in which the pressure loading is referred to Eq. (10),

$$\left(\begin{bmatrix} K_{nn} & K_{ni} \\ K_{in} & K_{ii} \end{bmatrix}-\omega^2\begin{bmatrix} M_{nn} & M_{ni} \\ M_{in} & M_{ii} \end{bmatrix}\right)\begin{Bmatrix} x_n \\ x_i \end{Bmatrix}=\begin{Bmatrix} f_n \\ f_i \end{Bmatrix}+\begin{Bmatrix} -[N]^TP \\ 0 \end{Bmatrix}, \tag{15}$$

where $M_{pq}\,(p,q=n,i)$ are global mass matrices corresponding to the n and i degrees of freedom, x_n and x_i are the corresponding displacement variables, f_n and f_i are the associated external alternating forces, P is the discretized surface pressure, and N is a shape factor matrix as indicated in Eq. (10). The variable x_i denotes the linear displacements other than x_n, including all the rotational displacements.

The boundary conditions for a symmetric structure deformed about a symmetric plane (Fig. 1) are no linear displacement perpendicular to the symmetric plane and no rotations except the direction normal to the symmetric plane. The geometrical constraints are easily achieved by a proper coordinate transformation in Eq. (12) and the associated transformed matrix in Eq. (11).

The pressure normal derivative is related to the normal displacement as

$$\frac{\partial p}{\partial n}=\rho_0\omega^2x_n, \tag{16}$$

where ρ_0 is the fluid density. Combining Eqs. (7) and (16) leads to

$$\{P\}=[H]^{-1}[G]\rho_0\omega^2\{x_n\}. \tag{17}$$

Thus, Eq. (15) becomes

$$\begin{bmatrix} D_{nn} & D_{ni} \\ D_{in} & D_{ii} \end{bmatrix}\begin{Bmatrix} x_n \\ x_i \end{Bmatrix}+\begin{bmatrix} \rho_0\omega^2[N]^T[H]^{-1}[G] & 0 \\ 0 & 0 \end{bmatrix}\begin{Bmatrix} x_n \\ x_i \end{Bmatrix}=\begin{Bmatrix} f_n \\ f_i \end{Bmatrix}, \tag{18}$$

where $D_{pq}\,(p,q=i,n)$ are identified as

$$D_{pq}=K_{pq}-\omega^2M_{pq}. \tag{19}$$

IV. SYMMETRY OF THE ACOUSTIC LOADING MATRIX

The acoustic loading in Eq. (18) is shown in the following to be symmetric owing to a reciprocity principle for surface acoustics of a body immersed in fluid. Acoustic reciprocity principles have various versions applied to various applications. Chen and Ginsberg[15] proposed a reciprocity principle for surface acoustics for a vibrating radiator for his studying acoustic radiation mode from a perspective of complex acoustic power. Chen *et al.*[9] later applied the reciprocity to coupled BEM/FEM formulation for submerged elastic shells of which the acoustic loading matrix is symmetric. The symmetry stems from the reciprocity, which is

$$\int_S p_1 v_2 \, ds = \int_S p_2 v_1 \, ds, \tag{20}$$

where p_1, p_2 and v_1, v_2 are two arbitrary sets of surface pressures and normal velocities for a radiator immersed in an acoustic medium of infinite extent. This situation is in parallel to the case that the Maxwell–Betti reciprocity[16] leads to a symmetric stiffness matrix in finite element methods.

A surface reciprocity is still valid for the present analysis for the acoustic domain considered herein to be the left-half side shown in Fig. 1. There are three surfaces bounded the domain: the vibrating surface *S*, the symmetric plane from the intersection with the surface *S* extending downward to infinity, and the pressure release water surface. An identity for the acoustic domain is held for a nonviscid linear acoustics[17]

$$\nabla \cdot (p_1 \vec{v}_2) = \nabla \cdot (p_2 \vec{v}_1), \tag{21}$$

where p_1, p_2 and $\vec{v}_1$, $\vec{v}_2$ are two arbitrary pressure and acoustical particle velocity distributions over the space domain. The reciprocity emerges as the above identity is integrated over the domain. Transforming the volume integration transforming into a surface integral yields Eq. (20) where surface integrals over the symmetric plane, water surface, and a large quarter sphere denoting far field vanish due to the boundary conditions and Sommerfeld radiation condition. The symmetry of the loading matrix in Eq. (18) is verified as one takes v_1 to be a nodal value V_i at the ith nodal surface point with other nodal values to be zero, and v_2 to be a nodal velocity V_j at the jth nodal surface point with other nodal values to be zero. The corresponding nodal surface pressures are provided by Eq. (7), which are

$$\begin{aligned} \{P_1\} &= -\rho_0 i\omega([H]^{-1}[G])_i V_i, \\ \{P_2\} &= -\rho_0 i\omega([H]^{-1}[G])_j V_j, \end{aligned} \tag{22}$$

where $([H]^{-1}[G])_1$ denote the lth column vector. Note from Eq. (10) that a shape factor matrix N_{ij} associated with the surface integral arises because of the shape function interpolations for surface pressure and normal velocity. The discretized representation of Eq. (20) for v_1 and v_2 chosen as described above is

$$\sum_l -\rho_0 i\omega([H]^{-1}[G])_{li} V_i V_j N_{lj} = \sum_l -\rho_0 i\omega([H]^{-1}[G])_{lj} V_j V_i N_{li}, \tag{23}$$

where $([H]^{-1}[G])_{pq}$ denotes the matrix element in (p,q) position. The above equation yields the symmetry

$$(N^T[H]^{-1}[G])_{ji} = (N^T[H]^{-1}[G])_{ij}. \tag{24}$$

V. A FORMULATION USING A GLOBAL COORDINATE SYSTEM

It is useful to comment the formulation of Eq. (18) when a global coordinate system is used rather than the surface normal coordinate. This becomes evident when one introduces a transformation between the surface normal coordinate and the global coordinate system,

$$\begin{Bmatrix} x_n \\ x_i \end{Bmatrix} = \begin{bmatrix} T_1 & 0 \\ T_2 & T_3 \end{bmatrix} \begin{Bmatrix} x_w \\ x_d \end{Bmatrix}, \tag{25}$$

where x_w denotes three linear displacement component vectors in the global system, which contacts the water, and x_d represents displacements other than x_w, including all the rotational displacements. The matrix T_1 relates the normal displacement x_n to the global linear displacements on the wetted surface. Let N_w denote the modal numbers attached with the wetted surface. The dimension of T_1 is N_w by $3N_w$. The matrix T_2 transforms the linear displacements x_w of the wetted surface into the components tangent to the wetted surface. The matrix T_3 represents the relationship between the two coordinate systems for the degrees of freedom not contacting with fluid and all the rotational displacements. Substituting Eq. (25) into Eq. (18) and premultiplying the transpose of the overall transformation matrix in Eq. (25) leads to an equation using the global coordinate system, which is

$$\begin{aligned} &\begin{bmatrix} T_1 & 0 \\ T_2 & T_3 \end{bmatrix}^T \begin{bmatrix} D_{nn} & D_{ni} \\ D_{in} & D_{ii} \end{bmatrix} \begin{bmatrix} T_1 & 0 \\ T_2 & T_3 \end{bmatrix} \begin{Bmatrix} x_w \\ x_D \end{Bmatrix} \\ &\quad + \begin{bmatrix} \rho\omega^2 T_1^T [N]^T [H]^{-1} [G] T_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{Bmatrix} x_w \\ x_D \end{Bmatrix} \\ &\quad = \begin{bmatrix} T_1 & 0 \\ T_2 & T_3 \end{bmatrix}^T \begin{Bmatrix} f_n \\ fi \end{Bmatrix}. \end{aligned} \tag{26}$$

The first term of the above equation represents the assembled stiffness matrix derived by the global system, and the second term is associated with the acoustic loading matrix in terms of the global system. Obviously, this loading matrix is symmetric. In Eq. (26), only the T_1 matrix is necessary during the computing process for the acoustic loading, while the first term of the multiplication of the three matrices is obtained directly by the global coordinate formulation. Equation (18) is used as numerical demonstrations in the present paper. To verify the proposed numerical method, we derive an analytical solution of a spherical thin shell with the semisphere immersed in fluid for numerical comparisons.

VI. ANALYTIC SOLUTION OF A SPHERICAL THIN SHELL WITH A SEMI-SPHERE IMMERSED IN FLUID

Figure 3 illustrates the geometry and loading force of the shell. The sphere has a lower semi-sphere submerged in the fluid. The characteristic function expansion of the acoustic field vibrating at circular frequency ω is[18]

$$p = \sum_{m=0}^{\infty} a_m L_{2m+1}(\cos\theta) h_{2m+1}(kr), \quad \frac{\pi}{2} \leqslant \theta \leqslant \pi, \tag{27}$$

where a_m is the expansion coefficient, L_{2m+1} denotes Legendre functions of order $2m+1$, and h_{2m+1} represents spherical Hankel functions of order $2m+1$. The elevation angle ranges from $\pi/2$ to π where $\pi/2$ corresponds to the fluid surface and π represents the bottom apex. The Legendre functions of the above expansion are odd orders owing to the boundary condition that the Legendre functions at $\theta = \pi/2$ are zero for the water surface. The set of odd-order Legendre functions forms a complete function space for

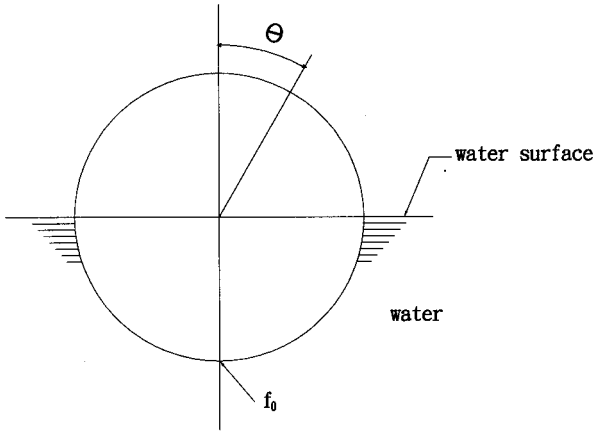


FIG. 3. A spherical thin shell with the lower semi-sphere immersed in water.

$\pi/2 \leqslant \theta \leqslant \pi$ with zero boundary condition at $\theta=\pi/2$ and finite values at $\theta=\pi$. Since the response is symmetric about the circumferential direction, the shell's displacements are described by the normal displacement w and tangential displacement u. These two displacements are expanded as[19]

$$w(\eta)=\sum_{n=0}^{\infty} w_n L_n(\eta),$$
$$u(\eta)=\sum_{n=0}^{\infty} u_n(1-\eta^2)^{1/2}\frac{dL_n}{d\eta}, \tag{28}$$

where w_n and u_n are the corresponding expansion coefficients, L_n is a Legendre function of order n, and $\eta=\cos\theta$. These displacements satisfy the spherical shell equation,

$$L_{uu}u+L_{uw}w+\Omega^2 u=0,$$
$$L_{wu}u+L_{ww}w+\Omega^2 w=-p\frac{a^2(1-\nu^2)}{Eh} + f_0\delta(\eta+1)\frac{(1-\nu^2)}{2\pi Eh}, \tag{29}$$

where the operators L_{uu}, L_{uw}, L_{wu}, and L_{ww} are defined in Ref. 19, $\Omega=\omega a/c_p$, $c_p^2=E/\rho_s(1-\nu^2)$, ν is the Poisson ratio of the shell material, ρ_s denotes the density of the shell, E is the Young's modulus, and h and a represent the thickness of the shell and its radius. The delta function $\delta(\eta+1)$ indicates a point force applied at $\eta=-1(\theta=\pi)$. The right-hand side of the second equation is the surface pressure load and the external point force exerted at the bottom apex. The surface pressure is related to the normal displacement w oscillating at circular frequency ω by the momentum equation,

$$\rho_0\frac{\partial}{\partial t}\left(\frac{\partial w}{\partial t}\right)=-\rho_0\omega^2 w=-\frac{\partial p}{\partial r} \quad \text{at } r=a, \quad 0\leqslant\eta\leqslant -1, \tag{30}$$

where the variable η ranges from 0 to -1 because the wetted surface varies θ from $\pi/2$ to π. Equations (29) and (30) constitute the coupled equations for the unknown variables u, w, and p. These coupled equations are solved by substituting the series expansions in Eqs. (27) and (28) into these equations. Equation (28) is the characteristic function expansions for the operators defined in Eq. (29), leading to

$$\sum_{n=0}^{\infty}(A_n u_n+B_n w_n)(1-\eta^2)^{1/2}\frac{dL_n(\eta)}{d\eta}=0,$$

$$\sum_{n=0}^{\infty}(\alpha_n u_n+\beta_n w_n)L(\eta) = \begin{cases} -\dfrac{a^2(1-\nu^2)}{Eh}\displaystyle\sum_{m=0}^{\infty} a_m L_{2m+1}(\eta)h_{2m+1}(ka) \\ \qquad +f_0\delta(\eta+1)\dfrac{(1-\nu^2)}{2\pi Eh}, \qquad \text{for } -1\leqslant\eta\leqslant 0, \\ 0, \qquad \text{for } 0\leqslant\eta\leqslant 1, \end{cases} \tag{31}$$

$$\rho_0\omega^2\sum_{n=0}^{\infty} w_n L_n(\eta)=\sum_{m=0}^{\infty} a_m L_{2m+1}(\eta)h'_{2m+1}(ka)k, \quad \text{for } -1\leqslant\eta\leqslant 0,$$

where h'_{2m+1} denotes the derivative of h_{2m+1} and the coefficients A_n, B_n, α_n and β_n are identified as

$$A_n=\Omega^2-(1+\beta^2)(\nu+\lambda_n-1),$$
$$B_n=-[\beta^2(\nu+\lambda_n-1)+(1+\nu)],$$
$$\alpha_n=-\lambda_n[\beta^2(\nu+\lambda_n-1)+(1+\nu)], \tag{32}$$
$$\beta_n=\Omega^2-2(1+\nu)-\beta^2\lambda_n(\nu+\lambda_n-1),$$

in which $\beta^2=\frac{1}{12}(h/a)^2$ and $\lambda_n=n(n+1)$. In the derivations of the first two equations of Eq. (31), the following identity has been used to simplify the expressions,

$$\frac{d}{d\eta}\left[(1-\eta^2)^{1/2}\frac{dL_n}{d\eta}\right]=-n(n+1)L_n. \tag{33}$$

The first equation of Eq. (31) becomes

$$A_n u_n+B_n w_n=0, \quad \text{for } n=0,1,2,\ldots. \tag{34}$$

The algebraic relationships for u_n, w_n, and a_m associated with the second equation of Eq. (31) become apparent when one realizes that the set of Legendre functions L_n, $n=0,1,2,3,\ldots$, forms a complete function space for η defining between ± 1. Thus, using the orthogonal condition for L_n by multiplying L_n on both sides of the equation and integrating over η between ± 1 yield

$$\alpha_n u_n+\beta_n w_n=\frac{2n+1}{2}\left(-\frac{a^2(1-\nu^2)}{Eh}\right)\times\sum_{m=0}^{\infty} a_m C_{nm}h_{2m+1}(ka) + \frac{2n+1}{2}\cdot\frac{(1-\nu^2)}{2\pi Eh}(-1)^n f_0, \quad \text{for } n=0,1,2,\ldots, \tag{35}$$

where the C_{nm} is identified as

$$C_{nm}=\int_{-1}^{0} L_n(\eta)L_{2m+1}(\eta)\,d\eta. \tag{36}$$

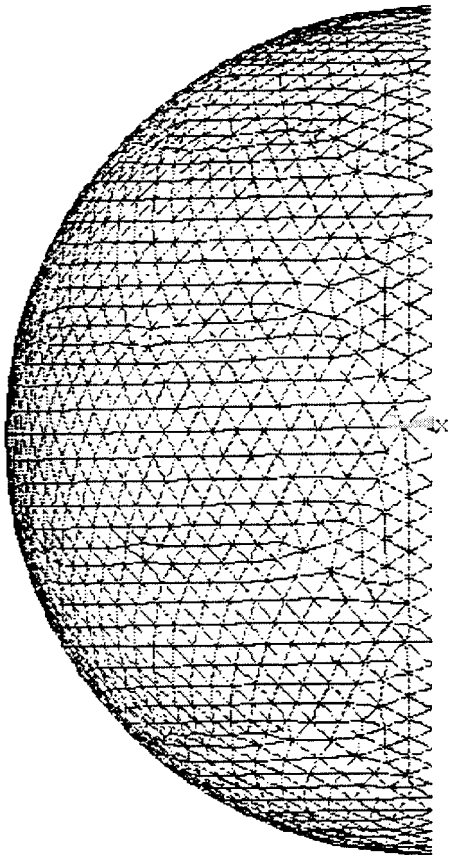


FIG. 4. Finite element mesh of a half-spherical shell.

Similarly, the set of the odd Legendre functions L_{2m+1}, $m=0,1,2,...$, spans a complete function space for η ranging from 0 to -1 with the boundary conditions that a zero value and finite value are required at $\eta=0$, and -1, respectively. The equation for w_n and a_m for the third equation of Eq. (31) emerges when the orthogonal conditions for the set of L_{2m+1} is applied, that is to multiply L_{2m+1} on the both sides of the third equation of Eq. (31) and to integrate over η between -1 to 0. This leads to

$$\frac{a_m h'_{2m+1}(ka)}{[2(2m+1)+1]ka} = \sum_{n=0}^{\infty} w_n C_{nm}, \quad m=0,1,2,\ldots. \tag{37}$$

Equations (34), (35), and (37) determine the expansion coefficients u_n, w_n and a_m by solving them simultaneously. Once the coefficients are solved, the displacements of the shell and the pressure are computed accordingly by Eqs. (27) and (28).

VII. NUMERICAL RESULTS AND COMPARISONS WITH THE ANALYTICAL SOLUTIONS

Figure 4 displays the finite element mesh of a half spherical shell. There are 1751 nodal points and 3404 elements. An external point force of one Newton is applied at the lower apex of the shell. Only half of the shell is modeled owing to the symmetries of the geometry, external forces, and acoustic loading. The boundary conditions for the nodal points on the symmetric plane of the shell, which form a closed circle of nodal points, are null normal displacements to the plane and null rotations on the tangential direction of the closed circle, as well as the direction normal to the spherical surface. The acoustic loading on the shell is one quarter sphere surface contacting with a water fluid, which is described by Eq. (6) with which the Green's function is given by Eq. (1). The function $\Omega(x)$ is provided in Eqs. (3)–(7) according to the locations of the surface field points on the surface. The coupled BEM/FEM matrix equation is computed in Eq. (18) with unknown nodal variables $\{x_n\}$ and $\{x_i\}$. The material constants selected are Young's modulus $E=2.07\times10^{11}$ Pa (Newton/square meter), Poisson ratio $\nu=0.3$, shell density $\rho_S=7669$ kg/m^3, the water density 1000 kg/m^3, and sound speed of the water $c=1524$ m/s. The radius of the spherical shell is 1 m and the thickness to radius ratio is 0.03. Equation (18) is a symmetric, banded matrix equation in which the acoustic loading matrix is regarded as a large element connected with each structural dynamic stiffness element whose orientation is normal to the wetted surface. An efficient means of storing the matrix elements is a variable-band (skyline) method in which a scheme to renumber nodal points is used to minimize the average bandwidth and to increase the computational speed.[20] Two vibration frequencies $ka=1.6$ and $ka=2.4$ are selected to compare numerical results with the analytical solutions derived in the preceding section.

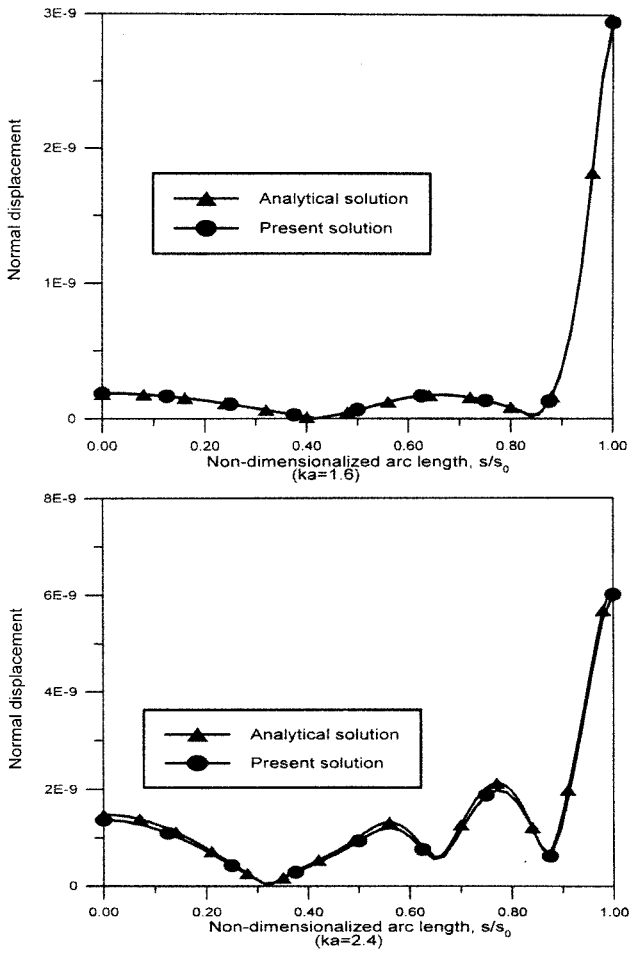


FIG. 5. Comparison of normal displacements along an arc length of the sphere with an external force exerted on the bottom apex vibrating at frequencies $ka=1.6,2.4$.

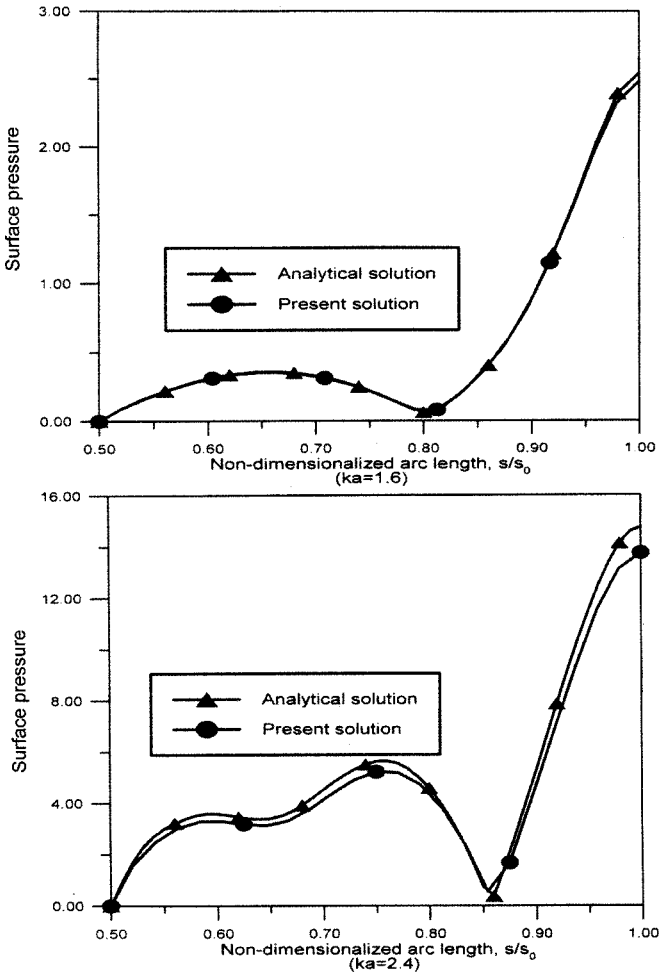


FIG. 6. Comparison of surface pressures for vibrating frequencies $ka = 1.6, 2.4$.

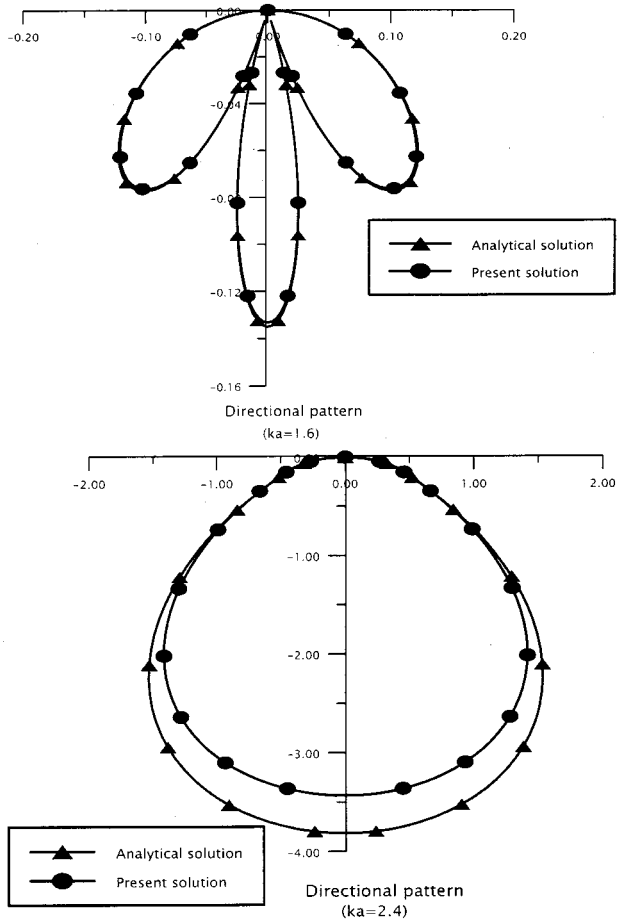


FIG. 7. Comparison of directional pattern of far field pressures for $ka = 1.6, 2.4$.

The expansion terms used in Eqs. (27) and (28) are selected such that the amplitudes of the coefficients solved by the coupled equations of Eqs. (34), (35), and (37) converge to zero. Since the pressure expansion of the Legendre functions L_{2m+1} is odd orders, a reasonable selection of natural mode expansions for shell displacements in Eq. (28) is twice the number of pressure expansions used in Eq. (27). Because the response is axisymmetric, the results presented herein are along an arc length measured from the top apex. Figure 5 compares the normal displacements obtained numerically and analytically where the nondimensional arc length is the ratio of arc distance from the top apex S to the total arc length S_0. According to our results, a good correlation is found for $ka = 1.6$ and minor discrepancy for $ka = 2.4$.

Once the displacements x_n and x_i are obtained from Eq. (18), the surface pressure is computed according to Eq. (17). Figure 6 compares the computed surface pressure with the analytical solution, where the nondimensional arc length only accounts for the lower sphere immersed in water for which the abscissa ranges from 0.5 to 1.0. Again, a close agreement for $ka = 1.6$ and minor discrepancies for $ka = 2.4$ are observed. The pressure at a field point is computed by Eq. (2), where the normal pressure gradient relates to the normal displacement x_n by Eq. (16). Herein, directional patterns are presented for a far field by factoring out the simple source term e^{ikr}/r, where r is a very large distance from the origin of the sphere. Figure 7 compares the numerical far field solution with the analytical solution versus the angle θ, in which 90 degrees corresponds to water surface and 180

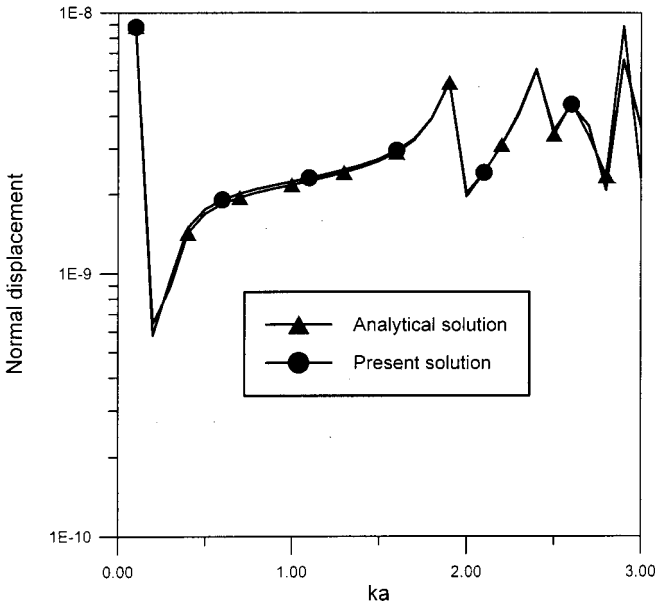


FIG. 8. Frequency dependence for the normal displacements on the bottom apex for using the numerical and analytical methods.

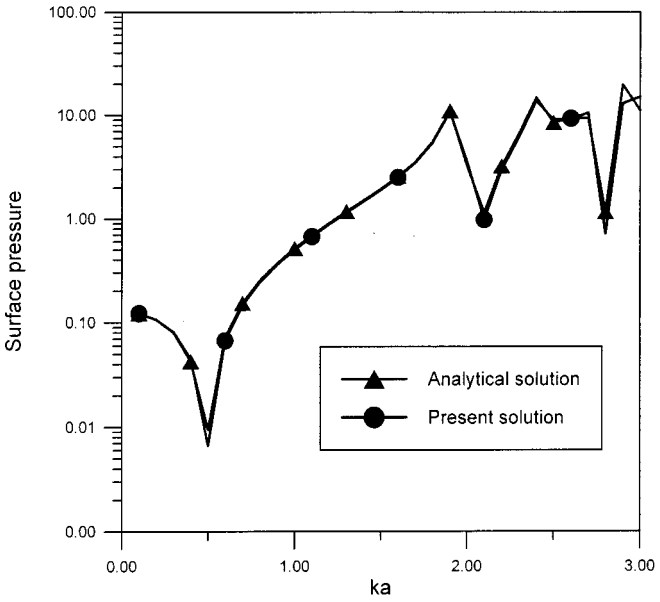


FIG. 9. Frequency dependence for the surface pressures computed on the bottom apex for using the numerical and analytical methods.

degrees denotes the bottom of the sphere. Similar correlations of numerical results with the analytical solutions are observed in Figs. 5 and 6.

Finally, Figs. 8 and 9 are the frequency scanning of the normal displacement and surface pressure at the bottom apex obtained by the numerical and analytical methods. Good correlations are achieved at low frequencies and minor discrepancies are observed with increasing frequencies.

VIII. COMPUTATIONAL ASPECTS

The numerical method is implemented on a Pentium II PC. The major computational loads occur at forming the acoustic loading matrix, $\rho_0\omega^2[N]^T[H]^{-1}[G]$ in Eq. (18) and the inversion of Eq. (18). The former execution requires a full complex matrix inversion and multiplication. The memory used is dominated by storing the matrix elements in Eq. (18). The efficiency of storing the elements and inversion of the matrix depends on the bandwidth. Rearranging the ordering of nodal points based on Cuthill–Mckee algorithm[20] reduces the bandwidth. Table I lists the total degrees of freedom, the averaged wavefront, and the maximum wavefront of the present numerical example. The averaged wavefront is substantially smaller than the maximum wavefront, indicating the effectiveness of the proposed numerical method.

TABLE I. Bandwidth of the coupled equation [Eq. (15)] for the numerical example.

Total degrees of freedom	10 218
Average wavefront of the matrix (after renumbering)	165
Maximum wavefront of the matrix (after renumbering)	252

IX. CONCLUSION

This work presents a numerical method for computing the response and acoustic radiation for semi-immersed elastic structures by using the coupled BEM/FEM. The structure and external applied forces are symmetric about a symmetric plane so that the responses are symmetric about the plane. A Green's function that satisfies the pressure release boundary condition at fluid surface and null normal velocity on the symmetric plane is applied to a boundary integral equation. In addition, the surface domain of the integral equation is defined on the half of the wetted surface. Particular emphasis is placed on the field point approaching the surface domain for the limiting surface integral value of normal derivative of the Green's function. The value depends on the field point located at the intersection of the fluid surface with the vibrating surface, at the intersection of the symmetric plane with the vibration surface, or on the vibrating surface. The acoustic loading matrix derived from the discretized boundary integral equation is symmetric owing to a reciprocity principle for the surface acoustics identified under the conditions of pressure release and null normal velocity on the fluid surface and the symmetric plane. This symmetric formulation in conjunction with structural stiffness matrices oriented to the normal direction of the wetted surface provides a symmetric banded matrix equation, thus making the technique of variable-band storage and the associated matrix solution available, and markedly increasing numerical efficiency. Effectiveness of the numerical method is verified by an analytical solution of a spherical thin shell semi-immersed in water. A good correlation is found at low frequencies and minor discrepancies with increasing frequency.

ACKNOWLEDGMENT

The author would like to thank the National Science Council of the Republic of China, Taiwan for financially supporting this research under Contract No. NSC 89-2611-E-019-010.

[1] G. C. Everstine and F. M. Henderson, "Coupled finite element/boundary approach for fluid-structure interaction," J. Acoust. Soc. Am. **87**, 1938–1947 (1990).
[2] R. A. Jeans and I. C. Mathews, "Solution of fluid-structure interaction problems using a coupled finite element and variational boundary element technique," J. Acoust. Soc. Am. **88**, 2459–2466 (1990).
[3] R. A. Jeans and I. C. Mathews, "A unique coupled boundary element/finite element method for the elastoacoustic analysis of fluid-filled thin shell," J. Acoust. Soc. Am. **94**, 3473–3479 (1993).
[4] A. F. Seybert, T. W. Wu, and W. L. Li, "A coupled FEM/BEM for fluid-structure interaction using ritz vectors and eigenvectors," J. Vibr. Acoust. **115**, 152–158 (1993).
[5] D. F'elix, "Numerical model for acoustic radiation for an immersed elastic structure and application to a thin cylindrical shell," J. Acoust. Soc. Am. **95**, 3332–3338 (1994).
[6] J. S. Patel, "Radiation and scattering from an arbitrary elastic structure using consistent fluid structure formulation," Comput. Struct. **9**, 287–291 (1978).
[7] D. T. Wilton, "Acoustic radiation and scattering from elastic structures," Int. J. Num. Mec. Eng. **13**, 123–138 (1978).
[8] I. C. Mathews, "Numerical techniques for three-dimensional steady-state fluid-structure interaction," J. Acoust. Soc. Am. **79**, 1317–1325 (1986).
[9] P. T. Chen, S. H. Ju, and K. C. Cha, "A symmetric formulation of coupled BEM/FEM in solving responses of submerged elastic structures for large degrees of freedom," J. Sound Vib. **233**(3), 407–422 (2000).

[10] T. Mikami and J. Yoshimura, ''The collocation method for analyzing free vibration of shells of revolution with either internal on external fluids,'' Comput. Struct. **44**, 343–351 (1992).
[11] M. Amabili, ''Flexural vibration of cylindrical shells partially coupled with external and internal fluids,'' J. Vibr. Acoust. **119**, 476–484 (1997).
[12] M. Amabili, ''Vibrations of circular tubes and shells filled and partially immersed in dense fluids,'' J. Sound Vib. **221**, 567–585 (1999).
[13] A. F. Seybert and T. W. Wu, ''Modified Helmholtz integral equation for body sitting on an infinite plane,'' J. Acoust. Soc. Am. **85**, 19–23 (1989).
[14] O. C. Zienkiewicz, *The Finite Element Method* (McGraw-Hill, New York, 1997), Chaps. 2 and 13.
[15] P. T. Chen and J. H. Ginsberg, ''Complex power, reciprocity, and radiation for submerged bodies,'' J. Acoust. Soc. Am. **98**, 3343–3351 (1995).
[16] I. S. Sokolnikoff, *Mathematical Theory of Elasticity* (McGraw-Hill, New York, 1995), pp. 390–393.
[17] A. D. Pierce, *Acoustics, An introduction of its Physical Principles and Applications* (McGraw-Hill, New York, 1981), pp. 195–198.
[18] P. M. Morse, *Vibration and Sound* (Acoustical Society of America, Woodbury, NY, 1976), VII. 27.
[19] M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction* (Acoustical Society of America, Woodbury, NY, 1993), Chap. 7.
[20] A. George and J. W.-H. Liu, *Computer Solution of Large Sparse Positive Definite System* (Prentice–Hall, Englewood Cliffs, NJ, 1981), pp. 48–90.

Estimation of sound power of baffled planar sources using radiation matrices

Mingsian R. Bai[a] and Mingchun Tsao
Department of Mechanical Engineering, National Chiao-Tung University, Hsin-Chu 30010, Taiwan, Republic of China

(Received 2 October 2001; revised 24 May 2001; accepted 11 June 2002)

A technique is presented in this paper for the analysis of acoustic radiation of baffled planar sources. This method is based on the concept of the radiation matrix which is a representation of the sound field radiated by complex sources. Both sound power and far-field pressure can be estimated using this method. Numerical simulation and experimental investigation were carried out to verify the proposed method. In comparison with the conventional pressure-based method ISO 3745, the proposed technique produced more accurate estimation of sound power. With comparable number of measuring positions as the conventional methods, the proposed approach does not require special measuring environments. Computation complexity can be significantly reduced if only dominant radiation modes in the low-frequency range are retained. Estimation of far-field pressure can also be obtained using the proposed method.
[DOI: 10.1121/1.1499133]

PACS numbers: 43.20.Rz, 43.40.Rj [MO]

I. INTRODUCTION

Sound power is an important index in acoustic radiation analysis because it is a single global quantity which can be used to characterize the strength of the sound generated by a source. Estimation of sound power is also a common practice of identifying dominant sources prior to a noise control project. Many standard methods are available for sound power estimation.[1] Some of them are based on pressure measurement; the others are based on intensity measurement. Some of them require free-field environment; the others require diffuse field environment.

In this paper, a technique different from the earlier methods is presented. This method is based on the concept of *radiation matrix* which is a representation of the sound field radiated by complex sources. More precisely, radiation modes are ''vibration'' patterns on the surface of a structure that radiate orthogonal sound fields into the space.[2–6] In the past, the ideas of radiation matrix and radiation modes have mainly been employed in the application of *Active Structural Acoustical Control* (ASAC)[2,3] to control low-frequency noise radiated by vibrating surfaces. In this paper, however, this idea of radiation modes is exploited to deal with a different problem—estimation of sound power and far-field pressure radiated from complex baffled planar sources. In addition, an experimental procedure using the radiation mode concept is established in this work, in an attempt to achieve improved estimation over conventional approaches.

Complexity of the calculation can be further reduced if only the dominant radiation modes are retained in the computation. A crucial step of this procedure is the replacement of the frequency-dependent radiation modes by frequency-independent ones.[7,8] In ASAC, the ''radiation filter'' consists of a constant modal filtering matrix, cascaded with frequency-dependent radiation efficiency filters.

As demonstrated by Berkhoff,[9] the input to the radiation filter can be based on either surface velocity data or near-field pressure data. This paper adopts primarily the former approach because, from our experience, it provides more reliable results than the latter approach. The pressure-based methods are thus omitted for brevity. With comparable number of measuring positions as the conventional methods, the proposed techniques do not require special measuring environment, e.g., a reverberant room or an anechoic room because only surface velocity is required as the input data.

This paper mainly focuses on baffled planar sources. Two kinds of sound sources, a baffled rigid piston and a baffled point-driven flexible plate, are chosen as the test objects in the simulation. The surface vibration of the plate is calculated by using ANSYS.[10] Sound power and far-field pressure are calculated. The proposed techniques are verified by numerical simulation as well as experiments. A baffled vibrating plate is chosen as the test source. The proposed methods are also compared with the ISO 3745 (Ref. 11) which is a widely accepted standard for sound power estimation.

II. RADIATION MATRIX-BASED TECHNIQUES

A. Velocity-based full radiation mode method (VFRM)

For a sound source with surface A, the radiated sound power W can be calculated by

$$W = \frac{S}{2}\mathrm{Re}\{\mathbf{v}^H\mathbf{p}\}, \tag{1}$$

where S is the area of each element, $\mathbf{p}$ and $\mathbf{v}$ are pressure vector and velocity vector, respectively, on the surface, and the superscript ''H'' denotes the Hermitian transpose operator.

On the surface of the source, the pressure vector $\mathbf{p}$ can be related to the velocity vector $\mathbf{v}$ by the following equation:

[a]Electronic mail: msbai@cc.nctu.edu.tw

$$\mathbf{p}=\mathbf{Z}\mathbf{v}, \tag{2}$$

where $\mathbf{Z}$ is the radiation impedance matrix that in general can be obtained using numerical methods such as boundary element methods.[12] There exist a number of different definitions of impedance in acoustics. The matrix $\mathbf{Z}$ is termed radiation matrix simply because it relates the pressure and velocity distributions on a acoustically radiating surface of a source. The dimension of the matrix $\mathbf{Z}$, however, is the same as specific acoustic impedance, i.e., pressure/velocity. For a baffled planar source, however, the matrix $\mathbf{Z}$ can simply be calculated by using the following formula:[9]

$$\mathbf{Z}=\rho_0 c\begin{bmatrix} d & -\frac{jkS}{2\pi}\frac{e^{ikr_{12}}}{r_{12}} & \cdots & -\frac{jkS}{2\pi}\frac{e^{ikr_{1N}}}{r_{1N}} \\ -\frac{jkS}{2\pi}\frac{e^{ikr_{21}}}{r_{21}} & d & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ -\frac{jkS}{2\pi}\frac{e^{jkr_{N1}}}{r_{N1}} & \cdots & \cdots & d \end{bmatrix}, \tag{3}$$

where

$$d=\frac{1}{2}\left(k\sqrt{\frac{S}{\pi}}\right)^2-j\frac{8}{3\pi}\left(k\sqrt{\frac{S}{\pi}}\right), \tag{4}$$

ρ_0 is the air density, c is the sound speed, $k=\omega/c$ is the wave number, and $r_{mn}=r_{nm}$ is the distance from the element m to the element n, $1\leqslant m,\ n\leqslant N$. Equation (4) is based on a numerical approximation of the Rayleigh integral in which the field point approaches the source surface. The off-diagonal elements are contributions due to point sources of the elemental radiators, with given volume velocity which lead to pressure on the other elements on the surface. For the diagonal terms, the field point and the source point may coincide, which results in a singular problem. The difficulty can be circumvented if the integration is not over a square area but over a circular area. The approach leads to an approximation of the diagonal terms. Note that the expressions of the diagonal terms are somewhat different from that in Ref. 9.

Equations (2) and (3) lead to

$$W=\mathbf{v}^H\mathbf{R}\mathbf{v}, \tag{5}$$

where the radiation resistance matrix $\mathbf{R}=S\,\mathrm{Re}\{\mathbf{Z}\}/2$ can be calculated by

$$\mathbf{R}=\frac{\omega^2\rho_0 S^2}{4\pi c}\begin{bmatrix} 1 & \frac{\sin(kr_{12})}{kr_{12}} & \cdots & \frac{\sin(kr_{1N})}{kr_{1N}} \\ \frac{\sin(kr_{21})}{kr_{21}} & 1 & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\sin(kr_{N1})}{kr_{N1}} & \cdots & \cdots & 1 \end{bmatrix}. \tag{6}$$

Complexity of the computation can be further reduced if only the dominant modes are retained in the computation. This is accomplished by exploiting the fact that $\mathbf{R}$ is a real-symmetric and positive-definite matrix and the following eigenvalue decomposition (EVD) is always possible[13]

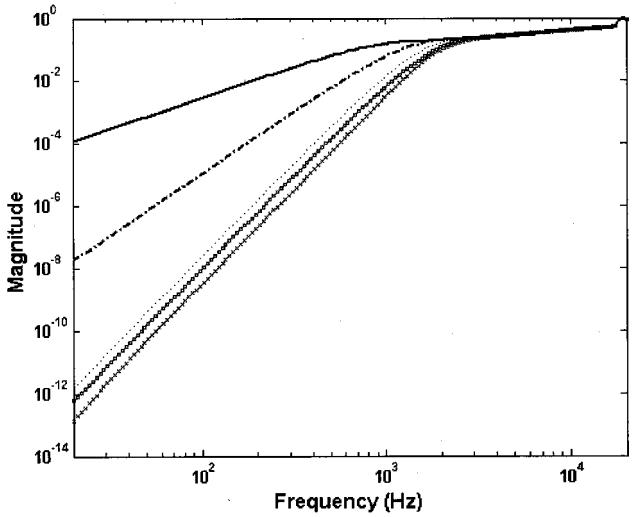


FIG. 1. The radiation efficiencies of the first six radiation modes of the baffled piston (—, first singular value; - - -, second singular value; -·-, third singular value; ···, fourth singular value, □, fifth singular value; +, sixth singular value).

$$\mathbf{R}=\mathbf{Q}^T\mathbf{\Lambda}\mathbf{Q}, \tag{7}$$

where $\mathbf{\Lambda}=\mathrm{diag}(\lambda_1,\lambda_2,...,\lambda_L)$ are eigenvalues that have the physical meaning as the radiation efficiencies, the radiation modal filter $\mathbf{Q}=[\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_L]$ is a group of orthogonal surface vibration patterns associated with the radiation modes. These radiation modes are associated with surface velocity distributions that produce orthogonal sound pressure modes which are generally different from the actual acoustical modes that satisfy boundary conditions. Although these radiation modes are always orthogonal in the mathematical sense, they are generally different from the structural vibration modes. In practice, only a limited number, say L, of radiation modes that are efficient in low frequencies. The computation can be simplified by considering only these dominant modes. From Eqs. (5) and (7),

$$W=\mathbf{y}^H\mathbf{\Lambda}\mathbf{y}=\sum_{i=1}^{L}\lambda_i|y_i|^2, \tag{8}$$

where vector $\mathbf{y}=\mathbf{Q}\mathbf{v}$. Each diagonal element represents the radiation efficiency corresponding to each radiation mode. Note that the radiation efficiency can be related to radiation resistance (real part of radiation impedance) normalized with $\rho_0 c$. Modes with large λ_i values are the modes capable of propagating efficiently into far field, whereas modes with small λ_i values are the *evanescent modes* existing only in near field. Figure 1 shows the radiation efficiencies of the first six radiation modes for a baffled rigid piston with 0.1 m radius. Note that, in the low-frequency range ($ka<1$, a is the radius), only two to six modes are dominant. At the extreme ($ka\ll 1$), the first mode (the volume velocity mode) is the only efficient mode. At the other extreme ($ka\gg 1$), all radiation modes contribute equally to the radiation with unity efficiency. This motivates the use of low-order radiation modes

to represent the sound field for low frequency range such that the computation can be simplified.

However, the radiation modal matrix is frequency dependent, which requires the computation procedure to be repeated for each frequency. A remedy to this difficulty lies in the fact that the radiation modes are slowly varying with respect to frequency and the low-frequency modes are ''nested'' by the high-frequency modes.[2] Figures 2(a) and (b) show the radiation modes of a baffled piston with a radius 0.1 m for $kl=0.1$ and $kl=2$. Now, suppose sound power is to be estimated by measuring the surface velocity at M positions on a source. The size of matrix $\mathbf{R}$ will be $M\times M$, which calls for the need of M^2 filters. Instead of using frequency-dependent matrices, the computation load can be reduced by using the radiation modal matrix, $\mathbf{Q}_{\max}$, at the highest frequency of interest. Note that $\mathbf{Q}_{\max}$ is a constant matrix with the dimension $M\times L$, where $L\ll M$ represents the number of dominant radiation modes. On the other hand, the radiation efficiency of the ith mode can be approximated as

$$\lambda_i(j\omega)=\mathbf{q}_{i,\max}\mathbf{R}(j\omega)\mathbf{q}_{i,\max}^T, \tag{9}$$

where $\mathbf{q}_{i,\max}$ is the ith column of $\mathbf{Q}_{\max}$, λ_i is still a function of frequency and generally exhibits a high-pass behavior. Hence, sound power can be estimated by the velocity-based dominant radiation mode method (VDRM):

$$W=\mathbf{v}^H\tilde{\mathbf{R}}\mathbf{v}, \tag{10}$$

where the matrix $\tilde{\mathbf{R}}$ is approximated as

$$\tilde{\mathbf{R}}=\mathbf{Q}_{\max}^T\tilde{\mathbf{\Lambda}}\mathbf{Q}_{\max}, \tag{11}$$

where $\tilde{\mathbf{\Lambda}}=\mathrm{diag}[\lambda_1(j\omega),\lambda_2(j\omega),\ldots\lambda_L(j\omega)]$. In this setting, only a constant matrix $\mathbf{Q}_{\max}$ and L filters are needed in the computation.

B. Modified velocity-based dominant radiation modes method (MVDRM)

The VDRM is based on the idea of dominant radiation mode which is only true in low frequencies. In high frequencies, direct use of VDRM will give rise to errors because all modes contribute equally to sound radiation. To correct the problem for high frequencies, we use another approach to estimate the sound power. For high frequencies, e.g., $ka>4$, the waves radiated from a planar source behave like plane waves and Eq. (5) should be modified into

$$W=\rho_0 cA\frac{\sum_{i=1}^{N}\nu_{\mathrm{rms},i}^2}{N}, \tag{12}$$

where ρ_0 is the air density, A is the total surface area of the source, c is the sound speed, ν_{rms} is the root-mean-square normal velocity, and N is the number of measurements.

C. Far-field pressure

Apart from sound power, far-field sound pressure can also be calculated by using the idea of radiation modes. The Rayleigh integral can be discretized into a matrix equation

$$\mathbf{p}=\mathbf{E}\mathbf{v}, \tag{13}$$

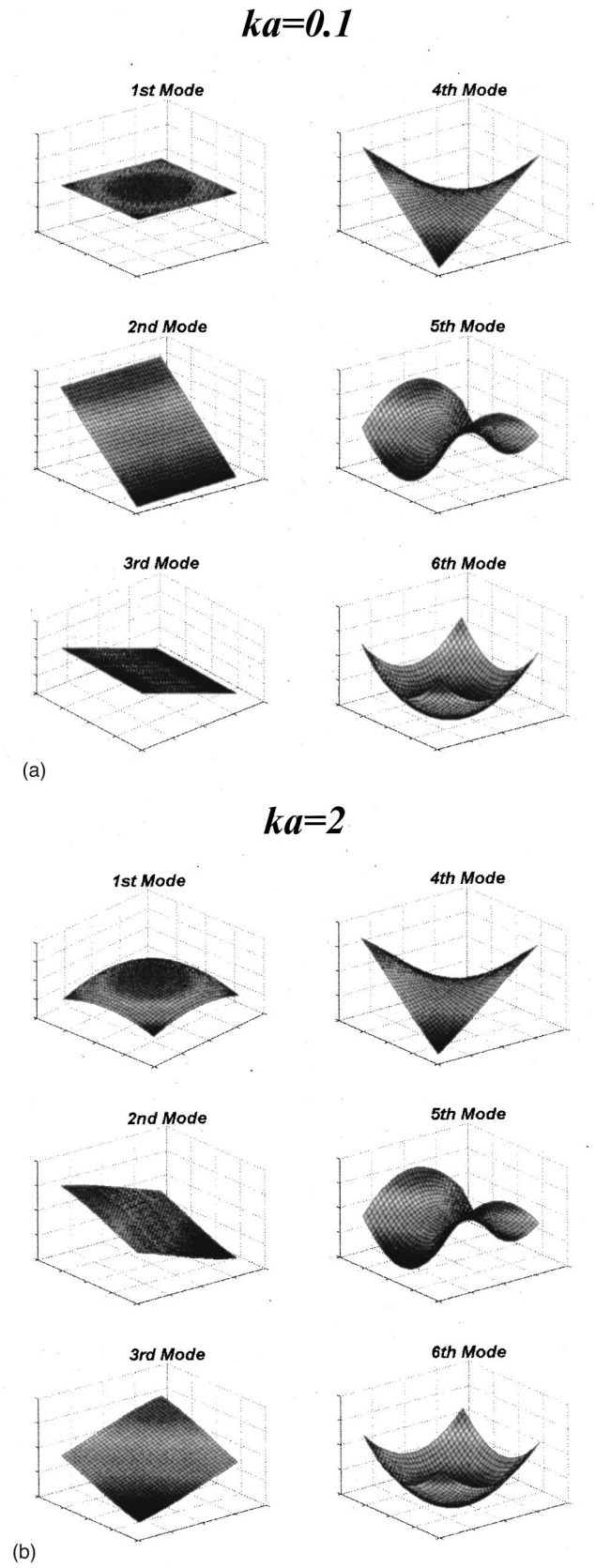


FIG. 2. The first six radiation modes. (a) $ka=0.1$; (b) $ka=2$.

where $\mathbf{p}$ is the vector of far-field pressure, $\mathbf{E}$ is the propagation matrix, and $\mathbf{v}$ is the vector of surface velocity.

For a baffled planar radiator, this propagation matrix $\mathbf{E}$ can be approximated as[9]

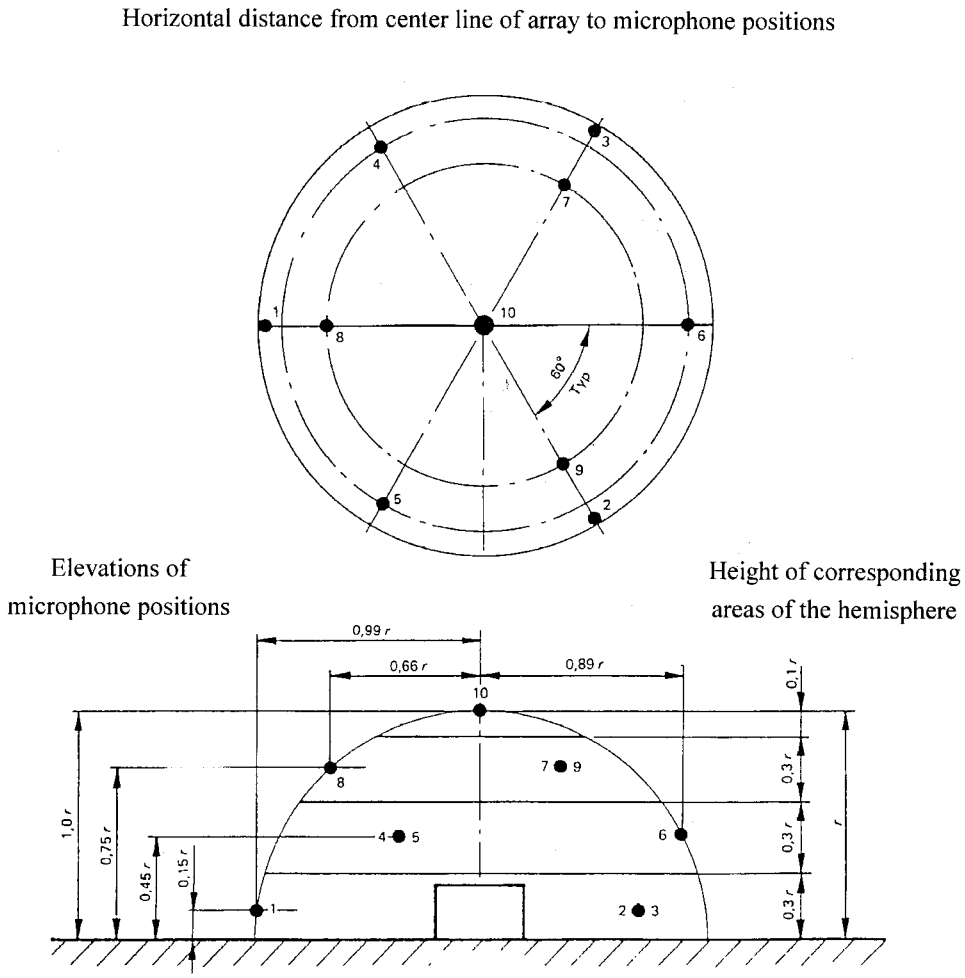


FIG. 3. Microphone positions on a hemispherical surface surrounding a source whose sound power level is to be measured according to ISO 3745 (cited from Ref. 11).

$$\mathbf{E}=j\frac{\rho_0 ckS}{2\pi}\begin{bmatrix} \frac{e^{-jkr_{11}}}{r_{11}} & \frac{e^{-jkr_{12}}}{r_{12}} & \cdots & \frac{e^{-jkr_{1N}}}{r_{1N}} \\ \frac{e^{-jkr_{21}}}{r_{21}} & \frac{e^{-jkr_{22}}}{r_{22}} & \cdots & \frac{e^{-jkr_{2N}}}{r_{2N}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{e^{-jkr_{M1}}}{r_{M1}} & \frac{e^{-jkr_{M2}}}{r_{M2}} & \cdots & \frac{e^{-jkr_{MN}}}{r_{MN}} \end{bmatrix}, \quad (14)$$

where r_{MN} is the distance from the elemental radiator N to the field point M.

Similar to power estimation, SVD[13] can be applied to the propagation matrix $\mathbf{E}$, i.e.,

$$\mathbf{E}=\mathbf{U}\mathbf{\Lambda}\mathbf{V}^H, \quad (15)$$

where the diagonal matrix $\mathbf{\Lambda}$ has singular values as its diagonal, $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices. In low-frequency range, only several dominant modes are needed to represent the propagation process.[14] That is,

$$\tilde{\mathbf{E}}=\tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}\tilde{\mathbf{V}}^H. \quad (16)$$

Thus, the far-field sound pressure $\mathbf{p}$ and $\mathbf{v}$ can be approximated as

$$\tilde{\mathbf{p}}=\tilde{\mathbf{E}}\mathbf{v}. \quad (17)$$

III. NUMERICAL AND EXPERIMENTAL INVESTIGATIONS

In this section, numerical simulation and experimental investigation were carried out to verify the proposed method.

A. Sound power estimation

1. Case 1—baffled rigid piston

In this example, we use a baffled rigid piston as a sound source to verify the proposed methods, VFRM, VDRM, and MVDRM. The piston source is chosen because its power can be calculated analytically.[15] The radius of the piston is 0.1 m which is excited with surface velocity $1\cdot e^{j\omega t}$ cm/s.

To justify the proposed techniques, a widely used ISO standard 3745 for sound power estimation in free-field environment is also employed in the comparison. The ISO procedure requires the mean value calculated from the pressure measurements at 10 positions on hemispherical surface surrounding the source of interest, as shown Fig. 3. The method is essentially based on the assumption of a point source, which is approximately true in the far field. Figure 4 compares the estimations of sound power obtained using the theoretical model and the ISO standard, respectively. Deviations from the theoretical value are found in high-frequency range (>1 kHz), using the ISO method. The total sound power within the band 20–16 kHz are found to be 88.1 dB and 96.3 dB for the theoretical value and ISO 3745, respectively. The

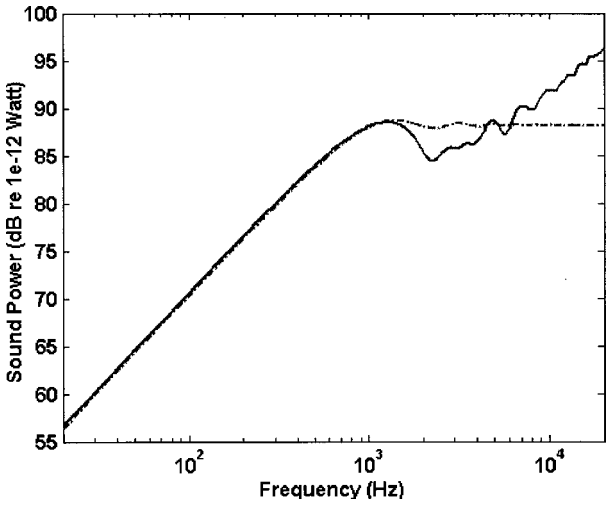


FIG. 4. The comparison of sound power radiated by the baffled piston between the theoretical calculation and ISO 3745 (-·-, theoretical value; —, ISO 3745).

errors in power estimation using ISO 3745 can be explained as follows. ISO 3745 is a essentially pressure-based method under the free-field condition. The measurement should be conducted in a sufficiently far field so that can be approximated by a point source field. However, even under the free-field condition, there are possibly three types of errors in the measurement: (1) the near field error, (2) a limited number of measuring points, and (3) the error in sound pressure measurement. A detailed investigation on these issues can be found in Ref. 16.

The proposed power estimation method requires velocity measurements at the grid points on the source surface. First, the sound power radiated by the piston is estimated by VFRM. A 4×4 rectangular mesh (16 elements) is used. Figure 5(a) compares the estimations of sound power obtained using the theoretical model and VFRM, respectively. The approximation appears close to the theoretical value in low-frequency range ($ka<4$, $f<1000$ Hz), while some deviations are found in high-frequency range.

Next, VDRM is applied to estimate the sound power for the same frequency range ($ka<4$, $f<1000$ Hz). Similar estimation of the radiated sound power is obtained by using only the first six radiation modes, as shown in Fig. 5(b).

The foregoing results indicate that VFRM and VDRM are able to give accurate estimation only in the low-frequency range. For the high-frequency range, the modified method MVDRM should be used to produce estimations with less error. Figure 5(c) compares the estimations of sound power obtained using the theoretical model and MVDRM, respectively. The total sound power within the band 20–16 kHz are found to be 88.1 dB, 86.5 dB, 86.1 dB, and 87.9 dB for the theoretical value, VFRM, VDRM, and MVDRM, respectively. From the result, MVFRM is found to produce the best estimation of sound power in both the low and high frequencies.

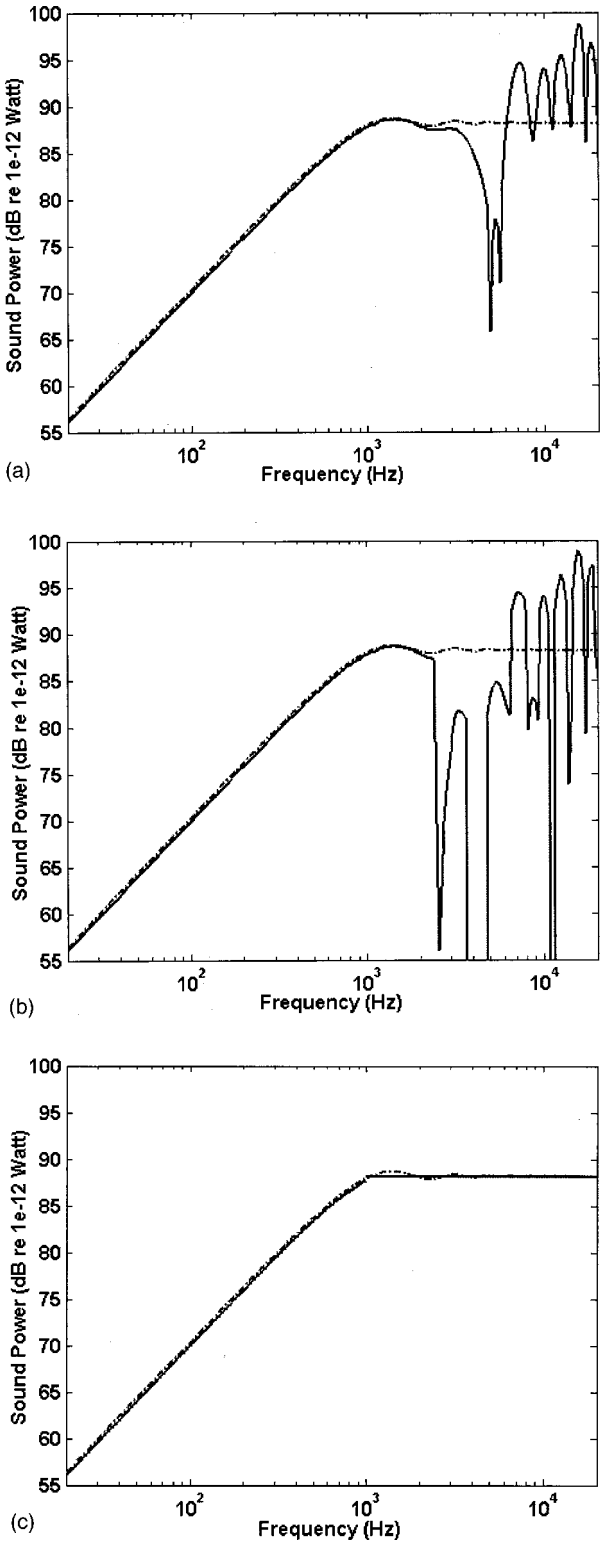


FIG. 5. The comparison of sound power radiated by the baffled piston between the theoretical calculation and the proposed methods. (a) The theoretical calculation versus VFRM (4×4 mesh) (-·-, theoretical value; —, VFRM); (b) the theoretical calculation versus VDRM by using the first six modes (4×4 mesh) (-·-, theoretical value; —, VDRM); (c) the theoretical calculation versus MVDRM (4×4 mesh) (-·-, theoretical value; —, MVDRM).

2. Case 2—vibrating plate

The rigid piston used for the last simulation has uniform surface velocity distribution. A more complex sound source, a vibrating flexible plate shall be used as the test sound source in this case. The parameters of the plate are area $=0.2\times0.15$ m^2, density $\rho=246$ kg/m^3, Young's modulus $E=2.28\times10^9$ N/m^2, Poisson ratio $\nu=0.33$, and thickness h

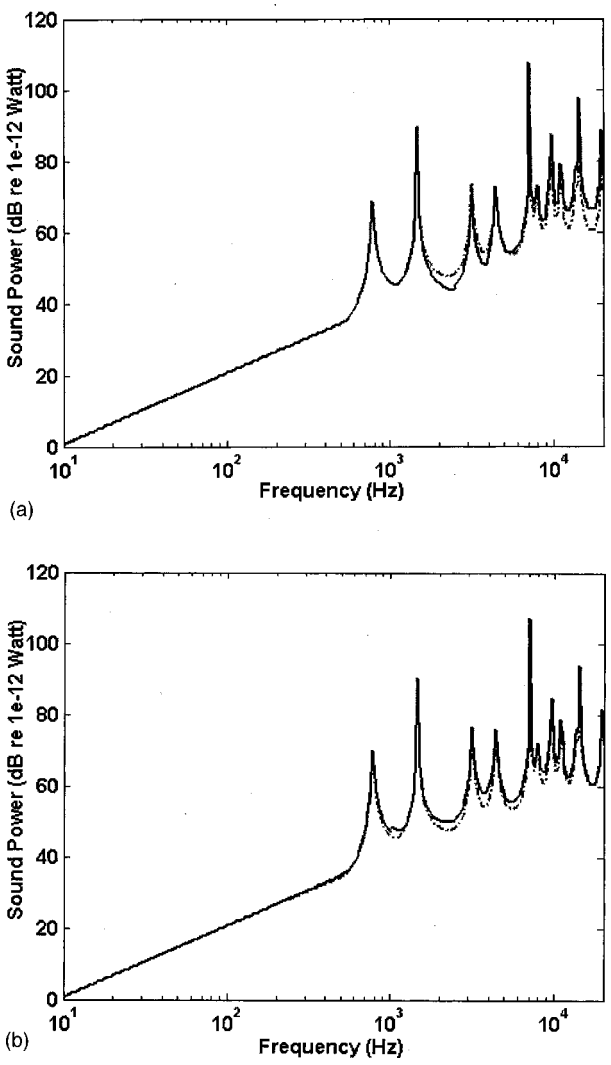


FIG. 6. The sound power radiated by the vibrating plate. (a) Theoretical calculation versus ISO 3745 (-·-, theoretical value; —, ISO 3745); (b) the theoretical calculation versus MVDRM by using the first six modes (3×3 mesh) (-·-, theoretical value; —, MVDRM).

$=1$ cm. The plate is excited at the center by a time-harmonic concentrated force, $0.5e^{j\omega t}N$. Four boundaries of the plate are assumed free. Finite element software, ANSYS, is employed to calculate the normal velocity distribution on the surface. Sound power is then estimated by using a 32×32 **R** matrix. Unlike rigid pistons, there is no analytical solution of sound power for plates, and thus we are content with the use of the fine mesh (32×32) as the ''theoretical'' solution.

Figure 6(a) compares the estimations of sound power obtained using the theoretical model and ISO 3745, respectively. In contrast to the piston, peaks due to resonance of the plate can be clearly seen in the spectrum. Deviations from the theoretical value are found in the high-frequency range (>1 kHz), using the ISO method.

Next, the modified method, MVDRM, is utilized to estimate the sound power radiated by the plate. Figure 6(b) compares the estimations of sound power obtained using the theoretical model and MVDRM, respectively. The total sound power within the band 20–16 kHz are found to be 77.6 dB, 76.2 dB, and 77.1 dB for the theoretical value, ISO 3745, and MVDRM, respectively. It is then concluded from the numerical comparison that, with comparable number of measurements (9 vs 10), MVDRM produces more accurate estimation of sound power than the ISO method.

B. The far-field radiation

In this section, calculation of far-field sound pressure using the proposed techniques will be examined, particularly in terms of propagation distance and the number of modes used in SVD of the matrix **E**.

The parameters of the baffled piston used in the simulations are identical to those mentioned earlier. The ''theoretical'' values of the piston can be calculated by suing a very fine mesh, e.g., 32×32. To examine the effect of propagation distance, a dimensionless parameter, the Fresnel number,[15] is defined as follows:

$$\mathrm{FN}=\frac{r\lambda}{a^2}, \tag{18}$$

where r is the axial distance from the piston surface to the field point, λ is the wavelength, and a is the radius of piston. Figure 7(a) compares the far-field sound pressure obtained using the theoretical model, **E** matrix based on 5×5 mesh with full modes, and **E** matrix based on 5×5 mesh with the first six modes, respectively, for $ka=0.5$ and $\mathrm{FN}=2$. The result of approach using the first six modes coincides with that using full modes, and so we found only a single dashed curve in the figure. Figure 7(b) compares the far-field sound pressure obtained using the theoretical model, **E** matrix based on 11×11 mesh with full modes, and **E** matrix based on 11×11 mesh with the first six modes, respectively, for $ka=30$ and $\mathrm{FN}=2$. As expected, to calculate the far-field sound pressure for a source with a large ka value requires a matrix **E** with higher dimension. Conversely, it takes only a few dominant modes to reconstruct far-field sound pressure for sources with a small ka value.

Like power estimation, we also apply the proposed technique to calculate the far-field sound pressure radiated by the vibrating plate. Finite element software, ANSYS, is employed to calculate the normal velocity distribution on the surface. Sound power is then estimated by numerically calculating the Rayleigh integral. Since there is no analytical solution of sound pressure for plates, we are content with the use of the fine mesh (32×32) as the ''theoretical'' solution. Figure 8(a) compares the far-field sound pressure obtained using the theoretical model, **E** matrix based on a 5×5 mesh with full modes, and **E** matrix based on a 5×5 mesh with the first six modes, respectively, for $ka=1.5$ and $\mathrm{FN}=5$, where a is the shorter width of the plate edges. Figure 8(b) compares the far-field sound pressure obtained using the theoretical model, **E** matrix based on a 11×11 mesh with full modes, and **E** matrix based on a 11×11 mesh with the first six modes, respectively, for $ka=20$ and $\mathrm{FN}=0.1$. As in the case of piston, to calculate the far-field sound pressure for a source with a large ka value requires a matrix **E** with higher dimension. Conversely, it takes only a few dominant modes to reconstruct far-field sound pressure for sources with a small ka value.

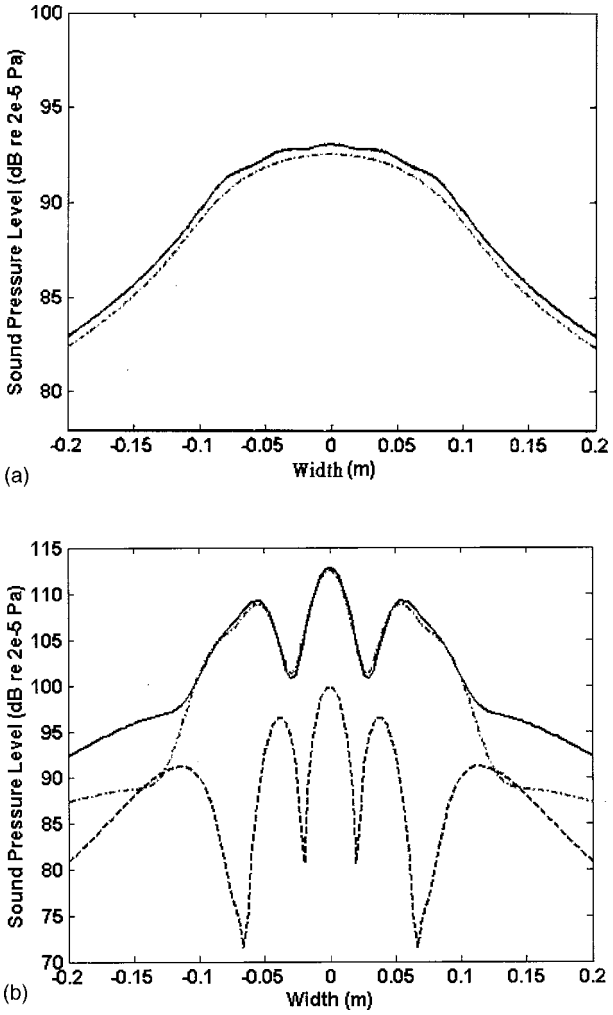


FIG. 7. Comparison of the far-field sound pressure radiated by the baffled piston, obtained using the theoretical model, E matrix with full modes, and E matrix with the first six modes. (a) 5×5 mesh, $ka=0.5$ and FN$=2$; (b) 11×11 mesh, $ka=30$ and FN$=0.2$ (-·-, theoretical value; —, **E** with full modes; — — —, **E** with six modes).

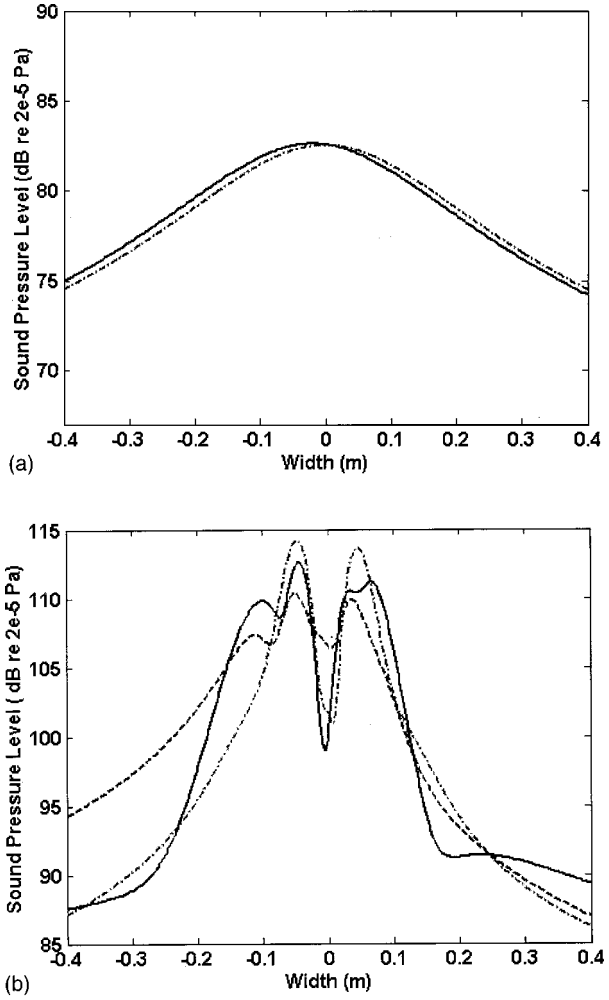


FIG. 8. Comparison of the far-field sound pressure radiated by the vibrating plate, obtained using the theoretical model, E matrix with full modes, and E matrix with the first six modes. (a) 5×5 mesh, $ka=1.5$ and $FN=5$; (b) 11×11 mesh, $ka=20$ and $FN=0.1$ (-·-, theoretical value; —, **E** with full modes, — — —, **E** with six modes).

C. Experimental investigation of the vibrating plate

In addition to the forgoing numerical simulations, experiments were carried out for a vibrating plate in order to further verify the proposed techniques. The experimental arrangement is shown in Fig. 9. The experimental rig consists of primarily a PU-foam panel embedded in a wooden baffle. The dimensions of the panel are $0.113\ \text{m}\times0.2\ \text{m}\times0.002$ m. The material properties of the panel are density $\rho=246\ \text{kg/m}^3$, Young's modulus $E=2.28\times10^9\ \text{N/m}^2$, Poisson ratio $\nu=0.33$. The panel is excited at the center by a small electromagnetic shaker. A fiber optic displacement sensor (PHILTEC, D125) is used to measure the surface velocities at the nine grid points on the vibrating panel.

In this experiment, the MVDRM method and the ISO 3745 are compared to the numerical model. Sound power calculated by using the aforementioned ANSYS and Rayleigh integral, with exciter response taken into account, serves as the ''theoretical'' value for comparison. The ISO method is conducted in an anechoic room. With comparable number of measuring positions as the conventional methods, the proposed method is conducted in an ordinary room because only surface velocity is required as the input data. A requirement of the proposed method is that the measurement must be conducted for planar sources baffled in a rigid infinite plane. Figure 10 shows the results obtained using these three approaches. The total sound power within the band 20–16 kHz are found to be 72.2 dB, 65.3 dB, and 69.4 dB

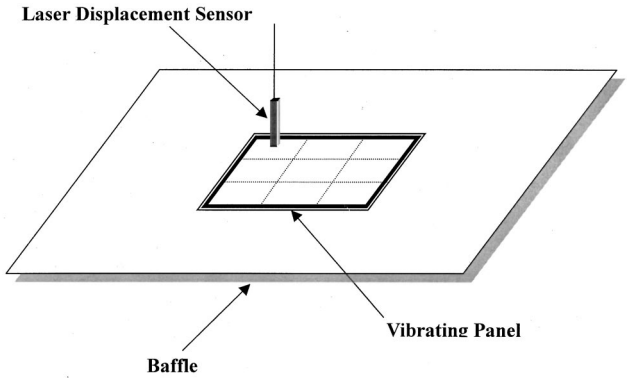


FIG. 9. The experimental arrangement of a baffled vibrating panel for sound power estimation.

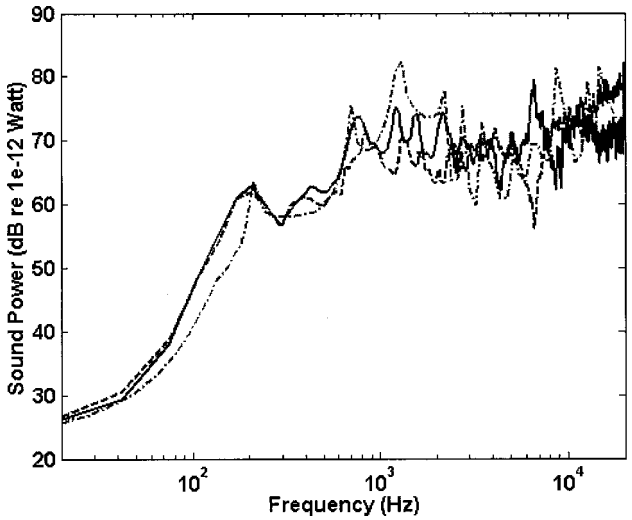


FIG. 10. Experimental results of sound power estimation for the baffled vibrating panel (- - -, theoretical value; — — —, ISO 3745; —, MVFRM).

for the theoretical value, ISO 3745, and MVDRM, respectively. Similar to the observation in numerical simulation, the experimental investigation indicates the proposed method provides a better estimation of sound power than the ISO standard.

IV. CONCLUSIONS

A technique based on radiation matrix is presented in this paper for the analysis of acoustic radiation of baffled planar sources. Numerical simulation and experimental investigation were carried out to verify the methods. From numerical and experimental results, the proposed method proves to be useful in estimating sound power and far-field pressure radiated by a complex source. In comparison with the conventional method, ISO 3745, the proposed techniques produced closer estimation of sound power. With comparable number of measuring positions as the conventional method, the proposed approach does not require special measuring environments. In addition, to calculate the far-field sound pressure for a source with a large *ka* value requires a matrix **E** with higher dimension. Conversely, it takes only a few dominant modes to reconstruct far-field sound pressure for sources with a small *ka* value.

As a limitation of the proposed techniques, the sound sources are restricted to planar sources baffled in a rigid infinite plane. By more sophisticated numerical schemes, e.g., the boundary element methods, the same rationale can be readily extended to deal with more practical problems such as radiation from three-dimensional sources. This aspect will be explored in future research.

ACKNOWLEDGMENTS

This work was supported by the National Science Council in Taiwan, Republic of China, under the Project No. NSC 89-2212-E009-057.

[1] J. R. Hasall and K. Zaveri, *Acoustic Noise Measurements* (Bruel & Kjaer, Denmark, 1988).
[2] R. L. Clark, W. R. Saunders, and G. P. Gibbs, *Adaptive Structures* (Wiley, New York, 1998).
[3] S. J. Elliott and M. E. Johnson, "Radiation modes and the active control of sound power," J. Acoust. Soc. Am. **94**, 2194–2204 (1993).
[4] G. V. Borgiotti and K. E. Jones, "The determination of the acoustic far field of a radiating body in an acoustic fluid from boundary measurements," J. Acoust. Soc. Am. **93**, 2788–2797 (1993).
[5] M. N. Currey and K. A. Cunefare, "The radiation modes of baffled finite plates," J. Acoust. Soc. Am. **98**, 1570–1580 (1995).
[6] B. S. Cazzolato, "Sensing systems for active control of sound transmission into cavities," Ph.D. thesis, University of Adelaide, 1999.
[7] G. V. Borgiotti and K. E. Jones, "Frequency independence property of radiation spatial filters," J. Acoust. Soc. Am. **96**, 3516–3524 (1994).
[8] W. T. Baumann, W. R. Saunders, and H. H. Robertshaw, "Active suppression of acoustic radiation from impulsively excited structures," J. Acoust. Soc. Am. **90**, 3202–3208 (1991).
[9] A. P. Berkhoff, "Sensor scheme design for active structural acoustic control," J. Acoust. Soc. Am. **108**, 1037–1045 (2000).
[10] P. Kohnke, Theory Reference, ANSYS, 1994.
[11] ISO 3745, Acoustics—Determination of sound power levels of noise sources—Precision methods for anechoic and semi-anechoic rooms, 1977.
[12] R. D. Ciskowski and C. A. Brebbia, *Boundary Element Methods in Acoustics* (Elsevier, London, 1991).
[13] B. Noble and J. W. Daniel, *Applied Linear Algebra* (Prentice-Hall, Englewood Cliffs, NJ, 1988).
[14] G. P. Gibbs, R. L. Clark, D. E. Cox, and J. S. Vipperman, "Radiation modal expansion: application to active structural acoustic control," J. Acoust. Soc. Am. **107**, 332–339 (2000).
[15] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
[16] G. Hübner, "Analysis of errors in measuring machine noise under free-field conditions," J. Acoust. Soc. Am. **54**, 967–977 (1973).

Remote measurement of material properties from radiation force induced vibration of an embedded sphere[a)]

Shigao Chen, Mostafa Fatemi, and James F. Greenleaf[b)]
Basic Ultrasound Research Laboratory, Department of Physiology and Biophysics, Mayo Clinic, Rochester, Minnesota 55905

(Received 4 September 2001; revised 22 February 2002; accepted 24 June 2002)

A quantitative model is presented for a sphere vibrated by two ultrasound beams of frequency ω_1 and ω_2. Due to the interference of two sound beams, the radiation force has a dynamic component of frequency $\omega_2-\omega_1$. The radiation impedance and mechanical impedance of the sphere are then used to compute the vibration speed of the sphere. Vibration speed versus vibration frequency is measured by laser vibrometer on several spheres, both in water and in gel phantom. These experimental results are used to verify the model. This method can be used to estimate the material properties of the medium (e.g., shear modulus) surrounding the sphere. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1501276]

PACS numbers: 43.20.Tb, 43.25.Qp, 43.30.Jx [MFH]

I. INTRODUCTION

The study of objects in terms of their mechanical response to external forces is of considerable interest in material science and medical diagnosis. Changes of elasticity of soft tissues are often related to pathology. The diagnostic value of a characteristic of an object depends upon the range of variation of that characteristic as a function of the state of the object. For soft tissues, the range of shear modulus can vary over three orders of magnitude [Sarvazyan *et al.*, 1998]. Thus, imaging techniques that exploit the shear modulus, or stiffness, of an object have great potential in medical application. Elasticity imaging [Gao *et al.*, 1996], a subject extensively investigated in recent years, is a quantitative method that measures the mechanical properties of tissue. The general approach is to measure the response of tissue to an excitation force and use it to reconstruct the elastic parameters of the tissue. These parameters are usually related to the shear modulus, or ''hardness'' of the tissues being imaged.

Vibro-acoustography [Fatemi *et al.*, 1999] is a new imaging modality, which can image the hardness of an object. Figure 1 is a brief diagram of a vibro-acoustography imaging system. The confocal transducer has a center disk and an outer ring that introduce two ultrasound beams to the same focal spot. These two ultrasound beams have slightly different frequencies: for example, $\omega_1=1.001$ MHz, and $\omega_2=0.999$ MHz. At the focal spot, the interference of these two beams will cause the object to vibrate at the beat frequency, in this case, at $\Delta\omega=2$ kHz. The resulting acoustic emission contains information about the local material properties of the object and can be detected by the acoustic hydrophone. Scanning the focal point of the transducer in a faster manner will generate a 2D image of the object.

Vibro-acoustography is particularly useful for imaging hard inclusions in soft material. For example, it has been used to image calcification in human arteries [Fatemi *et al.*, 1998], fractures in metal heart valves [Fatemi *et al.*, 2000], and microcalcifications in breast tissue [Fatemi *et al.*, 2002]. Other than imaging, another potential application of this technique is material characterization. In this paper, we explore this possibility. We speculate that by evaluating the response of the object to many excitation frequencies, we can estimate the mechanical properties, such as the complex shear modulus, of the object. A model is proposed for a spherical target embedded in a viscoelastic homogeneous medium. The goal is to estimate the mechanical properties of the medium by evaluating the displacement frequency response of an embedded spherical target. In this approach we excite the target (e.g., a sphere) at a fixed position, and monitor its response as we change the excitation frequency ($\Delta\omega$). By fitting the model to the data, material properties of the medium can be measured.

One potential medical application of this method is to estimate the mechanical properties of breast tissue, using microcalcification in the breast as a target. Breast microcalcifications are commonly found within both benign and malignant lesions [Fatemi *et al.*, 2002]. Previous studies suggest that malignant breast tissues are considerably stiffer than benign breast tissues [Krouskop *et al.*, 1998]. Therefore, information about tissue properties obtained from the vibration of microcalcification has diagnostic value. Other potential applications of this method can be nondestructive evaluation for materials such as polymers or other soft materials. The nature of this technique allows us to remotely access the mechanical properties of the medium around a target.

The focus of this paper is to develop a quantitative model for a sphere vibrated by two ultrasound beams of different frequency in a homogeneous medium, and to design experiments to verify the model. The paper is organized as follows. The dynamic radiation force used to vibrate the sphere is derived in Sec. II. Section III gives the radiation impedance of an oscillating sphere, from which the vibrating velocity of the sphere is calculated. The theory and experimental results are compared in Sec. IV. Section V discusses

[a)]Part of this work was presented in ''Vibro-acoustography of small spheres'' at the 140th meeting of the Acoustical Society of America, Newport Beach, CA, December 2000.

[b)]Electronic mail: jfg@mayo.edu

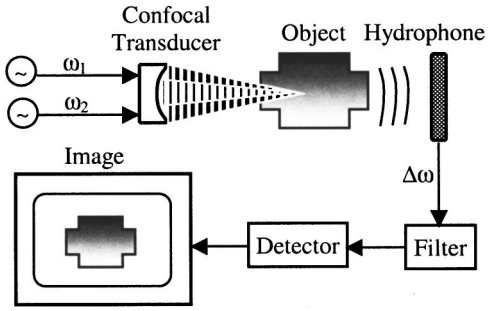


FIG. 1. Diagram of a vibro-acoustography imaging system. The confocal transducer has two elements: a center disk and an outer ring. The object is vibrating at the beat frequency $\Delta\omega$. Its emission is detected by the acoustic hydrophone to reconstruct the image.

some practical application issues of this research. The last section is the summary.

II. DYNAMIC RADIATION FORCE ON A SPHERE

The acoustic radiation force is a time-average force exerted by an acoustic field on an object. This force is caused by a change in the energy density of an incident acoustic field. Thus, an object in the wave path that absorbs or reflects sound energy is subjected to the acoustic radiation force. This force is usually a steady force, given that the intensity of the incident sound field does not change over time. In vibro-acoustography, the incident ultrasound is modulated such that its energy density changes sinusoidally at a low frequency (in the order of a few kilohertz). Thus, the radiation force on the object is oscillatory. For a general description of this dynamic radiation force on an arbitrary object, see Fatemi and Greenleaf (1999). In the following paragraphs, we will focus only on issues relevant to the application treated in this paper.

Consider a plane ultrasound wave incident on a sphere immersed in fluid. The radiation force on the sphere is along the incident wave direction, and its magnitude is

$$\langle F\rangle = \pi a^2 Y \langle E\rangle, \tag{1}$$

where πa^2 is the projected area of the sphere, Y is the radiation force function, and $\langle E\rangle$ is the time-averaged energy density of the ultrasound [Hasegawa *et al.*, 1969]. The symbol $\langle\ \rangle$ represents time average. The coefficient Y, which can be determined from the material properties of the sphere and the fluid medium, is usually expressed as a function of $k\cdot a$, where k is the wave number of the incident wave and a is the sphere's radius. Consider another situation where the incident pressure field $p(t)$ consists of two sine waves of slightly different frequency

$$\begin{aligned} p(t) &= P_0 \cos\omega_1 t + P_0 \cos\omega_2 t \\ &= 2P_0 \cos\left(\frac{\omega_1-\omega_2}{2}t\right)\cos\left(\frac{\omega_1+\omega_2}{2}t\right), \end{aligned} \tag{2}$$

which is equivalent to an amplitude-modulated sine wave. We have not included position as an argument in Eq. (2), since it is not relevant to the illustration here. For typical applications in this paper, $(\omega_1-\omega_2)/2$ is at least three orders of magnitude smaller than $(\omega_1+\omega_2)/2$. Thus, $k_1 a \approx k_2 a$, leading to $Y(k_1 a)\approx Y(k_2 a)$, which we can denote as $\overline{Y}$. The time-averaged energy density of this field is [Fatemi *et al.*, 1999]

$$\begin{aligned} \langle E\rangle &= \left[2P_0\cos\left(\frac{\omega_1-\omega_2}{2}t\right)\right]^2 \Big/ \rho c^2 \\ &= [1+\cos(\omega_1-\omega_2)t]\frac{2P_0^2}{\rho c^2}, \end{aligned} \tag{3}$$

where ρ and c are the density and sound speed of the fluid, respectively. Therefore, the radiation force on the sphere has a dynamic component of frequency $\omega_1-\omega_2$

$$\langle F_d\rangle = \pi a^2 \overline{Y}\langle E_0\rangle \cos(\omega_1-\omega_2)t, \tag{4}$$

where $\langle E_0\rangle = 2P_0^2/\rho c^2$.

In practice, the incident beams from the confocal transducer may differ from ideal plane waves, and their intensity may not be equal. The medium surrounding the sphere may be gelatin or soft tissue, instead of fluid. This will change the expression of $\pi a^2 \overline{Y}\langle E_0\rangle$, which results in a dynamic radiation force of different amplitude. But, the main conclusion remains valid: the dynamic radiation force is of frequency $\omega_1-\omega_2$, and its amplitude remains constant for all oscillating frequencies $(\omega_1-\omega_2)$ that are small. There are a few cases where the value of radiation force function Y is extremely sensitive to the change of ka, for example, at the resonant frequencies of a gold sphere in water [Chivers *et al.*, 1982]. Under these conditions, $\overline{Y}$ in Eq. (4) may not be independent of ω_1 and ω_2. But, in most of the situations it is safe to assume the dynamic radiation force is of constant amplitude when we sweep the oscillating frequency $(\omega_1-\omega_2)$.

III. IMPEDANCE AND VELOCITY IN VISCOUS MEDIUM

From the theory developed above, the dynamic driving force on the sphere is known. To find out the oscillating speed of the sphere, we need to compute the resistance it will meet when the sphere tries to move back and forth. In this section, a method is proposed to solve for the oscillating speed of the sphere, from which the acoustic emission from the sphere can also be derived.

A. Impedance of the sphere

For a rigid sphere oscillating back and forth at frequency $\Delta\omega$, the stress field around the sphere can be calculated. The net force on the sphere can be found by integrating the stress at the surface of the sphere. The radiation impedance of the sphere is equal to this force divided by the vibrating speed of the sphere, and represents the resistance the sphere will ''feel'' when pushing its surrounding medium back and forth. Oestreicher (1951) derived a radiation impedance formula for a rigid sphere oscillating in a viscoelastic medium

$$Z_r = -i\frac{4\pi a^3}{3}\rho\Delta\omega \times \frac{\left(1-\frac{3i}{ah}-\frac{3}{a^2h^2}\right)-2\left(\frac{i}{ah}+\frac{1}{a^2h^2}\right)\left(3-\frac{a^2k^2}{aki+1}\right)}{\left(\frac{i}{ah}+\frac{1}{a^2h^2}\right)\frac{a^2k^2}{aki+1}+\left(2-\frac{a^2k^2}{aki+1}\right)}. \quad (5)$$

Here, $k=\sqrt{\rho\Delta\omega^2/(2\mu+\lambda)}$, $h=\sqrt{\rho\Delta\omega^2/\mu}$, $\mu=\mu_1+i\Delta\omega\mu_2$, $\lambda=\lambda_1+i\Delta\omega\lambda_2$. Thus, the radiation impedance is determined by the radius of the sphere a, the density of the medium ρ, the shear elasticity and viscosity of the medium μ_1 and μ_2, and the volume elasticity and viscosity of the medium λ_1 and λ_2.

Another resistant force for vibration is the inertia of the sphere. For a sphere of mass m and oscillating velocity $\mathrm{V}e^{i\Delta\omega t}$, the force required to overcome the inertia of the sphere is $F=m[d(\mathrm{V}e^{i\Delta\omega t})/dt]=im\Delta\omega\,\mathrm{V}e^{i\Delta\omega t}$. We define the ''mechanical'' impedance of the sphere as

$$Z_m=\frac{-F}{\mathrm{V}e^{i\Delta\omega t}}=-im\Delta\omega. \quad (6)$$

The total impedance of the sphere is the summation of these two: $Z=Z_r+Z_m$.

B. Oscillating speed of the sphere

The dynamic radiation force drives the sphere to vibrate, and the impedance represents the resistance towards vibration. Dividing the driving force by the impedance yields the vibrating speed

$$V=\frac{\langle F_d\rangle}{Z_r+Z_m}. \quad (7)$$

Once the vibrating speed of the sphere is known, the acoustic emission from the sphere can be calculated. The relevant formula can be found in the literature [Morse, 1981]. But, our experiments are done in a water tank. The reverberation of low-frequency sound in the tank makes it difficult to measure acoustic emission from the sphere accurately. So, experiments are designed to measure directly the vibrating velocity of the sphere instead.

Next, we show how the impedance of the sphere can change the vibration of the sphere. Some analysis of Eq. (5) will help us understand its influence on the vibration amplitude of the sphere. For soft tissues, $\lambda_1\sim10^9$ Pa, $\lambda_2\sim0$ Pa·s; $\mu_1\sim10^4$ Pa, $\mu_2\sim10^{-2}$ Pa·s [Frizzell *et al.*, 1977; Oestreicher, 1951]. Therefore, ak is very small for the a and $\Delta\omega$ considered ($a<1$ mm and $\Delta\omega\leqslant2\pi\cdot10^3$ rad/s), and can be safely set to zero. After this step, Z_r is only a function of ah. Then, in the limit of large ah, one obtains purely mass impedance, and purely elastic impedance in the other limit. The absolute value of the total impedance Z_r+Z_m is minimum at some resonance frequency approximately defined by the cancellation of total mass and elasticity contribution. Since the amplitude of the dynamic radiation force that is driving the sphere is constant for all vibrating frequencies, the vibration amplitude of the sphere will demonstrate a pattern of resonance versus vibration frequency. And, the resonant pattern is determined by μ_1 and μ_2, not by λ_1 and λ_2. Computer simulation can demonstrate the result of this analysis more clearly. In the following simulations, it is assumed that the amplitude of the driving force on the sphere is 10^{-5} newton.

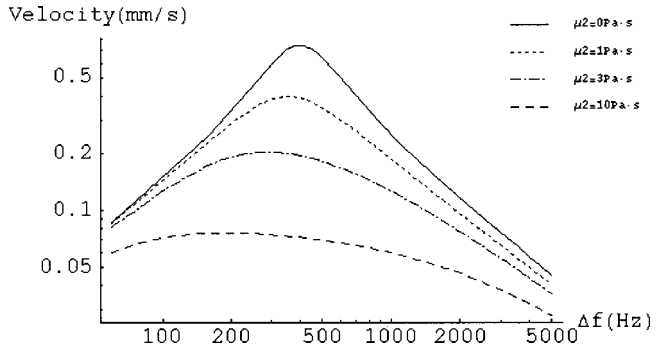


FIG. 2. Simulation of a sphere vibrated in gel. Parameters for the sphere: radius=0.59 mm, density=7667 kg/m^3. The gel has a shear elasticity of 4 kPa, and shear viscosity of 0, 1, 3, and 10 Pa·s, respectively.

Figure 2 is a simulation of the vibration speed versus vibrating frequency for a sphere in gel. The radius of the sphere is 0.59 mm. The gel has a shear elasticity of 4 kPa and its shear viscosity is set at 0, 1, 3, and 10 Pa·s, respectively. The velocity shows a resonance near 400 Hz. Low viscosity corresponds to high resonant peak.

In Fig. 3, the value of shear viscosity is fixed and μ_1 is changed over 30-fold. When μ_1 decreases, the resonant peak is shifted to lower frequency and its magnitude becomes higher. In water, the peak is at frequency 0. In that case, only the right side of the peak is visible. The shear properties of the gel can change the velocity profile dramatically. Thus, it is possible to determine μ_1 and μ_2 by fitting the measured velocity profile. The size of the sphere can also alter the velocity profile. Smaller spheres have higher resonant peaks.

IV. COMPARISON OF THEORY AND EXPERIMENT

A. Experimental setup

The setup of experiment is shown in Fig. 4. The transducer vibrates the sphere back and forth at frequency Δf in the water tank. A laser vibrometer is used to measure the vibration speed of the sphere. The laser vibrometer reflects a

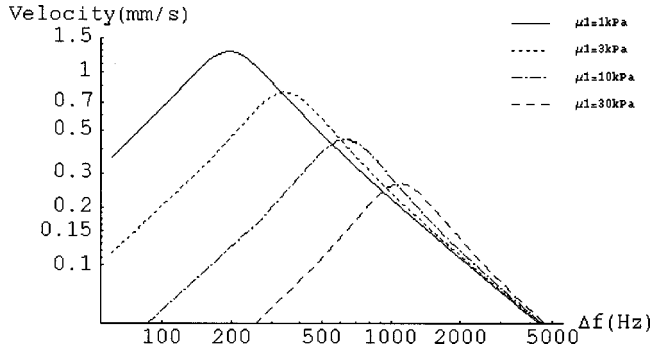


FIG. 3. Simulation of a sphere vibrated in gel. Parameters for the sphere: radius=0.59 mm, density=7667 kg/m^3. The gel has a shear viscosity of 0.1 Pa·s, and shear elasticity of 1, 3, 10, and 30 kPa, respectively.

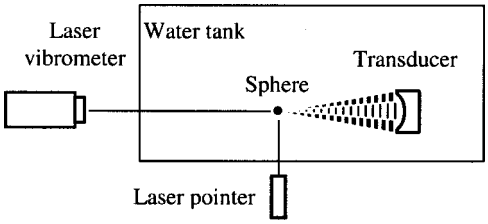


FIG. 4. Diagram of experimental setup. The laser vibrometer detects the motion of the sphere, which is vibrated by the transducer at frequency Δf. The laser beams of the vibrometer and the laser pointer are used to position the sphere at the focal spot of the transducer. The water tank is covered with sound-absorbing pads.

beam of light from the sphere, and detects the vibrating velocity of the sphere based on the Doppler frequency shift of the reflected laser. Four stainless-steel spheres (provided by New England Miniature Ball LLC, Norfolk, CT) of different size are used in the experiment. We put each sphere at the focal spot of the transducer and measure its vibration response versus frequency Δf. The laser beams of the vibrometer and the laser pointer are used to align each sphere to make sure that they stay at the focal spot of the transducer. For experiments of spheres vibrated in water, each sphere is suspended with very fine suture material in a bifilar arrangement. For experiments of spheres vibrated in gel, the spheres are embedded in a gel phantom of $8\times10\times12$ cm^3. The gel phantom is made from Bloom 300 gelatin powder (Sigma-Aldrich), with a concentration of 10%.

B. Results in water

In water $\mu_1, \mu_2 \to 0$, resulting in $ah \to \infty$, the impedance formula of the sphere can be simplified, and the velocity is

$$V \approx \frac{i3\langle F_d \rangle}{4\pi^2 a^3 (2\rho_s + \rho_0)} \times \frac{1}{\Delta f}. \tag{8}$$

Therefore, the amplitude of vibration velocity of the sphere should be inversely proportional to its vibrating frequency. Figure 5 is one set of measurement data for stainless-steel spheres vibrated in water. Since the velocity is inversely proportional to frequency, the curve should be a straight line within the log–log scale in this plot. All the dots are measurement points and each solid line is fitted by the LMS method according to Eq. (8). The experimental data agree

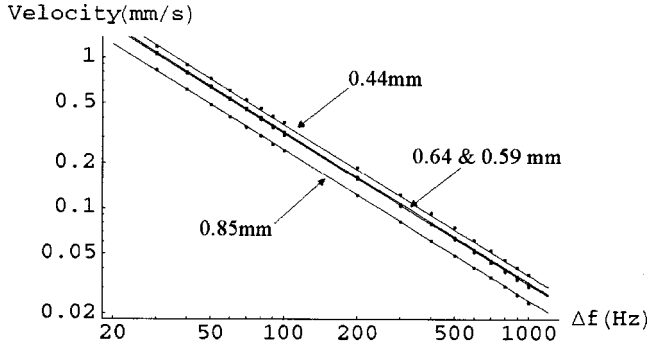


FIG. 5. Vibrations of 440C stainless-steel spheres of different size in water. Radius of the spheres: 0.44, 0.59, 0.64, 0.85 mm. Dots are measurement points and the solid lines are fitted by the LMS method.

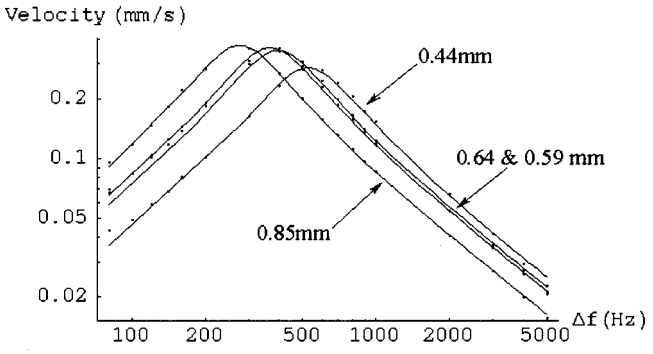


FIG. 6. Vibrations of 440C stainless-steel spheres of different size in gel. Radius of the spheres: 0.44, 0.59, 0.64, 0.85 mm. Dots are measurement points and the solid lines are fitted by the LMS method.

with the theory. Notice from Eq. (8) that one can also derive the dynamic radiation force on the sphere by the measured velocity. Hence, it is also possible to estimate the energy density of the incident ultrasound through Eq. (4), since $\overline{Y}$ can be deduced from computer simulation. This is similar to the idea of using a steady radiation force on a sphere as a primary method for determining the intensity of ultrasound fields [Hasegawa *et al.*, 1975; Dunn *et al.*, 1977].

C. Results in gel

In gel, μ_1 and μ_2 are no longer negligible. The formula of the vibrating speed of the sphere becomes more complicated. Computer simulations based on Eq. (5) are used to fit the data of measurement. Figure 6 is a set of measurement data for four spheres of different size in one single gel phantom. The dots are measurement points. By adjusting the value of μ_1, μ_2, and a, four solid lines are fitted to the points. As expected, the measurements show a resonance of velocity versus frequency. For spheres of different size, the optimal value of μ_1 lies within 3.9–4.2 kPa and μ_2 lies within 0.14–0.2 Pa·s. In Fig. 6, all four solid lines are plotted assuming $\mu_1 = 4.1$ kPa and $\mu_2 = 0.17$ Pa·s, which come from averaging optimal parameters of four spheres. Independent measurement of the gel's shear modulus μ_1 by MRE [Muthupillai *et al.*, 1995] yields 2.5~4.5 kPa, for shear wave frequency ranging from 200 to 500 Hz. The same gel samples are also tested by the dynamic mechanical analyzer (TA Instruments, New Castle, DE), and the measurement of μ_1 is within 3.2–4.7 kPa for shear vibration frequency from 0 to 10 Hz. All three methods yield similar results. We do not have a ground truth for μ_2 of the gel. Our estimation is close to the typical values of μ_2 for soft tissues estimated by Frizzell (1977).

V. DISCUSSION

The method proposed in this paper is quite sensitive to changes of material properties. It is known that the shear modulus of gelatin phantom becomes larger as the gel ages. Experiments are repeated on the same gel phantom on different days. This method is able to detect a consistent increase of the stiffness of the gel every day. But, the change of shear viscosity from day to day is not as prominent.

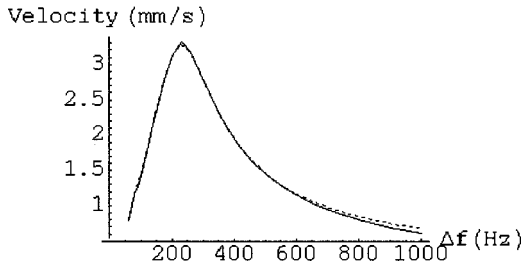


FIG. 7. Velocity measurement of a sphere in gel by laser vibrometer (dash line) and ultrasound Doppler method (solid line). Radius of the sphere: 0.59 mm. Estimation of the gel's shear modulus is 1384 Pa from the laser method and 1374 Pa from the ultrasound method.

For a laser vibrometer to measure the velocity of an object, the medium around the target must be optically clear. This will limit the applications of this method. Ultrasound Doppler technique has been used to detect the motion of soft tissue under external mechanical vibration [Yamakoshi *et al.*, 1990]. We also have used ultrasound Doppler to measure the velocity of the vibrating sphere. The method employed to extract the velocity information is similar to that described in Yamakoshi's paper. The frequency of the Doppler transducer is 5 MHz, while the frequency of the ultrasound used to vibrate the sphere is about 1 MHz. The received rf signal from the Doppler transducer is filtered out of the 1-MHz component before demodulation. Figure 7 shows the velocity profiles of a sphere in gel measured by laser vibrometer and the ultrasound Doppler technique simultaneously. Both methods obtain similar results. Therefore, ultrasound can be used to measure the vibration instead of laser.

The model can produce good quantitative results for calibrated targets, like the sphere used in the experiments. For targets of different shapes, the concept presented in this paper is still valid. However, we need to calculate the radiation impedance of that particular target. Resonance of vibrating amplitude versus frequency is still expected, and can be used to deduce mechanical properties of the medium around that target. To apply this model to breast calcification, some practical issues need to be addressed. First, the target here is not exactly spherical, though many appear to be so. More work needs to be done to access the impact of violating the spherical assumption: how sensitive is the radiation impedance to the shape of the target, and how spherical are breast calcifications? Given the high contrast between normal fat tissues and malignant breast tissues, it is likely that assessment of the tissue properties around the microcalcification using this method can help in screening of breast lesion. Another practical issue is estimating the size of the calcification. Currently we are developing a system that combines x-ray mammography and vibro-acoustography. The size of the calcification may be estimated by x-ray mammography. Then, this information may be used to estimate tissue properties by our method.

Another assumption of this model is that the medium surrounding the sphere is homogeneous. In some applications, this assumption might be true only within a small distance from the target. The particle displacement in the medium surrounding the sphere can be estimated, using Eq. (15) in Oestreicher's paper (1951). This displacement field diminishes as the distance from the sphere increases. For a 0.5-mm-radius sphere vibrated in the gel phantom considered in Fig. 6, the particle displacement at 5 mm from the sphere is estimated to be about 5% of that at the surface of the sphere. Thus inhomogeneity beyond 5 mm from the sphere probably can be tolerated. This ''effective'' distance becomes larger, as the medium's shear elasticity and the sphere's radius increase. This knowledge will indicate the confidence we can have for the material property estimation.

VI. SUMMARY

In this paper, we present a quantitative model for a sphere vibrated by two beams of ultrasound in a homogeneous viscous medium. The acoustic radiation force on the sphere is shown to have an oscillating component at the beat frequency of the two ultrasound beams. This driving force is used with the radiation impedance and the mechanical impedance of the sphere to calculate the vibrating velocity of the sphere. The theory predicts a distinct velocity profile (vibrating speed versus vibration frequency) for a sphere in water and in gel, which is confirmed by experiment. This method can be used to determine the local material properties of the medium surrounding the sphere. It can also be used to calibrate oscillatory driving force on the sphere.

ACKNOWLEDGMENTS

The authors are grateful to Randall Kinnick for his experimental support. This work was supported by the Army Medical Research and Material Command under Grant No. DAMAD 17-98-1-8121, and by Grant No. HL61451 from National Institutes of Health.

Chivers, R. C., and Ansosn, L. W. (**1982**). ''Calculations of the backscattering and radiation force functions of spherical targets for use in ultrasonic beams assessment,'' Ultrasonics **20**, 25–34.

Dunn, F., Averbuch, A. J., and O'Brien, W. D. (**1977**). ''A primary method for the determination of ultrasonic intensity with the elastic sphere radiometer,'' Acustica **35**, 58–61.

Fatemi, M., and Greenleaf, J. F. (**1998**). ''Ultrasound-stimulated vibro-acoustic spectrography,'' Science **280**, 82–85.

Fatemi, M., and Greenleaf, J. F. (**1999**). ''Vibro-acoustography: An imaging modality based on ultrasound-stimulated acoustic emission,'' Proc. Natl. Acad. Sci. U.S.A. **96**, 6603–6608.

Fatemi, M., Rambod, E., Gharib, M., and Greenleaf, J. F. (**2000**). ''Nondestructive testing of mechanical heart valves by vibro-acoustography,'' 7th International Congress on Sound and Vibration, Garmisch-Partenkirchen, Germany, 4–7 July 2000.

Fatemi, M., Wold, L. E., Alizad, A., and Greenleaf, J. F. (**2002**). ''Vibro-acoustic tissue mammography,'' IEEE Trans. Med. Imaging **21**(1), 1–8.

Frizzell, L. A., Carstensen, E. L., and Dyro, J. F. (**1977**). ''Shear properties of mammalian tissues at low megahertz frequencies,'' J. Acoust. Soc. Am. **60**(6), 1409–1411.

Gao, L., Parker, K. J., Lerner, R. M., and Levinson, S. F. (**1996**). ''Imaging of the elastic properties of tissue—a review,'' Ultrasound Med. Biol. **22**, 959–977.

Hasegawa, T., and Yosioka, K. (**1969**). ''Acoustic-radiation force on a solid elastic sphere,'' J. Acoust. Soc. Am. **46**, 1139–1143.

Hasegawa, T., and Yosioka, K. (**1975**). ''Acoustic radiation force on fused silica spheres and intensity determination,'' J. Acoust. Soc. Am. **58**, 581–585.

Krouskop, T. A., Wheeler, T. M., Kallel, F., and Garra, B. S. (**1998**). ''Elastic moduli of breast and prostate tissues under compression,'' Ultrason. Imaging **20**, 260–274.

Morse, P. (**1981**). *Vibration and Sound* (Acoustical Society of America), pp. 318–319.

Muthupillai, R., Lomas, D. J., Rossman, P. J., Greenleaf, J. F., Manduca, A., and Ehman, R. L. (**1995**). ''Magnetic resonance elastography by direct visualization of propagating acoustic strain waves,'' Science **269**, 1854–1857.

Oestreicher, H. L. (**1951**). ''Field and impedance of an oscillating sphere in a viscoelastic medium with an application to biophysics,'' J. Acoust. Soc. Am. **23**, 707–714.

Sarvazyan, A. P., Redenko, O. V., Swanson, S. D., Fowlkers, J. B., and Emelianov, S. Y. (**1998**). ''Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics,'' Ultrasound Med. Biol. **24**, 1419–1435.

Yamakoshi, Y., Sato, J., and Sato, T. (**1990**). ''Ultrasonic imaging of internal vibration of soft tissue under forced vibration,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**(2), 45–53.

Guided wave modes in porous cylinders: Experimental results

C. J. Wisse
Delft University of Technology, Department of Applied Earth Sciences, Mijnbouwstraat 120, 2628 RX Delft, The Netherlands and Eindhoven University of Technology, Department of Applied Physics, Fluid Dynamics Laboratory, Postbus 153, 5600 MB Eindhoven, The Netherlands

D. M. J. Smeulders[a)]
Delft University of Technology, Department of Applied Earth Sciences, Mijnbouwstraat 120, 2628 RX Delft, The Netherlands

M. E. H. van Dongen
Eindhoven University of Technology, Department of Applied Physics, Fluid Dynamics Laboratory, Postbus 153, 5600 MB Eindhoven, The Netherlands

G. Chao
Delft University of Technology, Department of Applied Earth Sciences, Mijnbouwstraat 120, 2628 RX Delft, The Netherlands and Eindhoven University of Technology, Department of Applied Physics, Fluid Dynamics Laboratory, Postbus 153, 5600 MB Eindhoven, The Netherlands

(Received 28 November 2001; revised 29 May 2002; accepted 31 May 2002)

In this paper guided wave modes in porous media are investigated. A water-saturated porous cylinder is mounted in the test section of a shock tube. Between the porous sample and the wall of the shock tube a water-filled annulus exists. For very small annulus width, bulk waves are generated and one-dimensional modeling is sufficient. Otherwise two-dimensional effects become important and multiple guided wave modes occur. Using a newly developed traversable positioning system in the shock tube, the frequency-dependent phase velocities and damping coefficients in the 1–120 kHz frequency range were measured. Prony's method was used for data processing. Agreement was found between the experimental data and the two-dimensional modeling of the shock tube which was based on Biot's theory. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1497621]

PACS numbers: 43.30.Jx [RR]

I. INTRODUCTION

Several laboratory experiments show the predictive power of Biot's theory[1,2] on wave propagation in porous media. Plona[3] demonstrates the existence of the slow compressional wave in synthetic rock samples. Using a spectral ratio technique for synthetic rock samples, Johnson *et al.*[4] measured phase velocities of the three Biot waves and damping coefficients of the slow wave. Kelder and Smeulders[5] observed the existence of the slow compressional wave in natural rock. The mentioned experiments are limited to the ultrasonic frequency range (>500 kHz). In this frequency range inertia effects in the pores are dominant and both the fast and the slow compressional wave are propagative. However, below the so-called crossover frequency, viscous effects in the pores dominate over the inertia effects and the slow wave is strongly damped and shows diffusive behavior. Therefore, it is interesting to obtain experiments in a frequency range where both viscous and inertia effects are important. A second limitation of these experiments is that they are concerned with plane wave propagation. Two-dimensional effects like surface wave phenomena do not occur. Surface waves are investigated by Kelders *et al.*[6] They present experimental evidence of surface waves at 40 kHz above thin porous layers of plastic foam.

The experiments of Winkler *et al.*[7] cover a broad frequency range as well as surface wave phenomena. They report experiments on Stoneley waves in both artificial and natural (Berea sandstone) samples in the 10–90 kHz frequency range. Using oils with different viscosities as a saturant liquid both inertia and viscous effects are important in this frequency range. Using Prony's method,[8,9] Winkler *et al.* obtained frequency-dependent phase velocities and damping coefficients of the Stoneley wave. Excellent agreement was found between modeling and experimental results. The wave velocities of the compressional and shear waves were measured under saturated conditions. These measured wave velocities were used as an input for the modeling. They only report data of the Stoneley wave, most likely because other mode types were weakly excited.

Hsu *et al.*[10] investigated the characteristics of tube waves in an elastic solid formation, with and without a cylindrical mandrel. Both porous and elastic mandrels were used. Using synthetic porous samples as a mandrel, they found agreement with their modeling for the phase velocity of the tube wave in the 10–100 kHz frequency range. Between 50 and 90 kHz a bulk mode could be recorded. Only those modes with appreciable energy in the fluid annulus could be detected. Attenuation data were also presented, but the scatter is such that it is difficult to compare the results to their modeling.

The current paper presents data of both bulk and surface waves in the 1–120 kHz frequency range for natural and synthetic rock. For our experiments we used a shock tube. Van der Grinten *et al.*,[11,12] Sniekers *et al.*,[13] and Smeulders and van Dongen[14] performed shock tube experiments on

a)Electronic mail: d.m.j.smeulders@ta.tudelft.nl

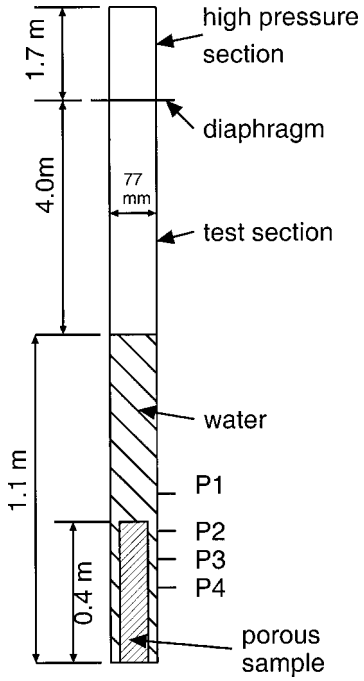


FIG. 1. Shock tube configuration. P1, P2, P3, and P4 indicate the positions of the pressure transducers. The transducer positions are shown only in relative order and not to scale.

water-saturated, partially saturated, and dry samples, and they presented their data in the time domain. In the current paper we present an extension of the work of these authors as we also present data of Stoneley waves, and we present data in the frequency domain. By a unique modification of the conventional shock tube configuration we could measure frequency-dependent phase velocities and damping coefficients.

The approach of this paper is as follows. First, we discuss the basics of the shock tube experiments and we present data in the time domain. Subsequently, we present the modification of the shock tube configuration. Finally, we discuss the frequency-dependent phase velocities and damping coefficients. The results are compared with the full modeling of the shock tube configuration which is based on Biot's theory.

II. SHOCK TUBE

The shock tube is shown in Fig. 1. It consists of a high pressure section and a low pressure section, separated by a diaphragm. The length of the shock tube is 6.8 m, while the inner diameter of the test section is 77.0 mm. The length of the test section is 5.1 m. The test section is this long in order to obtain a fully developed shock wave. The shock tube sections are made out of steel and have a wall thickness of 24 mm in order to minimize compliance effects.

A porous cylinder is mounted in the test section of the shock tube and carefully saturated with water of properties given in Table I. The water level is some 66 cm above the top of the sample. For the diameter of the sample two cases have been considered. In the first case a diameter is chosen very close to the inner diameter of the test section: 76.5 mm. This means that the sample-to-tube surface ratio is 0.987. In the second case a diameter of 70.0 mm is used. The surface ratio is then only 0.826 and thus a relatively large water-filled annulus exists between the surface of the sample and the shock tube wall.

TABLE I. Fluid parameters.

Density of the fluid ρ_f (kg/m^3)	998
Viscosity of the fluid η (mPa s)	1.0
Liquid bulk modulus K_f (GPa)	2.2

A wave experiment proceeds as follows. The pressure in the high pressure section is raised to a value in the range of 2–5 bars and the diaphragm is ruptured by means of an electric current. A shock wave in air is generated which is transmitted into the water column on top of the sample. In Fig. 2, the wave types are sketched. The step wave (cw_c) in the water partially reflects, partially transmits into the sample, where bulk waves (c1, c2) are generated. Surface waves (s) are generated on the surface between the sample and the liquid-filled annulus. Because of normal incidence no bulk-type shear waves are induced. The shock tube wall is equipped with four Kistler 603B pressure transducers, at the positions indicated by P1, P2, P3, and P4. The first transducer P1 is mounted some 15 cm above the top of the cylinder, the second transducer P2 at 0.5 cm below the top. The distance between P2 and P3, and P3 and P4 is 5 cm. Data acquisition is performed using a Lecroy 6810 wave form recorder.

III. TIME DOMAIN RESULTS

We first used a synthetic rock sample with a diameter of 76.5 mm. It consists of sintered crushed glass. The rock properties are given in Table II. Typical pressure signals are given in Figs. 3 (solid lines). At P1 the arrival of the step wave is observed at $t=0.51$ ms followed by its reflection from the top of the sample at $t=0.72$ ms. The negative pressure peak at 0.51 ms is caused by compliance effects of the shock tube wall. The pressure rise at $t=1.19$ ms is caused by

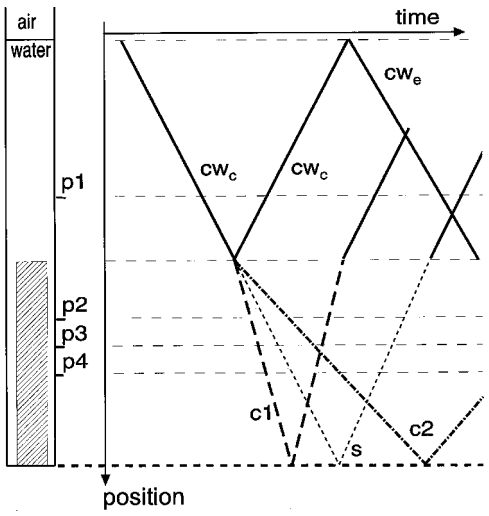


FIG. 2. Space–time plot of the incident, reflected and transmitted waves. The compressional waves in water are denoted by cw_c, the expansion wave by cw_e. The fast compressional wave is denoted by c1. the slow compressional wave by c2, and the surface wave by s. The transducer positions are shown in relative order and not to scale.

TABLE II. Synthetic samples parameters.

Diameter (mm)	76.5	70.0
Density of the solid ρ_s (kg/m^3)	2590	2570
Porosity ϕ	0.52	0.53
Permeability k_0 (10^{-12} m^2)	3.2	3.4
Tortuosity α	1.7	1.7
Constrained modulus K_p (GPa)	2.5	2.1
Shear modulus G (GPa)	0.88	0.88
P-wave velocity (m/s)	1190	1129
S-wave velocity (m/s)	582	585

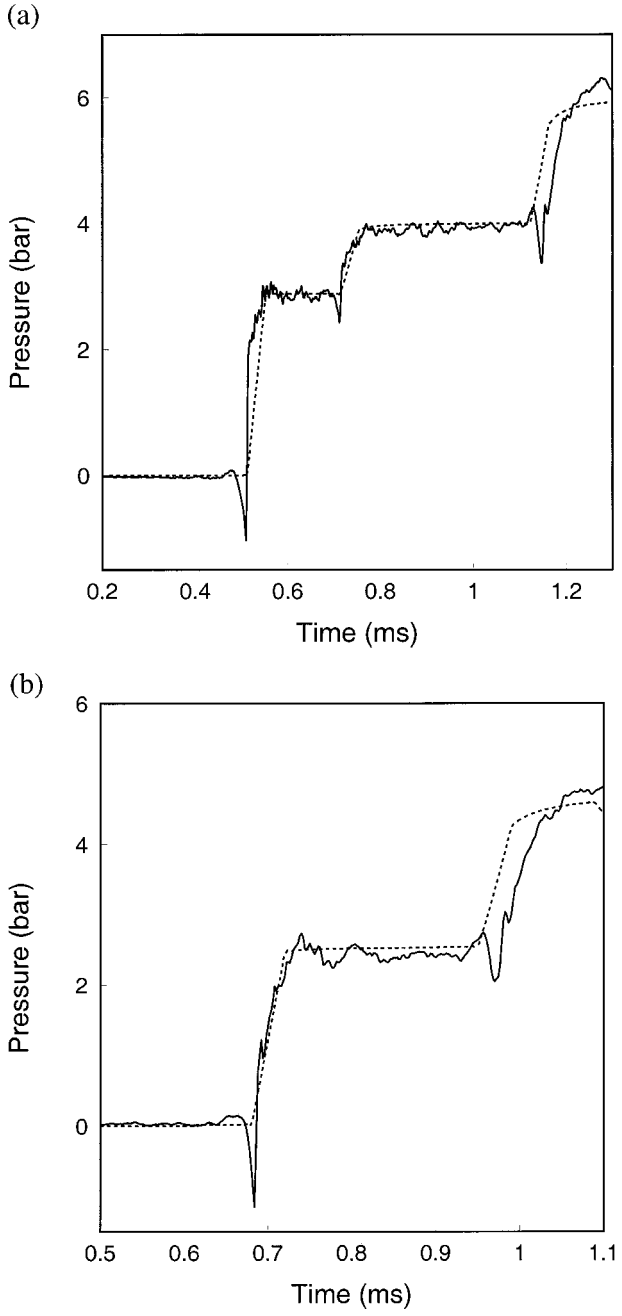


FIG. 3. Pressure signals at P1 (a) and P4 (b). Sample diameter: 76.5 mm. Sample length: 40 cm. The experiments were performed on a synthetic material. P1 is mounted at $z=-152$ mm and P4 at $z=136$ mm, where $z=0$ corresponds to the top of the sample. Solid lines refer to the experiments while the dashed lines correspond to 1D computations based on Biot's theory.

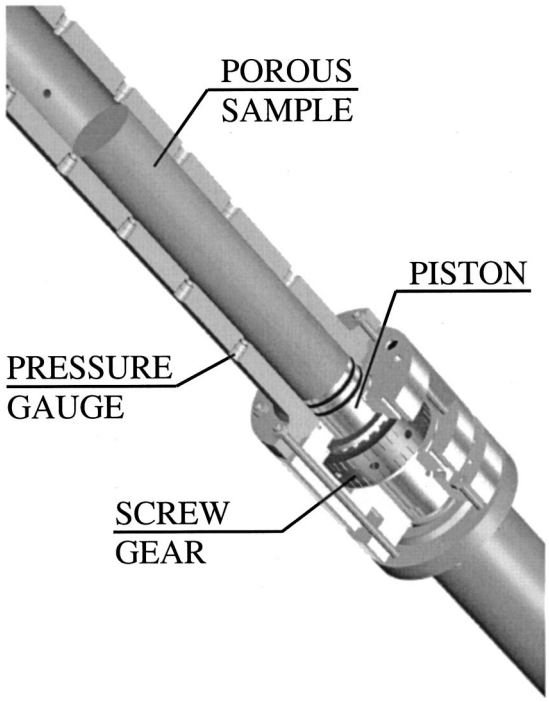


FIG. 4. Porous sample in test section of the shock tube.

the reflection of the fast wave from the lower end-plate of the shock tube. At P4, the arrival of the fast compressional wave is recorded at $t=0.68$ ms, followed by a second pressure wave arrival which corresponds to its reflection from the end plate. The slow compressional wave is not observed due to the large damping in the amplitude caused by attenuation effects.

The measurements are compared with one-dimensional (1D) computations based on Biot's theory. We developed a fast Fourier transform-based computer code which describes the transmitted and reflected waves in the shock tube.[15] This code is an extension of the code that was developed van der Grinten *et al.*[11,12] The predicted pressure signals are given in Figs. 3 (dashed lines). Excellent agreement with the experimental data was found.

The situation changes completely when we measure samples with a diameter of 70.0 mm. In this case, we find that two-dimensional effects become important and that the 1D modeling is no longer appropriate. The results in the frequency domain will show that surface wave effects are prominent. In order to properly measure surface wave propagation, we need accurate information on the pressure distribution over the sample length.

IV. SHOCK TUBE MODIFICATION

In order to record a large number of pressure signals at equidistant positions over the sample length, the shock tube was modified. In the conventional measurements as described before, only three signals at 5 cm spacing could be recorded.

The test section is shown in greater detail in Fig. 4. It is newly designed so that the position of the test sample in the shock tube can be varied with respect to the pressure gauges. By rotation of the screw gear (see Fig. 4), the piston moves

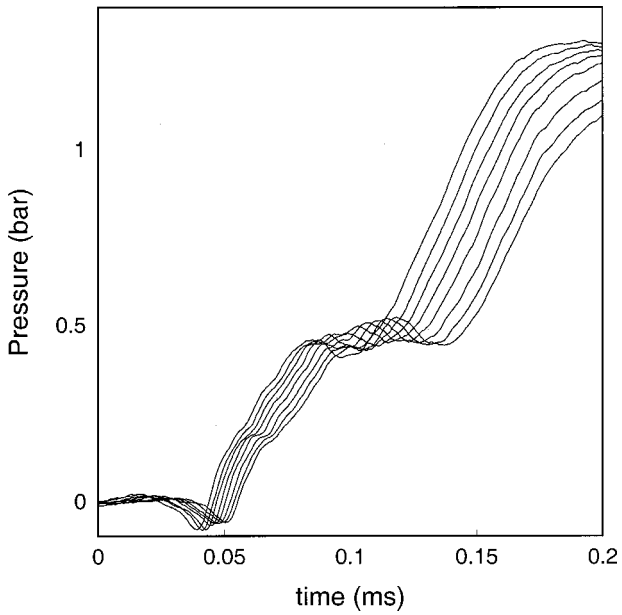


FIG. 5. Pressure signals for a Bentheim sample at equidistant axial positions. Sample diameter 70.0 mm, $\Delta z = 0.5$ cm.

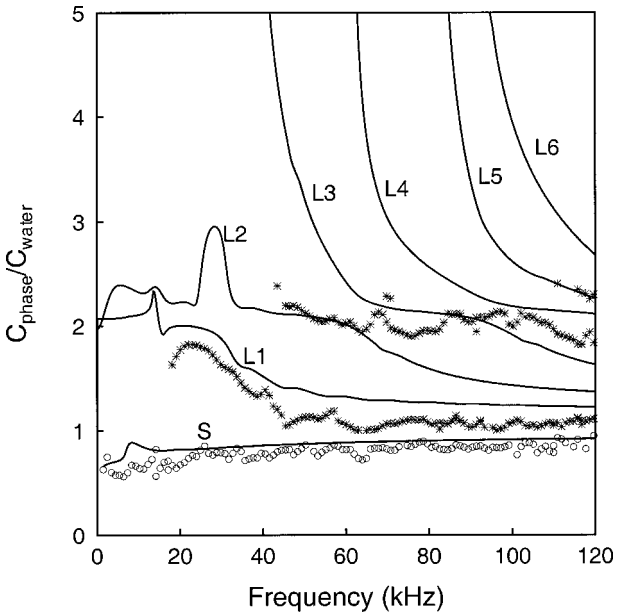


FIG. 6. Comparison between experiments and theory for Bentheim sandstone. The physical parameters were determined in independent laboratory experiments. Solid lines: wave mode computations; (○) experiments with a cylinder which was sealed at the top surface; (∗) experiments with open pores at the top surface. We used a Bentheim sandstone cylinder with a diameter of 70.0 mm.

the sample to a new position over a distance Δz with a system inaccuracy of a tenth of a millimeter. The position of the porous cylinder is measured electronically.

In this way, a large number of equidistant pressure signals can be recorded by combining different shock tube runs. The experimental procedure is organized as follows. A shock tube run as described before is performed. The pressure transducers are mounted as drawn in Fig. 1. Three pressure signals are recorded by P2, P3, and P4, while the transducer P1 is used as a reference. Subsequently, the sample is moved upwards over a distance of 0.5 cm. Then again a wave experiment is carried out and three more pressure signals are recorded. This procedure is repeated 10 times, which results in a record of 30 pressure signals 0.5 cm apart. A typical example of such a record is given in Fig. 5, where 8 signals are displayed for $\Delta z = 0.5$ cm. Here a Bentheim sample with a diameter of 70 mm is used. This graph also shows that the signal shape reproduces quite well. The pressure signals were recorded with a sample frequency of 2 MHz. Subsequently, wave mode decomposition can be performed in the frequency domain. Phase velocities and damping coefficients can be obtained.

V. FREQUENCY DOMAIN RESULTS

Once the large number of pressure signals is available, Prony's technique can be applied. It is based on the assumption that at any pressure gauge position $(n-1)\Delta z$ where $n=1,\ldots,N$, the recorded pressure $\hat{p}(\omega,n)$ can be written in terms of the pressure components $\tilde{p}_m$, where $m=1,\ldots,M$:

$$\hat{p}(\omega,n)=\tilde{p}_1 e^{-ik_1(n-1)\Delta z}+\cdots+\tilde{p}_M e^{-ik_M(n-1)\Delta z}. \quad (1)$$

For further details of Prony's method we refer to Lang *et al.*[9] and Marple.[8] The output of Prony's method is the ensemble of wave number k_m and pressure amplitudes $\tilde{p}_m$ at each frequency. This means that both the phase velocity and damping coefficient can be determined experimentally as a function of the frequency. Part of these wave numbers account for noise and are called spurious wave numbers. The spurious wave numbers have to be separated from the true wave numbers. Braun and Ram[16] and Grosh and Williams[17] present editing methods to obtain the true wave numbers. We previously demonstrated that modified versions of these editing methods could effectively be used for our experiments.[15]

We now compare the experimental and theoretical phase velocities and damping coefficients for natural Bentheim sandstone and synthetic rock. In Fig. 6, the solid lines indicate the theoretical phase velocities for the Bentheim sandstone cylinder whose properties were determined in independent laboratory experiments (Table III). The computations comprised the development of a Newton–Raphson solver in the complex wave number domain.[15] The compressibility of the grains and the elasticity of the shock tube wall are included in the numerics. We find that an infinite number of wave modes is generated that satisfy the boundary conditions at the interfaces between the sample and the annulus and between the annulus and the shock tube wall. Seven of these wave modes are drawn in Fig. 6. The modes *S*, *L*1, and *L*2

TABLE III. Bentheim parameters.

Diameter (mm)	70.0
Density of the solid ρ_s (kg/m^3)	2620
Porosity ϕ	0.22
Permeability k_0 (10^{-12} m^2)	2.7
Tortuosity	2.9
Constrained modulus K_p (GPa)	17
Shear modulus G (GPa)	8.0
Bulk modulus of the grains K_s (GPa)	36.5

are propagative over the entire frequency range, whereas the higher order modes $L3$–$L6$ have cut off frequencies.

The circles in Fig. 6 represent the experimental results of the Bentheim cylinder having a top surface sealing of a two-component epoxy resin in order to enhance surface wave generation and suppress the higher-order modes. For all experiments, the sidewall of the cylinder had open pores. The crosses in Fig. 6 represent an experiment where also the top surface had open pores in order to enhance the generation of the higher-order modes. We notice that the data represented by circles in Fig. 6 are in excellent agreement with the *S* mode predictions over the entire frequency range. The *S* mode corresponds to the pseudo-Stoneley wave in the high-frequency limit and is a bulk wave in the low-frequency limit. We found that the sealing at the top surface of the cylinder enhances the excitation of the *S* mode indeed. The data indicated by the crosses form two different lines. The lower line shows good agreement with the $L1$ mode. Both for the $L1$ and the *S* mode, the theory is slightly overestimating the results. This is probably due to the fact that the matrix bulk moduli were measured under dried conditions, whereas the actual experiments take place when the sample is water saturated. Below 20 kHz the data points were omitted, as the scatter in the data was high. A strong decrease of the phase velocity of the $L1$ mode in between 25 and 50 kHz is observed in the experimental results, as well as an asymptotic behavior above 50 kHz. The upper line of crosses in Fig. 6 is more difficult to explain, but clearly can be attributed to the higher-order wave phenomena. These modes generate relatively low pressure amplitudes in the annulus,[15] hence they are difficult to detect. The results of Fig. 6 show that we are able to excite multiple wave modes in one experiment over a broad frequency range. This is noteworthy, as most experimental setups are able to excite either one wave mode over a broad frequency range, or multiple wave modes in a limited frequency range.

Figure 7 shows the corresponding imaginary parts of the wave numbers. Again these plots result from a combination of 30 pressure signals, from which the noise has been removed in the output of Prony's method. The scatter in these measurements, however, is large. Only the results for the $L1$ mode suggest qualitative agreement between 30 and 40 kHz. Ellefsen *et al.*[18] and Hsu *et al.*[10] also report large bias in the attenuation coefficient in the presence of noise.

Experiments were carried out for a synthetic rock sample with a diameter of 70 mm. The parameters are given in Table II. The corresponding shear velocity is lower than the water velocity, and hence this rock sample is a so-called slow formation. The experimental and theoretical phase velocities are given in Fig. 8. For this experiment the smallest amount of scatter was obtained for a sample which was sealed at the top surface. The $L1$ mode is based on the wave motion in the annulus between the porous cylinder and the shock tube wall at high frequencies. At low frequencies it is based on longitudinal wave motion in the shock tube. The *S* mode for this slow formation corresponds to a so-called true Stoneley wave at high frequencies. Its phase velocity is lower than the Biot slow wave velocity. At low frequencies the *S* mode is related to bulk wave motion. For this slow

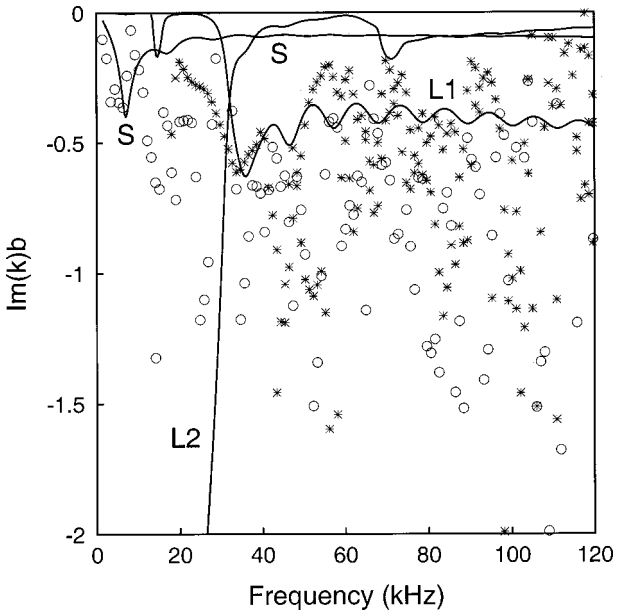


FIG. 7. Comparison between experiment and theory for the damping coefficient. Solid lines: wave mode computations; (○) experiments with a cylinder which was sealed at the top surface; (∗) experiments with open pores at the top surface. We used a Bentheim sandstone cylinder with a diameter of 70.0 mm. The damping coefficients are scaled by the inner radius of the shock tube *b*.

formation we also found so-called *M* modes. These modes have the velocity of the Biot fast wave as their high-frequency limit.[15] The experimental results in the frequency domain show one dominant wave mode over the entire frequency range with a slightly decreasing phase velocity (see Fig. 8). The experimental wave mode is in excellent agreement with the theoretical predictions for the $L1$ mode. Note that these predictions are based on parameters which have been determined in independent laboratory experiments. The amplitude of the surface wave was too small to be detected

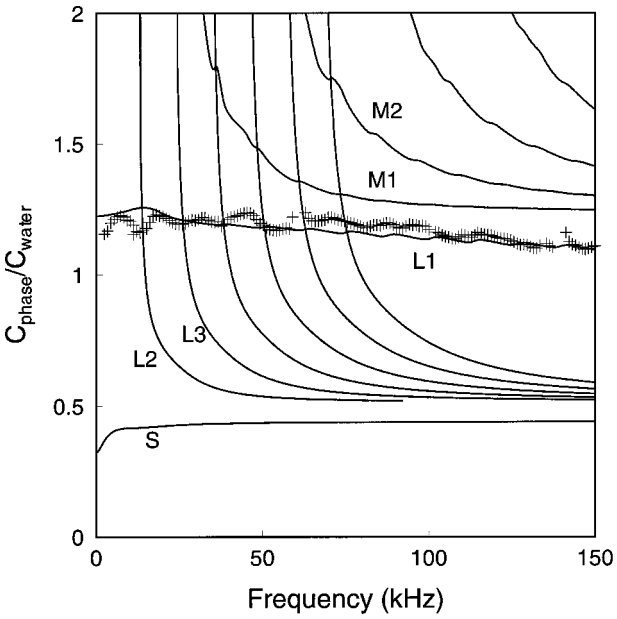


FIG. 8. Experimental and theoretical phase velocities for a synthetic rock with a diameter of 70.0 mm. The solid lines denote the computations for the *L*, *M*, and *S* modes. The top surface was sealed.

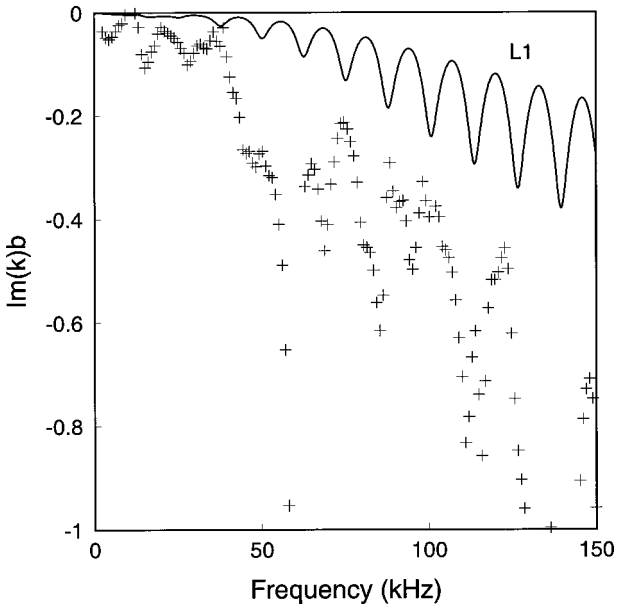


FIG. 9. Experimental and theoretical damping coefficients for a synthetic rock with a diameter of 70.0 mm. The solid line denotes the computation for the $L1$ mode. The top surface was sealed.

using Prony's method. This is consistent with the modeling of the S mode, which shows that the relative amplitude of the annulus pressure is small.[15]

The corresponding results for the damping coefficients are given in Fig. 9. Again, only qualitative agreement was obtained for the $L1$ mode although the signal-to-noise ratio was considerably better than in the Bentheim experiments. Nevertheless, experiments and theory show the same trend and the orders of the magnitudes are also in agreement.

VI. DISCUSSION AND CONCLUSIONS

We presented shock tube experiments on water-saturated porous cylinders. The sample properties were determined separately. A full modeling of the experimental configuration was implemented using linear Biot theory. We found that the dimensions of the annulus between the porous cylinder and the wall of the shock tube play an important role in the excitation of guided wave modes. The results in the time domain show that for a small annulus only bulk waves are generated and the one-dimensional Biot theory applies. For a larger annulus width, two-dimensional effects play an important role and multiple guided wave modes are excited. Using a newly developed traversable positioning system in the shock tube, we were able to record a precise distribution of the pressure over the sample length. From a modified Prony's technique, we measured frequency-dependent phase velocities in the 1–120 kHz frequency range. For Bentheim sandstone, we observed both the Stoneley wave and the first higher-order bulk mode. We report large bias in the attenuation coefficients of the wave modes. For the synthetic sample, the $L1$ mode could clearly be detected. Due to the high signal-to-noise ratio in this case we also measured reasonable data for the damping coefficient. Good agreement was found between our experimental data and the theoretical predictions for the phase velocities of both the Bentheim and the synthetic sample.

[1] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[2] 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).
[3] T. J. Plona, "Observation of a second bulk compressional wave in a porous medium at ultrasonic frequencies," Appl. Phys. Lett. **36**, 259–261 (1980).
[4] D. L. Johnson, T. J. Plona, and H. Kojima, "Probing porous media with first and second sound. II. Acoustic properties of water-saturated porous media," J. Appl. Phys. **76**, 115–125 (1994).
[5] O. Kelder and D. M. J. Smeulders, "Observation of the Biot slow compressional wave in water-saturated Nivelsteiner sandstone," Geophysics **62**, 1794–1796 (1997).
[6] L. Kelders, W. Lauriks, and J. F. Allard, "Surface waves above thin porous layers saturated by air at ultrasonic frequencies," J. Acoust. Soc. Am. **104**, 882–889 (1998).
[7] K. W. Winkler, H.-S. Liu, and D. L. Johnson, "Permeability and borehole stoneley waves: Comparison between experiment and theory," Geophysics **54**, 66–75 (1989).
[8] S. L. Marple, *Digital Spectral Analysis: With Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1987).
[9] S. W. Lang, A. L. Kurkjian, J. H. McClellan, C. F. Morris, and T. W. Parks, "Estimating slowness dispersion from arrays of sonic logging waveforms," Geophysics **52**, 530–544 (1987).
[10] C.-J. Hsu, S. Kostek, and D. L. Johnson, "Tube waves and mandrel modes: Experiment and theory," J. Acoust. Soc. Am. **102**, 3277–3289 (1997).
[11] J. G. van der Grinten, M. E. H. van Dongen, and H. van der Kogel, "A shock tube technique for studying pore-pressure propagation in a dry and water-saturated porous medium," J. Appl. Phys. **58**, 2937–2942 (1985).
[12] J. G. van der Grinten, M. E. H. van Dongen, and H. van der Kogel, "Strain and pore pressure propagation in a water-saturated porous medium," J. Appl. Phys. **62**, 4682–4687 (1987).
[13] R. W. J. M. Sniekers, D. M. J. Smeulders, M. E. H. van Dongen, and H. van der Kogel, "Pressure wave propagation in a partially water-saturated porous medium," J. Appl. Phys. **66**, 4522–4524 (1989).
[14] D. M. J. Smeulders and M. E. H. van Dongen, "Wave propagation in porous media containing a dilute gas-liquid mixture: Theory and experiments," J. Fluid Mech. **343**, 351–373 (1997).
[15] C. J. Wisse, "On frequency dependence of acoustic waves in porous cylinders," Ph.D. thesis, Delft University of Technology, 1999.
[16] S. Braun and Y. M. Ram, "Determination of structural modes via the Prony model: System order and noise induced poles," J. Acoust. Soc. Am. **81**, 1447–1459 (1987).
[17] K. Grosh and E. G. Williams, "Complex wave-number decomposition of structural vibrations," J. Acoust. Soc. Am. **93**, 836–848 (1993).
[18] K. J. Ellefsen, D. R. Burns, and C. H. Cheng, "Homomorphic processing of the tube wave generated during acoustic logging," Geophysics **58**, 1400–1407 (1993).

A systematic study of water-filled submerged elastic spherical shells and the resolution of elastic- and water-included resonances

Michael F. Werby
Naval Research Laboratory, Stennis Space Center, Mississippi 39529 and Department of Physics, The Catholic University of America, Washington, DC 20064

Herbert Überall
Department of Physics, The Catholic University of America, Washington, DC 20064

(Received 10 September 2000; accepted for publication 11 February 2002)

Water-filled elastic shells present more complicated backscattered echoes when compared to evacuated shells and are less well understood. The objective here is to determine and explain features peculiar to such objects and to use these features to aid in object characterization. Towards that end, an acoustic background suitable for such targets in conjunction with ''residual'' partial wave analysis obtained from subtracting the background from the elastic response is used. This analysis is aided by the determination and comparison of *plate modes* and resonance locations of *evacuated unloaded*, *evacuated loaded*, and *loaded water-filled* shells. Analysis of water-filled shells suggests that there exist isolated narrow, uniformly spaced eigenmodes in addition to the elastic modes. The broader elastic modes appear split in comparison to their unloaded counterparts. Otherwise, the elastic modes are not greatly affected by the included eigenmodes. The isolated (included) eigenmodes are extremely well modeled by a water inclusion in an infinite elastic matrix composed of the elastic shell material. The splitting of elastic modes proves to be due to interference of the *included eigenmodes* and the elastic body resonances. This leads to what appear to be separate branches of dispersion curves related to the elastic modes. Because of the numerous and usually evenly spaced included eigenmodes, dispersion curves are dominated by the abundance of these modes. A clear picture of the physical processes emerges from this analysis that explains all features of this event. Simple sets of rules that lead to tractable calculations are introduced that facilitate analysis of this interesting physical process.
[DOI: 10.1121/1.1467673]

PACS numbers: 43.30.Jx, 43.40.Ey [ADP]

I. INTRODUCTION

Our aim is to carry out a detailed analysis of acoustical scattering from loaded water-filled shells. Many of the details in this analysis are similar to those of scattering from evacuated shells, which in turn have common characteristics with the free vibrations of spherical elastic shells. The method of characterizing the modes on unloaded evacuated shells is quite analogous to the unloaded infinite plate problem. Thus, we discuss some of the details of those objects in order to add clarity to this analysis. Much of the insight that one may bring to the problem of spheres may be obtained from the much easier problem of infinite flat plates.[1,2] Even the terminology for waves on spheres has been borrowed to describe and classify their resonances or modal vibrations from those of plates.

In this study we make use of residual partial wave analysis (RPWA) to interpret details of loaded spherical shells. The RPWA consists of calculating the partial waves in the Rayleigh series for each elastic mode and subtracting that due to the partial waves of an acoustic background. Acoustic backgrounds should adequately describe scattering of the object in the absence of resonances. Thus, RPWA will bring out the resonances of the process. The practical use of residual partial wave analysis is described, which is crucial in this analysis. The RPWA indicates that in addition to elastic waves the water-filled shell scattering amplitude consists of sharp, uniformly spaced, very narrow resonances, which leads to the conclusion that they are associated with eigenmodes of the enclosed water. Thus, a model is proposed and used to describe the apparently isolated eigenmodes due to the disturbance of the water contained in the shell. In the following sections relevant commentaries on unloaded infinite plates, unloaded evacuated shells, and loaded evacuated shells are given to aid in the analysis of fluid-filled shells. The eigenvalue problem for a fluid in an infinite elastic matrix is developed and shown to account for some of the features of submerged fluid-filled shells. The analysis of residual partial waves is used to aid in the classification of resonances, and the dispersion curves of circumferential waves on fluid-filled submerged shells are interpreted along with the resonance mechanisms responsible for the details of these events.

II. USEFUL EXPRESSIONS FOR PROPER LAMB MODES ON UNLOADED PLATES

It is useful in this analysis to be familiar with results from infinite unloaded plates. The rationale for this brief examination is that it will be shown later that the critical fre-

quencies and behavior of shells is very close to that of results of the more manageable infinite plate. The equations that describe the motion of unloaded plates are obtained from the generalized Hooke's law. Two independent equations emerge that describe symmetric and antisymmetric plate modes. The expressions account for interface and proper Lamb modes. In the neighborhood of the critical frequency of Lamb modes, the general dispersion curves for large values of the phase velocity V_x for the Lamb modes reduce to[1]

$$(V_x/V_T)^2 \tan(df\,\pi/V_T) = -4(V_T/V_L)\tan(df\,\pi/V_L), \tag{1}$$

$$(V_x/V_T)^2 \tan(df\,\pi/V_L) = -4(V_T/V_L)\tan(df\,\pi/V_T), \tag{2}$$

where V_L and V_T are the compressional and shear bulk velocities, respectively. These expressions are used to obtain the critical frequencies, which are defined in the limit as the Lamb phase velocity V_x approaches infinity. Equation (1) leads to two expressions for symmetric modes

$$K_T d = \pi \frac{V_L}{V_T}(2n-1), \tag{3}$$

$$K_T d = 2\pi m \tag{4}$$

$(K_T = \omega/V_T)$, where n and m are positive integers. The critical frequencies for the antisymmetric modes are obtained from Eq. (2), which are

$$K_T d = 2\frac{V_L}{V_T}\pi m, \tag{5}$$

$$K_T d = \pi(2n-1). \tag{6}$$

Here, d is half the thickness of the plate. The terminology used to describe these modes follows from the spatial symmetry of the modes and the numerical ordering of each of the symmetries that follows from Eqs. (3)–(6). All modes from either Eq. (3) or (4) are listed as S_i (symmetric modes) of order $i>0$ and all modes that arise from Eqs. (5) and (6) are listed as A_i (antisymmetric modes) of order $i>0$. The factor that determines what the ordering is follows strictly from the numerical ordering of each symmetry series. Clearly, the ordering varies with material. For example, Eq. (4) defines an S_1 mode when $V_L > 2V_T$, while Eq. (3) defines the S_2 with the converse being the case otherwise. This is rather remarkable because of the peculiar and atypical nature of the S_1 resonance.

III. DISPERSION CURVES OF PHASE VELOCITIES OF UNLOADED, EVACUATED SPHERICAL SHELLS

This problem, like that of the infinite plate, is an eigenvalue problem. To pose the eigenvalue problem, we make use of the normal and transverse traction τ on the outer and inner surfaces of the shell[3,4]

$$\tau_{rr} = -\lambda k_d^2 \varphi + 2\mu\left\{\frac{\partial^2}{\partial r^2}\left[\varphi + \frac{\partial}{\partial r}(r\psi)\right] + k_s^2 \frac{\partial}{\partial r}(r\psi)\right\}, \tag{7}$$

$$\tau_{r\theta} = \mu\left\{2\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial \theta}\left[\varphi + \frac{\partial}{\partial r}(r\psi)\right]\right) + k_s^2 \frac{\partial \psi}{\partial \theta}\right\}, \tag{8}$$

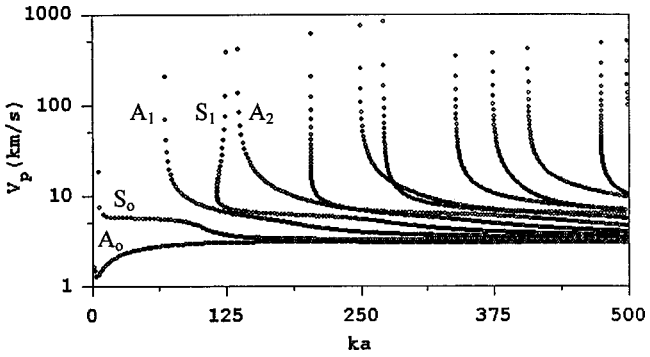


FIG. 1. Dispersion curves for evacuated unloaded 10% steel shell.

$$\varphi = \sum_k i^k \sqrt{2k+1}(b_k j_k(k_d r) + c_k y_k(k_d r)) P_k(\cos\theta), \tag{9}$$

$$= \sum_k i^k \sqrt{2k+1}(e_k j_k(k_s r) + g_k y_k(k_s r)) P_k(\cos\theta). \tag{10}$$

Here, the requirement is that $\tau_{rr} = \tau_{r\theta} = 0$ on the outer and inner surfaces of the shell for stress-free surfaces. These two conditions, along with the expansion of the displacement potentials and the boundary conditions at the two free surfaces, provide four equations and four unknowns. Since these expressions are all set to zero, because of linear independence of the expressions, a necessary condition is that the determinant of their coefficients is zero. We designate that determinant for each partial wave k by D_k. Usually one determines only the zeros of these functions. However, interesting insight can be obtained by plotting D_k as a function of ka. When its value crosses the real axis, there is of course an eigenvalue. However, it also graphically illustrates any symmetries of the Lamb modes, double-valued roots indicating the presence of two modes with the same critical frequency, and also illustrates a trend in the roots as the partial waves increase. Plots of the D_k's as a function of ka serve the same function as the residual partial wave analysis does for the loaded cases. Figure 1 illustrates the dispersion curves for an unloaded, evacuated steel shell. These curves are very much like the analogous plate dispersion curves. The location of the resonances is seen in Figs. 2(a)–(c) [plots of D_k vs ka] to agree with the predictions from modified versions of Eqs. (3)–(6). Figure 2(a) illustrates the critical frequency while Fig. 2(b) illustrates the trends in the resonances. Note that some of these progress rapidly while others seem to remain close to critical frequency. The latter waves have a rapid falloff in phase velocity.

IV. INTERFACE WAVE ON SUBMERGED ELASTIC SHELLS

Acoustic waves incident on submerged elastic shells cause the excitation of interface waves generated on the surface of the shell and proper Lamb modes generated in the body of the shell.[5–9] There are two interface waves: the antisymmetric A_0 or flexural mode and symmetric S_0 or dilatational mode. The phase velocity of A_0 with increasing frequency ranges from zero to the Rayleigh phase velocity.

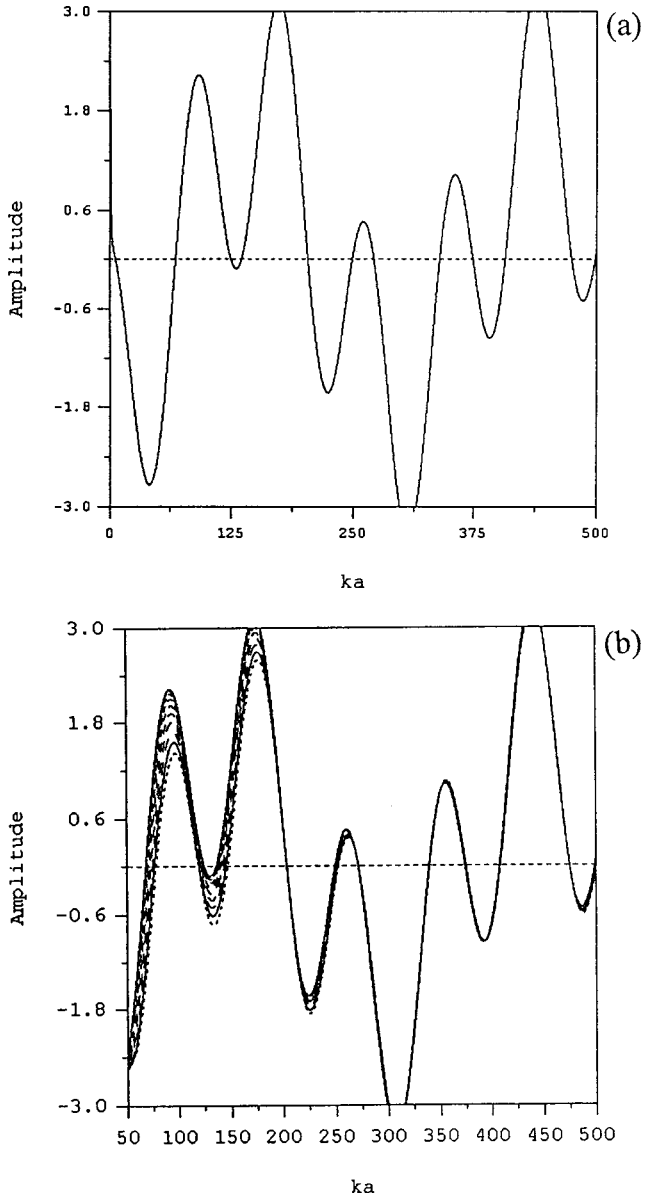


FIG. 2. (a) Plot of D_0 illustrating the roots of the unloaded shell; (b) Plot of D_0 through D_9 illustrating the trend of the roots of the unloaded shell.

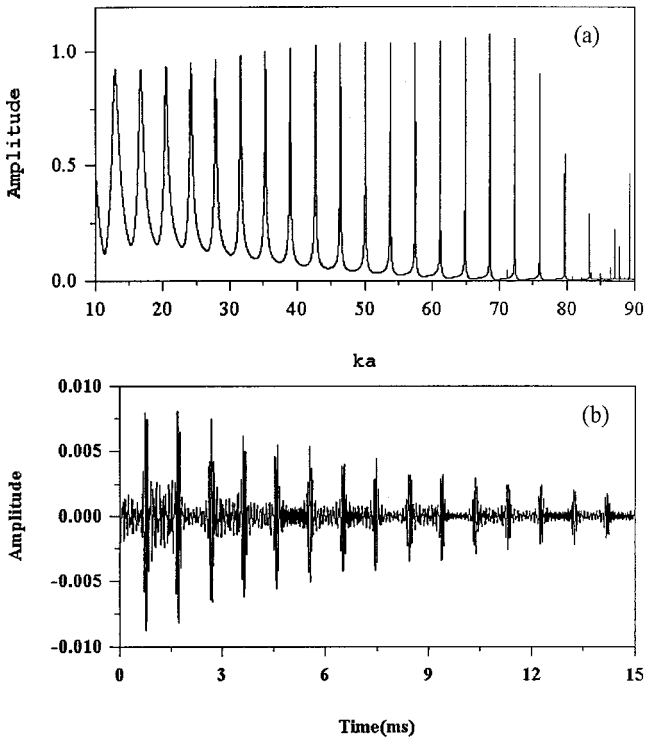


FIG. 3. (a) Residual form function for 1% evacuated steel shell over constant phase velocity region; (b) Pulse return signal for 1% steel shell over constant phase velocity region.

When there is water loading this wave does not radiate into the water until its phase velocity reaches that of the sound speed of water. This wave behaves like a flexural wave at low frequencies, is inhomogeneous, and at high enough frequencies it behaves in a manner similar to Rayleigh waves where it is by and large a surface wave with rapid attenuation away from the surface. The phase velocity of the S_0 wave varies from a value at the lower frequency of about three times the flat plate velocity[1,8] $C_P = V_T\sqrt{2/(1-\nu)}$ to the Rayleigh phase velocity V_R for large frequencies (ν = Poisson ratio). The increase in value of the phase velocity at the lower frequencies over the flat plate value is due to curvature and is typically about $2.7C_p$. However, its value quickly levels out to the plate velocity and reaches a plateau over a frequency zone with a width related to the Poisson ratio and the shell thickness.[10] The range in *ka* for which this is valid may be inferred from plate theory by making use of the approximate range that $\tanh(x) = x$ is satisfied, which is about 98% for $x = 0.25$ and about 92% for $x = 0.5$. This leads to an accuracy of 92% or better over the range of the zone. The upper limit of the zone is defined by[10] (V_w = sound speed in water)

$$ka = \frac{V_T}{V_w}\frac{a}{d}K_T d \leqslant \frac{0.5}{\sqrt{1 - V_x^2/V_L^2}} = 0.5\frac{V_T}{V_w}\frac{a}{d}\frac{2\nu}{1-2\nu}\sqrt{\frac{2\nu}{1-\nu}}. \tag{11}$$

For $\nu = 1/3$, $K_T d = 0.5$ for 98% accuracy and $K_T d = 1$ for a 92% accuracy, which is a factor of (2/3) lower than the critical frequency of the lowest proper antisymmetric Lamb mode and a factor of (1/3) lower than the lowest proper symmetric Lamb mode, which are discussed below. For a 1% shell, the *ka* zone is between 5–96.[7,11] This effect is illustrated in Fig. 3(a), which is the residual form function for a 1% steel evacuated shell. The S_0 resonances in this zone are narrow and of uniform spacing. The spacing between resonances for a sphere in water with the speed V_w and phase velocity V_p is simply $\delta ka = V_p/V_w \approx 3.8$ for typical materials. The group velocity is close to that of the plate velocity since the phase velocity is almost constant in this frequency zone. The value for $2N$, the number of resonances in the plateau zone, is obtained from $2N = (96\text{-}5)/3.8 = 32$. Because of these conditions, they particularly yield very interesting pulse returns over the plateau region due to a reinforcement mechanism that renders the pulse return to have a dramatic telltale signal. It is known to take the form[7]

$$P(t)=2^N\Delta_\omega\{\sin(\omega_0 t)\}^N\cos(V_g t-r)e^{-\Delta_\omega t}, \tag{12}$$

where ω_0 is the carrier or midfrequency value in the plateau zone and Δ_ω is the mean resonance width. Thus, one gets a long ringing resonance with a telltale beat pattern with an enhancement factor of 2^N. This is illustrated in Fig. 3(b), which is a pulse return with a bandwidth over the region in which the resonances are uniformly spaced. The resonance width is a function of material so that the exponential falloff varies with material. This plateau region is very important for target identification. As frequencies increase, the S_0 mode also approaches Rayleigh wave properties. Both A_0 and S_0 waves behave in a similar manner in their asymptotic zones. At that point, neither mode has a specific symmetry. In those zones, the two differ mainly in phase. The mechanism responsible for producing resonances is due to traveling waves excited on the surface that circumnavigate the shell along geodesic patterns traversing in paired opposite directions. At prescribed frequencies, the waves traveling in opposite directions interfere to form standing waves that account for the resonances over small intervals in frequency or duration.[3,5] These standing waves radiate energy back into the water (they are leaky). The superposition of these standing waves with the specular component, particularly when measuring backscattered signals, forms the resonance signatures due to the coherent sum of the two components, particularly when backscattered signals are measured. Since these resonances couple with the ambient fluid, they exhibit a resonance width identified with the energy leakage, which may also be associated with the ''lifetime'' of the resonance as readily derived from Fourier analysis. Moreover, they are characterized by a phase change of typically 180°, which is partly responsible for their resonance ''signature.'' At the inception of the A_0 interface wave there is an enormous resonance response. This occurs when the phase velocity of the A_0 wave equals that of the ambient fluid and therefore couples with the fluid and begins to radiate energy. Initially, these resonances are broad enough to have contributions from overlapping partial waves.[3] Also, the phase of the pressure is relatively constant until coincidence frequency. There, the phase goes from a real to a pure imaginary value[5,12] which leads to a jump in phase of $\pi/2$, so that the pressure varies from essentially constructive interference to partly destructive interference. Thus, over this range the A_0 resonances first manifest themselves as a large hump,[3] roughly three times or more larger in amplitude then the usual resonances; see Fig. 4(a). One may recall another interesting effect on flat plates that are fluid loaded on one side and evacuated on the other.[2,3] In the region about coincidence frequency the plate equations admit a real root at values just below coincidence frequency referred to as Stoneley waves. This effect also exists for shells and one observes narrow, subsonic water-borne waves circumnavigating the sphere (referred to here as pseudo-Stoneley waves) that actually produce strong, very narrow resonances.[2,3] Note that these waves are not related to Franz or creeping waves, which are circumferentially diffracted waves prominent at lower frequencies and are not related to resonances. The pseudo-Stonenely waves add coherently with the incipient A_0 resonances, leading to the appearance

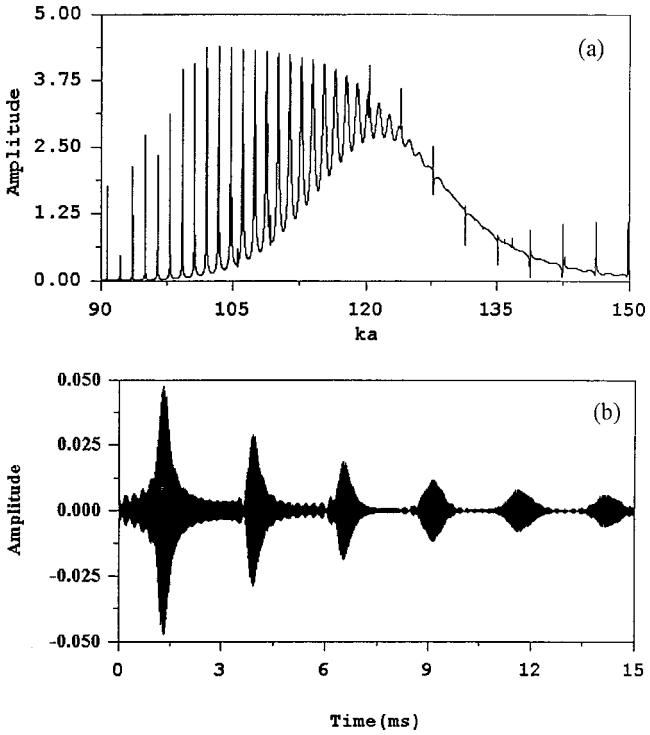


FIG. 4. (a) Residual form function for 1% evacuated steel shell about coincidence frequency. (b) Pulse return for 1% evacuated steel shell about coincidence frequency.

of a hump with large spikes on the upslope and only abated spikes apparent on the down slope. Figure 4(a) illustrates this effect for the residual form function for a 1% steel shell. The pulse returns in the frequency zone where pseudo-Stoneley resonances are prominent also lead to pronounced telltale signatures,[7,11] since they also satisfy the conditions that lead to Eq. (12). In this case the midfrequency value is somewhat higher, the enhancement factor is smaller, and the group velocity is about half that of the plate velocity for resonances compared with the plateau region for the S_0 resonances. Figure 4(b) illustrates this effect for a pulse signal with a bandwidth of a *ka* of 20 with the mid-*ka* at 120 for the 1% steel shell. Pseudo-Stoneley waves[3,7,11] typically have a group velocity of around 2.14 km/s. To get an idea of the *ka* location of coincidence frequency, one may use a simple, rather approximate expression from the literature to get[9,12]

$$[ka]_c=\sqrt{6}\frac{a}{d}\frac{V_T}{V_W}\sqrt{1-\nu}. \tag{13}$$

For a 1% shell for typical materials, this ranges from $ka=150$ for brass to $ka=76$ for WC and $ka=92$ for steel.

V. PROPER LAMB WAVES ON EVACUATED SUBMERGED ELASTIC SHELLS

In addition to the interface waves (A_0 and S_0), another class of waves is initiated in the body of the shell. These are the same type of modes as the Lamb modes discussed in Sec. III for plates. They interact with the ambient fluid as they travel around the annular region of the shell and they too initiate standing waves at the shell surface that radiate into

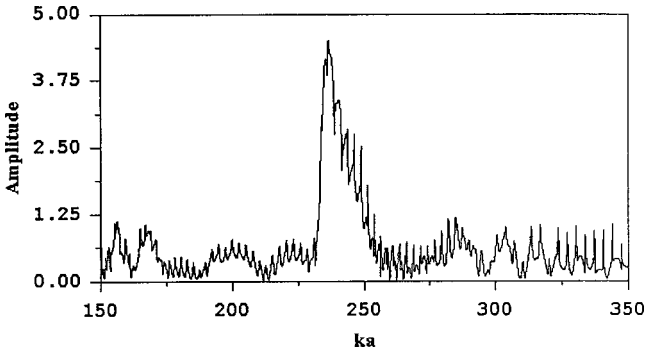


FIG. 5. Strong S_1 resonance for 5% steel shell.

the water at discrete frequencies. These waves, like the two lowest ''surface'' interface waves, may further be characterized as symmetric or antisymmetric relative to their motion with respect to a dissecting sphere at the midpoint of the outer and inner surfaces of the shell. As frequency increases the proper Lamb waves become more localized at the center of the annular region with asymptotic limits in phase velocity that approach that of the transverse bulk wave. Thus, as they behave less like surface waves they radiate more weakly. In their respective asymptotic zones, the interface waves behave as Rayleigh waves and specific proper Lamb modes become more concentrated at the center of the shell with phase velocities that approach the transverse phase velocity.[1] For both classes, the concept of symmetry no longer has meaning in their respective high-frequency limits.

The most pronounced Lamb resonance observed on shells is due to the first symmetric Lamb mode (S_1). Because of certain constraints in the shell equations,[10] this resonance has a cutoff value below critical frequency and is double-valued from the range of inception to critical frequency. If one plots the partial wave components, they too are double-valued in that region and since they are coherent over that range there is an additive effect that produces a very large resonance that is broad and about five times the strength of the usual resonance. Since this resonance is so pronounced and easily relates to one of the bulk velocities and the shell thickness, its measurement should be one of the most important aids in the detection and discrimination of submerged elastic shells. Furthermore, it is a function of uniform shell thickness and only one of the two bulk velocities and not shape. This effect is illustrated in Fig. 5 for a 5% steel shell. Another distinguishing property of all S_i modes is that there is a plateau in both phase and group velocity when the phase velocity is about equal to the longitudinal bulk velocity. There, the group velocity is also close to the bulk velocity. The *ka* value for which this occurs is[10]

$$ka = m\pi \frac{V_T}{V_W} \frac{a}{d} \sqrt{2(1-\nu)}. \tag{14}$$

For most materials, for a 1% thick shell the plateau occurs at $ka = 720m$ for all integers m. The systematics of all phenomena at this point such as the initial location of the resonances and their signatures is described here in concise form, since it must be used to infer the effect of the enclosed fluid problem. This allows us to interpret the remaining events due to water filling of shells. In summary, there are two pure interface waves that are symmetric (the S_0 mode) and antisymmetric (the A_0 mode), and two classes of an infinite number of proper Lamb modes in which some are symmetric (labeled S_i) and some are antisymmetric (labeled A_i) that proliferate with increasing frequency.

TABLE I. Partial-wave resonance frequencies (units *ka*) for proper Lamb modes on a submerged evacuated sphere.

Mode order	Antisymmetric	Symmetric
1	67.8	124.6
2	203.4	136.6
3	249.1	271.3
4	339.1	372.6
5	474.8	406.9
6	499.1	542.4

VI. RESIDUAL PARTIAL WAVES AND THE ANALYSIS OF PROPER LAMB MODES ON LOADED SHELLS

The defining geometry for a sphere does not yield such an easy analysis as that for infinite flat plates, nor does it even suggest the possibility of a good analogy. However, because proper Lamb modes occur at such high frequencies, one may take advantage of the asymptotic values of the spherical Bessel functions of the first and second kind providing one uses as the zero reference a sphere with diameter midway between the outer and inner surface of the elastic shell and arrives at essentially the same conditions for the critical frequencies expressed in Eqs. (1) and (2). To illustrate the close analogy of the resonances of each geometry, we determine the lowest resonances on a 10% thick shell. This works even for loaded spheres and plates. Below, we plot the residual partial wave amplitudes for several of the partial waves over the *ka* range from 0 to 500 to test the scheme. The following are the analogous rules that relate to spheres:

The symmetric modes

$$ka = \pi V_L / V_W (a/\delta a)(2n-1), \tag{15}$$

$$ka = 2\pi V_T / V_W (a/\delta a)m. \tag{16}$$

The antisymmetric modes:

$$ka = 2\pi V_L / V_W (a/\delta a)m, \tag{17}$$

$$ka = \pi V_T / V_W (a/\delta a)(2n-1). \tag{18}$$

Here, V_W, V_T, V_L, a, δa, and k are the speed of sound in water, the transverse and compressional bulk velocities, the shell radius, the shell thickness, and the wave number in water, respectively. The analogy is so close we may use the same terminology as that for flat plates. From Eqs. (15)–(18), Table I lists the partial wave resonance frequencies (units *ka*) of the first 12 proper Lamb modes for a 10% water-loaded, evacuated steel shell where $V_W = 1.5$, $V_T = 3.24$, $V_L = 5.95$ in units of km/s and $a/\delta a = 10$. These values agree nicely with those obtained from the partial wave

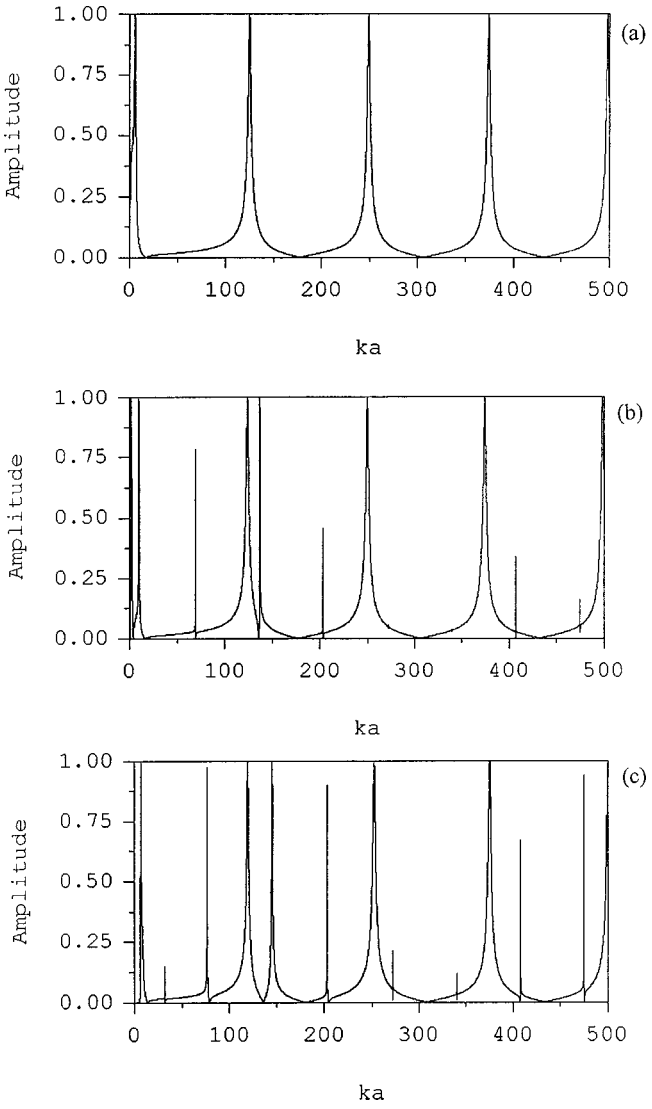


FIG. 6. (a) Residual partial wave amplitude for zeroth partial wave for submerged evacuated shell. (b) Residual partial wave amplitude for second partial wave for submerged evacuated shell. (c) Residual partial wave amplitude for ninth partial wave for submerged evacuated shell.

analysis of loaded shells (Fig. 6) and those obtained from the eigenvalue problem of the unloaded evacuated shell (Fig. 1). Note that the zeroth partial wave ($m=0$) can only sustain waves generated by the compressional wave. Not until the second partial wave will one see all of the waves generated by the transverse waves.[13] Some are too narrow to see. However, by the time one reaches the ninth partial wave all of the Lamb modes in that range are visible. In Figs. 6(a)–(c) we examine the residual partial waves for the fluid-loaded shell for different partial waves ($m=0$, 2, and 9).

VII. EIGENFREQUENCIES OF FLUID WAVES ENCLOSED IN A SUBMERGED ELASTIC SHELL

To completely validate this study we must account for the nature of the water disturbance in the elastic shell. Examinations of the partial waves that derive from scattering from fluid-filled submerged elastic shells suggest both elastic resonances and discrete eigenmodes. The elastic resonances are shown in the next section to overlie the elastic resonances of evacuated shells. Thus, it appears that the spectrum of resonances is composed of the elastic resonances and the discrete modes of the water enclosed in the spherical shell. To demonstrate this we must account for the included water modes. To this end, we treat a spherical fluid inclusion in an infinite elastic matrix. In the elastic material the governing equations for a spherically symmetric geometry are[4]

$$u_r=\frac{\partial}{\partial r}\left\{\varphi+\frac{\partial}{\partial r}(r\psi)\right\}+rk_s^2\psi, \tag{19}$$

$$u_r^0=\frac{\partial\varphi_0}{\partial r}, \tag{20}$$

$$P_0=\rho_0\omega^2\varphi_0, \tag{21}$$

where

$$u_r=u_r^0, \quad P_0=\tau_{rr}, \quad \text{and} \quad \tau_{r\theta}=0 \tag{22}$$

at the interface of the spherical water inclusion and the elastic matrix. The normal-mode expansions of the displacement potentials are

$$\varphi_0=\sum_k i^k a_k\sqrt{2k+1}j_k(k_0r)P_k(\cos\theta), \tag{23}$$

$$\varphi=\sum_k i^k b_k\sqrt{2k+1}h_k(k_dr)P_k(\cos\theta) \tag{24}$$

$$\psi=\sum_k i^k c_k\sqrt{2k+1}h_k(k_sr)P_k(\cos\theta). \tag{25}$$

Here, a_k, b_k, and c_k are unknown expansion coefficients and j_k, h_k are the regular and outgoing spherical Hankel functions and P_k the Legendre function of order k, respectively. Outgoing waves are required to satisfy the radiation boundary conditions. By making use of Eqs. (7), (8), (19)–(21), and the boundary conditions Eq. (22) with Eqs. (23)–(25), we obtain the following matrix relations for each partial wave k:

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & 0 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix}=0, \tag{26}$$

with the matrix elements given in the Appendix. Setting this expression to zero guarantees that it is a homogeneous problem with discrete eigenvalues. This requires that the determinant of the matrix is zero due to linear independence of the expansion coefficients, so that taking advantage of $m_{33}=0$ one may expand in minors to obtain

$$f_k=m_{13}(m_{21}m_{32}-m_{22}m_{31})-m_{23}(m_{11}m_{32}-m_{12}m_{31})=0 \tag{27}$$

for each order, k. The zeros of f_k define the discrete eigenfrequencies of the disturbance of the enclosed water. We use the following method to compare the eigenfrequencies generated from Eq. (27) with the actual values from excited water-filled shells. We plot f_k vs ka for several partial waves.

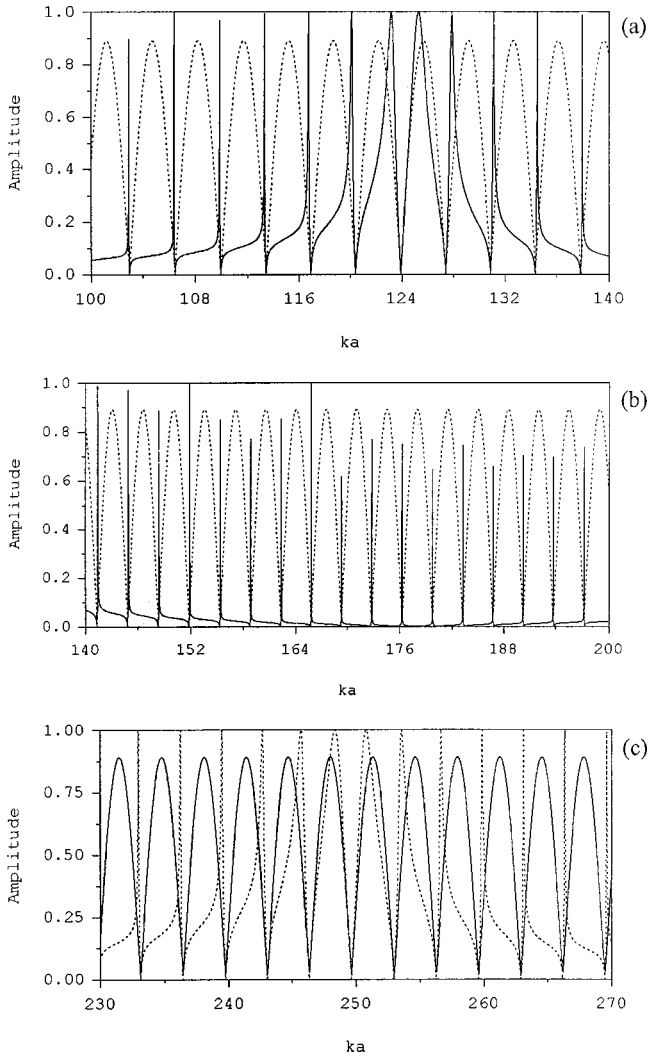


FIG. 7. (a) Comparison of included eigenmodes for fluid-filled elastic shell (solid curves) and water inclusion in an infinite matrix (dashed curves), $ka=100-140$. (b) Comparison of included eigenmodes for fluid-filled elastic shell (solid curves) and water inclusion in an infinite matrix (dashed curves), $ka=140-200$. (c) Comparison of included eigenmodes for fluid-filled elastic shell (solid curves) and water inclusion in an infinite matrix (dashed curves), $ka=230-270$.

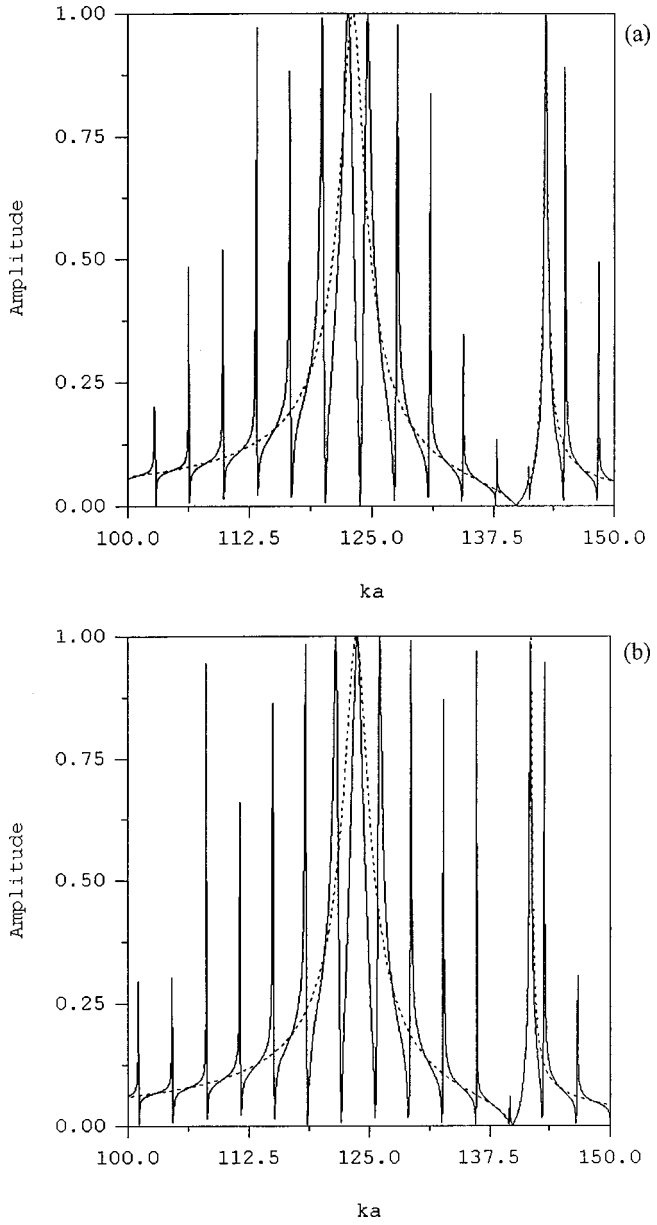
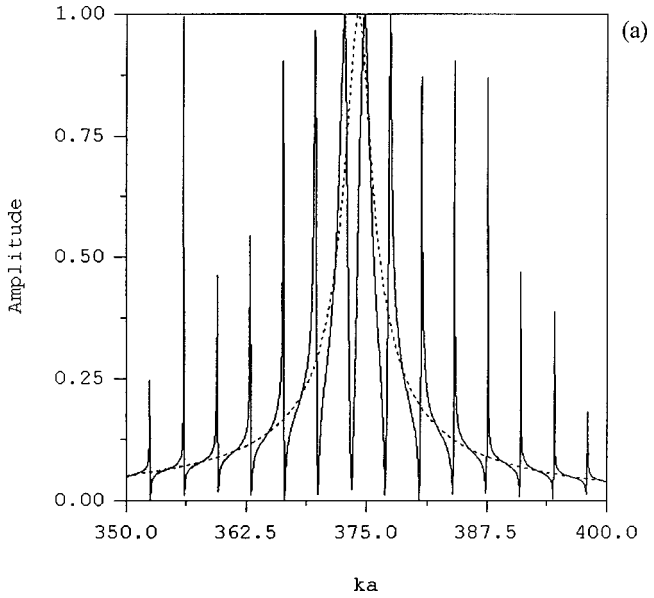


FIG. 8. (a) Comparison of partial waves for evacuated (dotted) and water-filled shells (solid curves), $ka=100-150$, partial wave 0. (b) Comparison of partial waves for evacuated-(dotted) and water-filled shells (solid curves), $ka=100-150$, partial wave 1.

The eigenfrequencies occur when $f_k=0$. Recall that the subtraction of the proper acoustic background from a specific partial wave results in *residual partial waves that are useful in illustrating resonance locations*. Plots of these quantities manifest the usual elastic resonances as well as separated spikes. We now know enough about submerged elastic evacuated shells that the elastic resonance locations are known. If we overlay this residual curve with the plots of f_k, then the spikes from the residual plots should coincide with the nulls of the plot of f_k from Eq. (27) for each partial wave if they are indeed pure eigenmodes as we proposed. Figures 7(a)–(c) are illustrations of the results. In each figure, the nulls of the dotted curves of f_k from Eq. (27) and the spikes from the residual plots exactly agree. Thus, the peaks of the residual plots (solid curves) come strictly from the elastic shell resonances and fall in exactly the correct place as for evacuated shells, with the exception of some broad resonances such as the proper symmetric Lamb modes. These modes (at $ka=124$ and 250) are seen to ''bifurcate'' [13] or ''sever'' due to interference with water-included eigenmodes. Thus, an included eigenmode splits the broad modes into two or more peaks when they overlap. In plots of the phase velocities, this can be the cause of an apparent repulsive manifestation[14] of the dispersion curves. At each ''severance'' the dispersion curves seem to split and are responsible for forming two separate branches of the elastic dispersion curves. Figures 8 and 9 illustrate a comparison of the plot of f in Eq. (27) with the residual partial waves. Note that at times broad elastic resonances appear to be severed [Figs. 8(a) and 9(a)] and at other times, when there is no overlap with the included eigenmodes, there is no severance [Figs. 8(b) and 9(b)]. What is clear from this is that the main elastic resonances are in their locations and that there are a large number of additional spikes that we must identify with the



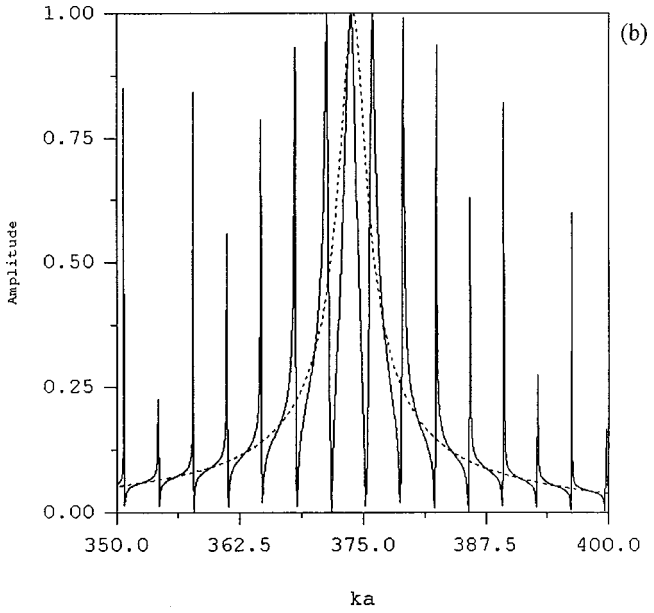


FIG. 9. (a) Comparison of partial waves for evacuated-(dotted) and water-filled shells (solid curves) $ka=350-400$, partial wave 0. (b) Comparison of partial waves for evacuated-(dotted) and water-filled shells (solid curves), $ka=350-400$, partial wave 1.

eigenmodes of the water inclusion. Another apparent and important feature is the interference between the broad resonances and the included eigenmodes. In these cases, the elastic resonances seem to split or are severed and form separate branches of the main (unloaded) elastic resonance. That is a manifestation of the plots; however, their origin is due to the interference phenomena.

VIII. THE RESONANCES OF FLUID-FILLED SUBMERGED ELASTIC SHELLS AND THEIR INTERPRETATION

Fluid-filled shells have been analyzed recently for the infinite cylinder problem.[15] Some of the dispersion curves illustrate rather interesting results. Here, we attempt to explain all features that are observed on these type of objects.

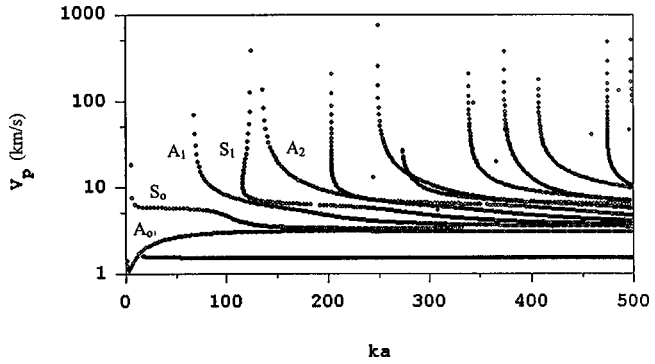


FIG. 10. Dispersion curves for evacuated water-loaded 10% steel shell.

We start for comparative purposes by illustrating (Fig. 10) the dispersion curves for an evacuated 10% steel shell. Figure 11 is a blowup of the dispersion curves (see also Ref. 16) illustrating the presence of two water-borne waves: one that begins and the other that ends at the point that the A_0 wave begins to radiate into the water. The lower wave (the *A* wave or the pseudo-Stoneley wave) is subsonic and terminates around coincidence frequency. The constant wave on the right with a phase velocity equal to that of sound in water is a symmetric (*S*) water-borne wave predicted[17] on plates or spherical shells, water loaded on one side and evacuated on the other; its partial waves are broad in width and weak in amplitude and at best contribute imperceptible resonance effects. A residual partial wave study has already been done for the water-filled shell in Sec. VI. In addition, in Sec. VII it was determined that the disturbance of the water filling is responsible for the very narrow, spiky resonances observed in the residual partial wave analyses. The determination of dispersion curves for this problem can be a formidable undertaking because of the large number of included eigenmodes. An analysis out to *ka* of 500 includes so many curves that it is impossible to illustrate them. Therefore, results are limited to *ka* of 100 for scatter plots and *ka* of 50 for continuous-line plots. Scatter plots are useful since they take raw discrete data (numerically generated here) and display trends. In the alternative it is possible to sort and keep track of each data point and associate that point with a specific mode. One may also examine the scatter plots and connect the discrete data with lines to illustrate the dispersion curves.

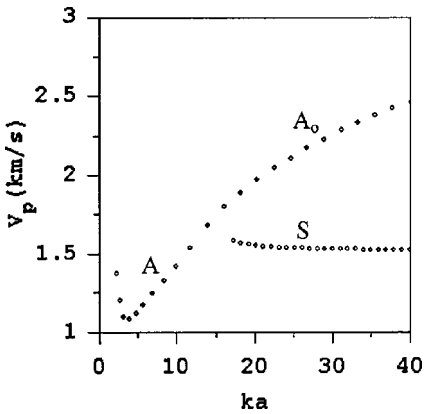


FIG. 11. Blowup of dispersion curves for evacuated loaded 10% steel shell for small *ka*.

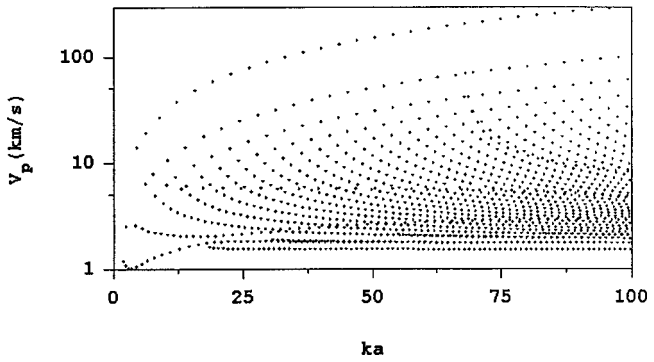


FIG. 12. Raw dispersion curves from raw data for water-filled submerged 10% steel shell.

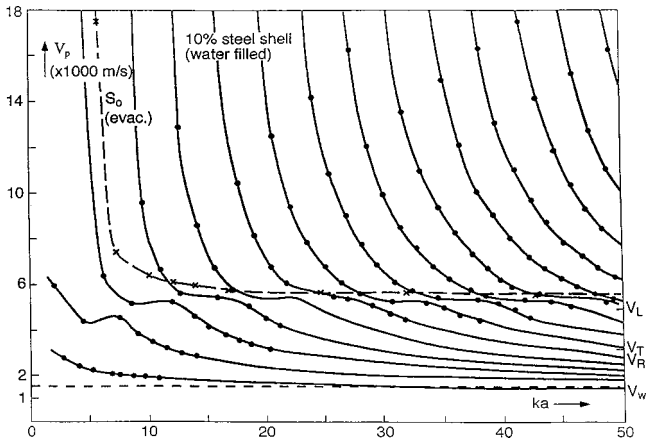


FIG. 13. Dispersion curve for water-filled submerged 10% steel shell.

The set of numerically generated data is inherently discrete and any connecting lines are included to ''guide the eye.'' Since the phenomena we deal with here have discontinuities at points related to the intermittent splitting of elastic resonances, then one must use additional knowledge to ensure meaningful plots. Further, determining what are roots and what are not in some brute force solution itself is a problem. However, residual partial wave analysis suggests a simplification of this problem by the following observation. In general, the partial wave amplitude may be expressed as follows:

$$f_i = \frac{j_i - j' A_i}{h_i - h_i' A_i}, \tag{28}$$

where A_i has been derived by Goodman and Stern[18] and is the ratio of two 6×6 determinants with complicated elements. In that formulation, A_i is real. The form of our acoustic background is[13]

$$b_i = \frac{g_i j_i - kaj'}{g_i h_i - kah_i'}. \tag{29}$$

To obtain the residual partial wave amplitude, we subtract b_i from f_i.

In the neighborhood of a resonance, the absolute value passes through unity. Numerically this requires that one use very small step sizes for the included eigenmodes. That means we take the expression

$$p_i = |f_i - b_i|, \tag{30}$$

and determine approximately the value of *ka* when p_i approaches unity to determine a resonance location. Away from a resonance p_i is always less than 1 and approaches zero. This significantly facilitates the calculation of the dispersion curves for this case. It is also possible to examine the relative phase of f_i and b_i and at a resonance there is a phase jump of π, and one may exploit that feature to determine resonance locations if one so chooses. The dispersion curves of water-filled shells obtained from this method are illustrated in Figs. 12 and 13. The notable feature comes from the large number of included eigenmodes that dominate the plots. They present fairly uniform dispersion curves; see Fig. 12. However, the elastic modes, including the region of broad resonances, are expected to have two branches at the point of inception of some of the modes, due to the severing of a single broad mode by an included eigenmode; see Figs. 8(a) and 9(a). The difference between Figs. 12 and 13 (apart from the obvious difference in *ka* range) is that Fig. 12 represents raw data and is a scatter plot from calculation. Figure 13 was calculated over a smaller range, with a coarser mesh and with lines drawn between discrete eigenvalues to guide the eye. Without the detailed analysis carried out here, that figure is subject to different possible interpretations. The analysis demonstrates that the appearance of repulsion[14] of some dispersion curves of this kind may be the result of destructive interference of broad elastic resonances and included eigenmodes. For water-filled shells, dispersion curves seem to lose value in analysis because of the proliferation of evenly spaced eigenmodes (Fig. 12) which tend to mask other effects and because an interpretation of these kinds of modes when equated with a phase velocity does not have the same usefulness as the elastic waves which are known to be circumferential and to be related to either interface or Lamb waves. The water-included eigenmodes are in fact related to modes in a uniform isotropic ideal fluid waveguide. The phase velocities of that problem are also uniform and evenly spaced.

IX. SUMMARY

This analysis clearly establishes that differential scattering cross sections of signals scattered from water-filled submerged shells are influenced by the coherent sum of the elastic response typical of an evacuated shell and the water-included eigenmodes that are well represented by a water inclusion in an infinite elastic matrix. Further, these two mechanisms interfere destructively and this is noticeable for the broader elastic resonances. The interference between broad elastic resonances and included eigenmodes renders a *severance effect*, leading to two or more peaks related to broad elastic resonances for specific partial waves. Since interference occurs in a specific way for each partial wave and only when there is sufficient overlap between the two phenomena, then some partial wave components will not be split. Corresponding dispersion curves for the elastic components at times will form two or more branches related to the splitting, or merge into a single branch when splitting is absent. Dispersion curves for water-filled shells without additional knowledge may not lead to a definitive understanding

of the entire physical process. The partial wave decomposition method, based on a new generalized acoustic background[13] and the observation that the included eigenmodes are well represented by water inclusions in an infinite elastic matrix, proved crucial in the interpretation of all events on water-filled shells.

ACKNOWLEDGMENTS

We are greatly indebted to the Acoustics Division under the leadership of Dr. E. Franchi of NRL Code 7100 as well as the director of NRL, Dr. T. Coffey, and the Office of Naval Research for the continued funding of this research.

APPENDIX: MATRIX ELEMENTS OF EQ. (26)

$$m_{11}=\{2n(n+1)-z_d^2\}h_n(z_d)-4z_d h_n' \quad m_{21}=z_d h_n'(z_d) \quad m_{31}=2[h_n(z_d)-z_d h_n'(z_d)]$$
$$m_{12}=2n(n+1)[z_s h_n'(z_s)-h_n(z_s)] \quad m_{22}=n(n+1)h_n(z_s) \quad m_{23}=-z_0 h_n'(z_0)$$
$$m_{13}=\frac{\rho_0}{\rho_1}z_0^2 h_n'(z_0) \quad m_{32}=2z_s h'(z_s)+[z_s^2-2n(n+1)+2]h_n(z_s) \quad m_{33}=0.$$

[1] I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[2] G. Quentin and M. Talmant, ''The plane plate model applied to scattering of ultrasonic waves from cylindrical shells,'' in *Proceedings of the International Conference on Elastic Wave Propagation*, edited by M. F. McCarthy and M. A. Hayes (Elsevier, North-Holland, Amsterdam, 1989).
[3] M. Talmant, H. Überall, R. D. Miller, M. F. Werby, and J. W. Dickey, ''Lamb waves and fluid-borne waves on water-loaded thin spherical shells,'' J. Acoust. Soc. Am. **86**, 278–289 (1989); M. F. Werby, ''Scattering from elastic shells at coincidence frequencies,'' Acoust. Lett. **15**, 39–42 (1991); M. F. Werby, and H. Überall, ''The excitation of waterborne waves at the interface of evacuated elastic spherical shells and pseudo-Stoneley resonances,'' J. Phys. IV 2, 1087–1090 (1992).
[4] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1951).
[5] H. Überall, L. R. Dragonette, and L. Flax, ''Relation between creeping waves and normal mode vibrations of a curved body,'' J. Acoust. Soc. Am. **61**, 711–715 (1977).
[6] H. Überall, I. K. Bjørnø, and L. Bjørnø, ''Dispersion of circumferential waves on evacuated, water-loaded spherical steel shells,'' Ultrasonics **37**, 673–675 (2000).
[7] M. F. Werby and J. W. Dickey, ''Transient Resonance Scattering from Bounded Objects'', in *Acoustic Resonance Scattering*, edited by H. Überall (Gordon and Breach Science Publishers, Newark, NJ, 1992), pp. 257–276. Bounded
[8] J. Ripoche and G. Maze, ''A New Acoustic Spectroscopy: The Resonance Scattering Spectroscopy by the Method of Isolation and Identification of Resonances (MIIR),'' in *Acoustic Resonance Scattering*, edited by H. Überall (Gordon and Breach Science Publishers, Newark, NJ, 1992), pp. 69–103.
[9] M. F. Werby, ''The isolation of resonances and the ideal acoustic background for submerged elastic shells,'' Acoust. Lett. **15**, 65–69 (1991); ''The acoustic background for submerged elastic shells,'' J. Acoust. Soc. Am. 90, 3279–87 (1991); ''Recent Developments in Scattering from Submerged Elastic and Rigid Targets,'' in *Acoustic Resonance Scattering*, edited by H. Überall (Gordon and Breach, Science Publishers, Newark, NJ, 1992), p. 277–303.
[10] M. F. Werby and H. Überall, ''The analysis and interpretation of special properties of higher order symmetric Lamb waves: The case for plates,'' J. Acoust. Soc. Am. 111, 2686–91 (2002).
[11] M. F. Werby and H. B. Ali, in *Computational Acoustics*, edited by D. Lee, A. Cakmak, and R. Vichnevetsky (Elsevier Science, North Holland, Amsterdam, 1990), Vol. 2, pp. 133–158.
[12] D. Ross, *Mechanics of Underwater Noise* (Pergamon, New York, 1976).
[13] M. F. Werby and N. A. Sidorovskaia, ''Modern Developments in the Theory and Application of Classical Scattering,'' in *Acoustic Interactions with Submerged Elastic Structures*, edited by A. Guran, A. De Hoop, D. Guicking, and F. Mainardi (World Scientific, Singapore, 2001), pp. 256–258.
[14] H. Überall, B. Hosten, M. Deschamps, and A. Gérard, ''Repulsion of phase velocity dispersion curves and the nature of plate vibrations,'' J. Acoust. Soc. Am. **96**, 908–917 (1994).
[15] X. L. Bao, H. Überall, P. K. Raju, A. C. Ahyi, I. K. Bjørnø, and L. Bjørnø, ''Waves on fluid-loaded shells and their resonance frequency spectrum,'' Proc. R. Soc. London, Ser. A (submitted).
[16] G. Kaduchak, C. S. Kwiakowski, and P. L. Marston, ''Measurement and interpretation of the impulse response for backscattering by a thin spherical shell using a broad bandwidth source that is nearly acoustically transparent,'' J. Acoust. Soc. Am. **97**, 2699–2708 (1995).
[17] J. P. Sessarego, J. Sageloli, C. Gazanhes, and H. Überall, ''Two Scholte-Stoneley waves on doubly fluid-loaded plates and shells,'' J. Acoust. Soc. Am. **101**, 135–142 (1997).
[18] R. Goodman and R. Stern, ''Reflection and transmission of sound by an elastic spherical shell,'' J. Acoust. Soc. Am. **34**, 338–344 (1962).

The significance of shape and orientation in single-particle weak-scatterer models[a)]

Constantin-C. Coussios[b),c)]
Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

(Received 11 September 2001; revised 11 June 2002; accepted 14 June 2002)

Particles that have a density and compressibility comparable to those of the surrounding medium (weak scatterers) are often modeled as spherical scatterers of equivalent volume. The simplicity and symmetry of the spherical model does not however account for the effect of the angle of incidence of the incident field onto the particle. By performing a comparative study of scattering models for the case of a fluid sphere, it is shown that models derived using the Born approximation closely match the exact solution for weak scatterers. This approximation is therefore applied to produce a model for the fluid disk, which is used to investigate the significance of particle shape and orientation. Scattering by a red blood cell of *given volume* modeled as a sphere is compared to the result obtained by approximating its shape as a disk of varying aspect ratio and orientation. The spherical model is shown to provide a good description for frequencies up to 20 MHz, beyond which particle shape becomes significant. This effect could go undetected if the scattered field is only observed at 90° relative to the direction of the incident field. Nevertheless, the significance of particle shape and orientation might form the basis of a novel detection technique. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1498852]

PACS numbers: 43.35.Bf, 43.20.Fn, 43.30.Ft [SGK]

I. INTRODUCTION

Models for the scattering of sound by particles, which have a density and compressibility that are only marginally different to those of the surrounding medium (*weak scatterers*), have been extensively developed in the past for a variety of particle shapes. Such models are often obtained as a special case of an exact analytical solution (Bowman *et al.*, 1987), derived by expanding the incident wave in spherical harmonic series, and using the fact that the normal velocity and acoustic pressure must be continuous at the scatterer surface. Alternatively, scattering models can be reached by using approximate integral techniques, which exploit the fact that the particle of interest is a weak scatterer, in order to approximate the total field inside the scattering volume with the incident field (Born approximation).

Exact analytical solutions present the advantage of being valid over the whole frequency range, whether the scatterer is weak or not, but are generally cast as an infinite series of spheroidal wave functions, which makes them difficult to comprehend. Those expressions can be considerably simplified in the long wavelength limit (Bowman *et al.*, 1987; Burke, 1966a, b), where the wavelength $\lambda_0 = 2\pi/k_0$ of the incident field is much larger than the characteristic dimension of the scatterer a. This type of solution was first developed by Rayleigh in 1872 for the case of a sphere (Rayleigh, 1872). Thus, particles for which the nondimensional size parameter $k_0 a = \omega a/c_0$ is considerably smaller than one and which can be safely modeled using a long-wavelength method came to be known as Rayleigh scatterers.

On the other hand, models obtained using integral techniques and the Born approximation yield simple results, but are only applicable to situations in which the magnitude of the scattered field in the region of the scatterer is much smaller than that of the incident field. This not only implies that the particle must be a weak scatterer, but also suggests a limit on the frequency range over which Born models are likely to remain accurate: those results are not reliable for values of $k_0 a$ close to one, or at wavelengths where phenomena such as resonance, or multiple scattering effects for multiple particles, are important (Shung and Thieme, 1993). These considerations have generated notable concern in the literature as to the range of values of $k_0 a$ over which such integral techniques are applicable.

Weak scatterer models, whether derived using exact methods or the Born approximation, are often applied to collections of randomly oriented particles. In this context, scatterers of arbitrary shape are commonly modeled as spheres of equivalent volume. Such a simple model may seem inadequate at first, considering that exact solutions are available for more complex shapes, like the prolate spheroid, the oblate spheroid, or the infinite cylinder. However, to use these results one would require *a priori* knowledge of scatterer orientation, which is unavailable in most experimental situations. In addition, spherical models appear to describe successfully scattering by collections of particles that are truly randomly oriented: drawing upon the work of Twersky (1964, 1988), Roos demonstrated that, for a red blood cell modeled as an oblate spheroid of aspect ratio 1:4, the scattered field averaged over all possible scatterer orientations agreed to within three significant figures with the value ob-

[a)]Part of this work was presented at the 141st Meeting of the ASA in Chicago, 4–8 June 2001, under the title "Scattering by a Red Blood Cell: The Importance of Particle Shape and Orientation."
[b)]Under the supervision of Shon E. Ffowcs Williams.
[c)]Electronic mail: coussios@email.uc.edu

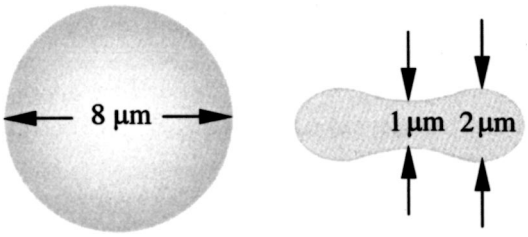


FIG. 1. Top view and cross section of a healthy red blood cell.

tained by treating the cell as a sphere of equivalent volume (Roos, 1983; Roos *et al.*, 1988).

Such observations have often led to the belief that shape and orientation are not a factor when modeling weak scatterers (Roos, 1988; Roy, 1987). The simplicity and symmetry of the sphere model does however prevent one from taking into account the effect of the angle of incidence of the incident field onto an asymmetric particle, such as a disk. Intuition tends to suggest that the scattered field would differ if the particle were impinged upon broad-side or thin-side, a consideration which would become important if a collection of particles were aligned in any particular plane. Although rare, such situations do present themselves when the scatterers are submitted to particular flow conditions or to directed electromagnetic fields. It has been suggested by many investigators (Caro *et al.*, 1978) that red blood cells, which normally present themselves as biconcave disks of dimensions shown in Fig. 1, will tend to orient themselves parallel to the planes of shear under laminar flow. To illustrate this, Fig. 2 shows a collection of red cells that have just been sheared by a microscope plate, causing them to align in the plane of the plate (Bessis, 1974).

Recent developments in ultrasonic transducer technology have made it possible to generate sound fields of increasingly short wavelengths, hence improving the imaging resolution. However, shorter wavelengths tend to push the nondimensional size parameter k_0a closer and closer to one, undermining the validity of long-wavelength models. With improved resolution, it is also likely that factors such as particle shape and orientation will become increasingly important. Two recent investigations (Allard *et al.*, 1996; Quin *et al.*, 1998) indicated that the power backscattered by flowing whole blood was dependent upon the angle of incidence of the incident field relative to the axis of the tube; in both

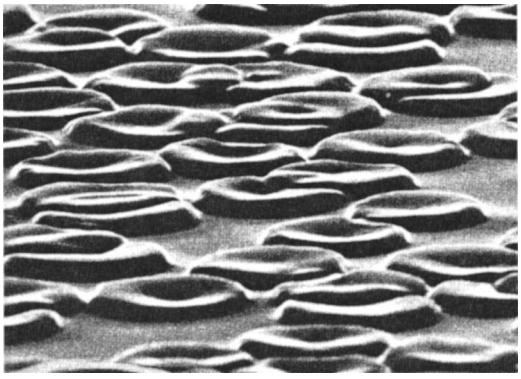


FIG. 2. Aligned red blood cells after shearing by a microscope plate.

TABLE I. Average dimensions of a normal red blood cell (a) and dimensions of the sphere (b) and disks, (c), (d) of equivalent volume used for modeling purposes.

Shape	Radius (μm)	Max thickness (μm)	Min thickness (μm)	Volume (μm^3)
(a) Biconcave	4	2	1	87
(b) Sphere	$a=2.75$	⋯	⋯	87
(c) Disk 1	$R_1=4.00$	$t_1=1.73$	$t_1=1.73$	87
(d) Disk 2	$R_2=5.26$	$t_2=1.00$	$t_2=1.00$	87

cases, it was suggested that the particular geometry of the suspended particles could account for this anisotropic effect.

In this context, the purpose of this investigation is to compare the modeling of an asymmetric particle of *given volume* by regarding it as a sphere, to the result obtained by closely approximating the particle shape. A red blood cell of mean corpuscular volume (MCV) equal to 87 μm^3, density $\rho_1=1.087\,\mathrm{g/cm^3}$, and compressibility $\kappa_1=0.355\times10^{-9}$ m^2/N, suspended in isotonic saline of density $\rho_0=1.004\,\mathrm{g/cm^3}$ and compressibility $\kappa_0=0.442\times10^{-9}\,\mathrm{m^2/N}$ (Roy, 1987) is used as a case study.

For the purpose of this analysis, the red blood cell is modeled as a sphere and as two disks of different aspect ratios, with dimensions shown in Table I. This investigation first compares spherical models obtained both by exact and by approximate methods, in order to determine the range of validity of the long wavelength and Born approximations. Drawing from these findings, a scattering model is derived for the disk, which is used to investigate the importance of scatterer shape and orientation.

II. SCATTERING BY A FLUID SPHERE

Scattering by a sphere is undoubtedly the most exhaustively treated case in the literature. In addition to being the simplest possible model, the sphere is one of the few shapes for which both exact and approximate solutions are available, hence permitting an extensive comparison of modeling techniques. The purpose of this section is to summarize those results and to investigate their likely relevance as models for particles of different shape and orientation.

A. The exact solution

An exact solution for the scattered field produced when a harmonic plane wave $p_i(\mathbf{r})=\mathcal{P}e^{i(k_0z-\omega t)}$,[1] impinges upon a fluid sphere was first obtained by Anderson (1950). At very large distances from the scatterer, this solution can be cast in the form

$$p_s(\mathbf{r})=\left\{\frac{\mathcal{P}e^{i(k_0r-\omega t)}}{r}\right\}\Phi, \tag{1}$$

where $\mathbf{r}=(r,\theta)$ and Φ is known as the angular distribution factor. For $k_0a\leqslant1$, Anderson's solution is found numerically to converge very quickly, and can hence be evaluated accurately by a small number of terms ($m\leqslant5$). Figure 3 shows a plot of the non-dimensional far field scattering cross section, $\sigma=|\Phi|^2/(\pi a^2)$, evaluated at 180° and 90° relative to the

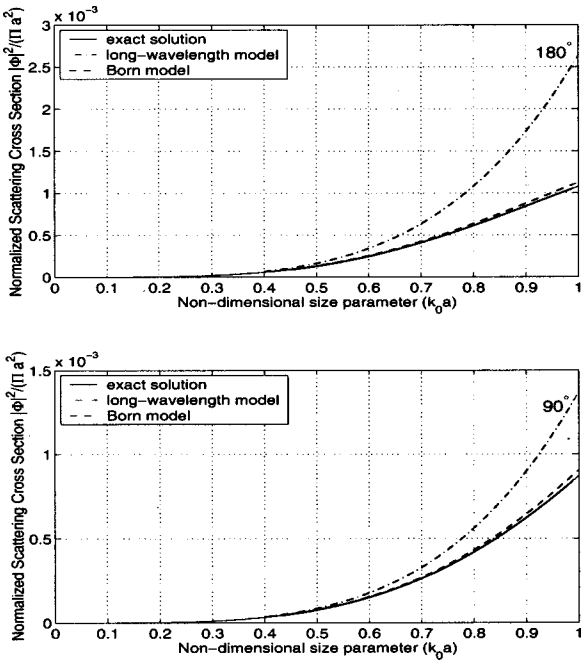


FIG. 3. Non-dimensional scattering cross section for a red blood cell modeled as a sphere, measured at 180° and 90° relative to the direction of the incident wave, as a function of the nondimensional size parameter k_0a (Roy, 1987).

direction of the incident wave, for a ''spherical'' red blood cell, at various values of the nondimensional size parameter k_0a.

B. The long wavelength approximation

The long wavelength solution for scattering by a sphere was historically derived much earlier than the exact result (Rayleigh, 1872). By assuming that the particle size is much smaller than the wavelength of the incident field, Rayleigh obtained the following simple expression for the angular distribution factor:

$$\Phi = \frac{1}{3}k_0^2 a^3 \left[\frac{\kappa_1 - \kappa_0}{\kappa_0} + \frac{3(\rho_1 - \rho_0)}{2\rho_1 + \rho_0} \cos\theta \right]. \tag{2}$$

Equation (2) shows that, in the long wavelength limit, the scattered field is proportional to the particle volume as well as to the square of the frequency. The monopole term describes the *axisymmetric* anomalous ''pulsation'' of the particle, which is due to the contrast in compressibility between the sphere and the surrounding medium. On the other hand, the directionally dependent dipole term expresses the fact that the scatterer, having a different inertia from the surrounding fluid, will have a different acceleration *in the direction of the incident wave.*

For a weak scatterer, it is found that the monopole—the more efficient radiator—often dominates the dipole term, as illustrated in Fig. 4. Most of the scattered field is generated by the monopole term of Eq. (2), the strength of which is, at any given frequency, solely dependent on the contrast in compressibility and the particle volume. On the other hand, the contribution of the dipole, which by virtue of its directionally sensitive physical origin would be most likely to be affected by particle shape and orientation, is very small by

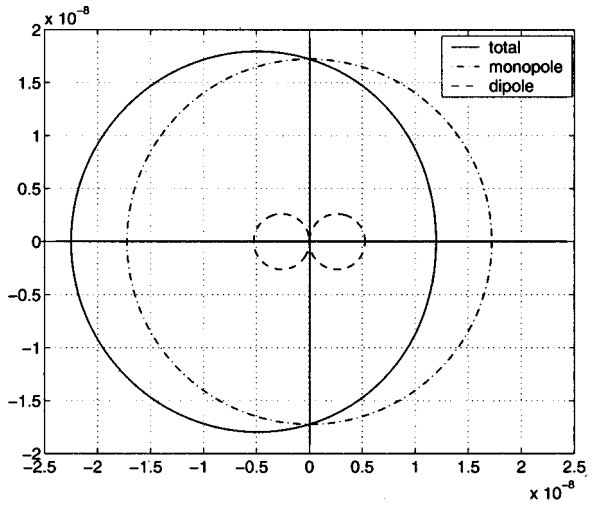


FIG. 4. Contribution of the monopole (-·-) and dipole (---) terms to the directivity pattern of the scattered field (—) produced by a *compact* red blood cell modeled as a fluid sphere for an incident plane wave coming in from the left.

comparison. This is the reason for which, in the long wavelength, a particle of arbitrary shape is accurately modeled by a sphere of equivalent volume.

The Rayleigh result for scattering by a sphere is plotted in Fig. 3. As expected from the assumptions involved, this model only mirrors the exact result accurately for $k_0a \leqslant 0.35$, beyond which the error exceeds 10%.

C. The Born approximation

The pressure field produced when a a harmonic wave $p_i(\mathbf{r},t) = p_i(\mathbf{r},\omega)e^{-i\omega t}$ impinges upon a fluid sphere may also be evaluated by viewing the scatterer as a collection of elementary point sources. The scattered field is then the integral of all contributions from the source elements distributed over the scattering volume V, giving (Morse and Ingard, 1968)

$$p_s(\mathbf{r},\omega) = \int\int\int_V [k_0^2 \gamma_\kappa(\mathbf{r}')p(\mathbf{r}',\omega)G(\mathbf{r},\mathbf{r}',\omega) + \gamma_\rho(\mathbf{r}')\nabla' p(\mathbf{r}_0)\cdot\nabla' G(\mathbf{r},\mathbf{r}',\omega)]dV', \tag{3}$$

where $G(\mathbf{r},\mathbf{r}',\omega)$ is the free-space Green's function and the terms

$$\gamma_\kappa(\mathbf{r}) = \frac{\kappa_e(\mathbf{r}) - \kappa_0}{\kappa_0}, \quad \gamma_\rho(\mathbf{r}) = \frac{\rho_e(\mathbf{r}) - \rho_0}{\rho_e(\mathbf{r})}, \tag{4}$$

respectively, represent the normalized contrasts in compressibility and density between the fluid element located at $\mathbf{r}$ and the surrounding medium.

The expression for the scattered field in Eq. (3) is not explicitly dependent on the incident wave p_i, but on the *unknown total field p* inside the scattering volume V. The Born approximation assumes that the scattered wave amplitude is much smaller than that of the incident wave, so that the total field within the scattering region can be approximated by the incident field: this will only be true for small γ_ρ and γ_κ (*weak scatterers*), and under conditions for which

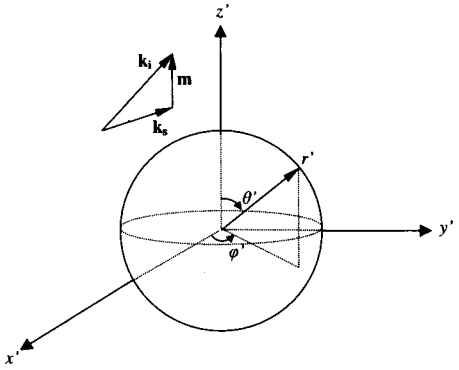


FIG. 5. System of coordinates for evaluating the field scattered by a sphere.

resonance and multiple scattering effects are not important. Using the Born approximation, the following expression is obtained for the angular distribution factor:

$$\Phi = \frac{k_0^2}{4\pi} \int \int \int_V [\gamma_\kappa(\mathbf{r}') + \gamma_\rho(\mathbf{r}') \cos\theta] e^{i\mathbf{m}\cdot\mathbf{r}'} dV', \quad (5)$$

where $\mathbf{k}_s = k_0 \mathbf{r}/r$ is the wave number vector in the direction of the observer, $\mathbf{m} = \mathbf{k}_i - \mathbf{k}_s$ and $|\mathbf{m}| = \mu = 2k_0 \sin(\theta/2)$. This is a general and well-known result (Morse and Ingard, 1968) giving the scattered field for a weak scatterer of arbitrary shape.

In the special case where the scattering volume V is a sphere of radius a and uniform mean density ρ_1 and compressibility κ_1, the functions γ_ρ and γ_κ are constants inside the scatterer. Based on the set of coordinates (r', θ', ϕ') defined in Fig. 5, Eq. (5) reduces to

$$\Phi = k_0^2 a^3 [\gamma_\kappa + \gamma_\rho \cos\theta] \frac{j_1(\mu a)}{\mu a}, \quad (6)$$

where j_1 denotes the first-order spherical Bessel function (Roy, 1987).

The nondimensional far-field scattering cross-section based on the Born approximation is plotted in Fig. 3 for observation angles of 180° (backscattering) and 90° relative to the incident field. Both the Rayleigh and Born models are found to be in close agreement with the exact solution for $k_0 a \leq 0.35$. This is not unexpected: in the long wavelength limit, $j_1(\mu a)/\mu a \approx 1/3$ and Eq. (6) reduces to

$$\Phi = \frac{k_0^2 a^3}{3} [\gamma_\kappa + \gamma_\rho \cos\theta]. \quad (7)$$

This result is found to be identical to the Rayleigh solution, given by Eq. (2), for $\kappa_1 \approx \kappa_0$ and $\rho_1 \approx \rho_0$. For a weak scatterer, the Born and Rayleigh models are in good agreement in the long wavelength limit.

However, unlike the Rayleigh model, the model derived using the Born approximation remains in close agreement with the exact theory up to wavelengths comparable to the size of the sphere: at $k_0 a = 1$ the Rayleigh model exceeds the exact solution by 60%, whilst the Born model overshoots by just 6% (Roy, 1987). It may seem surprising at first that the Born solution provides such an accurate representation for weak scatterers. However, the only limitation of the Born approximation lies in the assumption that the scattered field is much smaller than the incident field in the region of the scatterer. For a weak scatterer, this would only cease to be true due to multiple scattering effects, or if the particle itself were to resonate. Multiple particle scattering is not an issue when modeling scattering by a single particle. Resonance, on the other hand, need not be a wave mechanism, being possible even when the medium surrounding the particle is incompressible ($c_0 = \infty$): it is in fact the elasticity of the particle and the inertia of the exterior fluid that constitute the ''mass-spring'' oscillator (Dowling and Ffowcs Williams, 1983). The resonant frequency of a fluid sphere, ignoring surface tension effects, is well approximated by Minnaert's formula (Leighton, 1994; Crighton *et al.*, 1992)

$$\omega_M = \frac{c_1}{a} \sqrt{\frac{3\rho_1}{\rho_0}}. \quad (8)$$

Comparing this to the frequency ω_c, which corresponds to the nondimensional size parameter $k_0 a$ being equal to one, gives

$$\frac{\omega_M}{\omega_c} = \frac{c_1}{c_0} \sqrt{\frac{3\rho_1}{\rho_0}} \approx \sqrt{3} \quad (9)$$

for weak scatterers. The resonant frequency for a weak scatterer is therefore always located well above the frequency corresponding to $k_0 a = 1$. The Born approximation can be expected to yield theoretically accurate single particle weak-scattering models in the range $0 \leq k_0 a \leq 1$.

III. A MODEL FOR SCATTERING BY A FLUID DISK

To the author's best knowledge, no exact solution is available for the scattering of sound by a fluid disk *of nonzero thickness*. Detailed results are available for the disk of zero thickness (Bowman *et al.*, 1987), but those are of limited interest toward modeling the scattering of sound by a ''real'' particle: an infinitesimally thin particle will have no compressibility and therefore the contribution of the monopole term, which as shown in Sec. II is dominant for a weak scatterer, will be zero.

A model for the scattering of sound by a nonthin disk was first produced by Kou and Shung (1994), using the T-matrix method originally developed by Waterman (1969). They successfully demonstrated that the scattered field remains proportional to the square of the frequency up to $k_0 a = 1$, irrespective of particle shape, and that the angle of incidence of the incident field onto an asymmetric particle affects scattering significantly. However, their approach required the extensive calculation of Bessel and Hankel functions by recurrence and approximation formuli and results were presented solely for backscattering, whilst no simple expression was reached for the backscattering cross section.

The objective of the present study is to develop a simple, general, yet accurate model for the scattering of sound by a fluid disk, using the general result of Eq. (5) and the knowledge gained in Sec. II regarding the validity of the Born approximation. A disk of radius R, thickness t, and the system of coordinates defined in Fig. 6 are used for this purpose. Unlike the case of the sphere, the asymmetric shape of the disk implies that the z axis cannot be chosen parallel to the

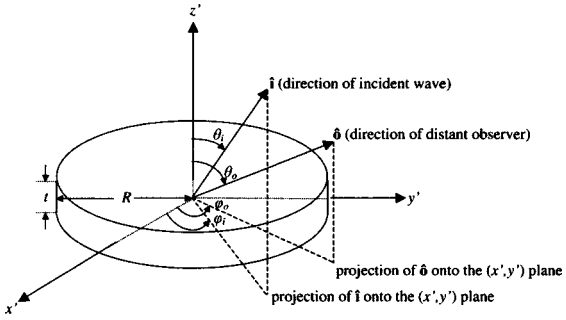


FIG. 6. System of coordinates for evaluating the field scattered by a disk of thickness t, lying in the (x',y') plane.

direction of the vector $\mathbf{m}=\mathbf{k}_i-\mathbf{k}_s$. The incident field impinges upon the disk in the arbitrary direction of the unit vector $\hat{\mathbf{i}}$, and the observer lies in the direction of the unit vector $\hat{\mathbf{o}}$. Expressing the vector $\mathbf{m}$ in the chosen coordinate system gives

$$\begin{aligned}&\mathbf{m}=m_1\hat{\mathbf{x}}+m_2\hat{\mathbf{y}}+m_3\hat{\mathbf{z}}=k_0(\hat{\mathbf{i}}-\hat{\mathbf{o}}),\\&m_1=k_0(\sin\theta_i\cos\phi_i-\sin\theta_o\cos\phi_o),\\&m_2=k_0(\sin\theta_i\sin\phi_i-\sin\theta_o\sin\phi_o),\\&m_3=k_0(\cos\theta_i-\cos\theta_o).\end{aligned}\tag{10}$$

and Eq. (5) becomes

$$\Phi=\frac{k_0^2}{4\pi}[\gamma_\kappa+\gamma_\rho\cos\theta]\int_{-t/2}^{t/2}\int_0^R\int_0^{2\pi}\rho'\times e^{i(m_1\rho'\cos\phi'+m_2\rho'\sin\phi'+m_3z')}d\phi'\,d\rho'\,dz',\tag{11}$$

where $\theta=(\hat{\mathbf{i}},\hat{\mathbf{o}})$, and ρ' is defined so that $x'=\rho'\cos\phi'$ and $y'=\rho'\sin\phi'$. The integral with respect to ϕ' in Eq. (11) can be evaluated by noting that $m_1=\sqrt{m_1^2+m_2^2}\cos\phi_1$ and $m_2=\sqrt{m_1^2+m_2^2}\sin\phi_1$, where ϕ_1 is the angle between the projection of $\mathbf{m}$ on the (x',y') plane and the x' axis. This gives

$$\begin{aligned}&\int_0^{2\pi}e^{i(m_1\rho'\cos\phi'+m_2\rho'\sin\phi')}d\phi'\\&=\int_0^{2\pi}e^{i\rho'\sqrt{m_1^2+m_2^2}(\cos\phi'\cos\phi_1+\sin\phi'\sin\phi_1)}d\phi'\\&=2\pi J_0(\sqrt{m_1^2+m_2^2}\rho')\end{aligned}\tag{12}$$

by definition (McLachlan, 1961), with J_n denoting the Bessel function of the first kind and order n. Finally, using the property $\int_0^a\alpha J_0(\alpha)d\alpha=aJ_1(\alpha)$ (McLachlan, 1961) gives

$$\Phi=k_0^2R^2\frac{t}{2}\frac{J_1(\sqrt{m_1^2+m_2^2}R)}{\sqrt{m_1^2+m_2^2}R}\frac{\sin(m_3t/2)}{m_3t/2}[\gamma_\kappa+\gamma_\rho\cos\theta]\tag{13}$$

with

$$\begin{aligned}&m_1^2+m_2^2=k_0^2[\sin^2\theta_i+\sin^2\theta_o-2\sin\theta_i\sin\theta_o\times\cos(\phi_i-\phi_o)],\\&m_3=k_0(\cos\theta_i-\cos\theta_o),\\&\cos\theta=\sin\theta_i\sin\theta_o\cos(\phi_i-\phi_o)+\cos\theta_i\cos\theta_o.\end{aligned}$$

This is essentially the result obtained by Ishimaru (1997) for the equivalent electromagnetic problem. It provides a full description of the scattered wave in the far field when a plane wave impinges upon a weak-scattering disk in an arbitrary direction.

In the long wavelength limit, only the first term of the Bessel function in Eq. (13) survives truncation, and $\sin(m_3t/2)/(m_3t/2)\to1$, giving

$$\Phi=\frac{k_0^2R^2t}{4}[\gamma_\kappa+\gamma_\rho\cos\theta],\tag{14}$$

or, in terms of the disk volume, $V=\pi R^2t$,

$$\Phi=\frac{k_0^2}{4\pi}V[\gamma_\kappa+\gamma_\rho\cos\theta].\tag{15}$$

Similarly, rewriting Eq. (7) in terms of the sphere volume, $V'=4\pi a^3/3$ yields

$$\Phi=\frac{k_0^2}{4\pi}V'[\gamma_\kappa+\gamma_\rho\cos\theta].\tag{16}$$

Direct comparison between Eqs. (15) and (16) demonstrates that a sphere and a disk of the same volume $V=V'$ will scatter identically at low frequencies.

As suggested by the observations in Sec. II.B, *particle shape is not important when modeling the scattering of sound by weak scatterers in the long-wavelength limit*; the strength of the scattered field depends solely on particle volume, the contrast in density and compressibility between the particle and the surrounding medium, and the frequency of the incident wave.

IV. THE SIGNIFICANCE OF PARTICLE ORIENTATION

It is not only particle shape, but also particle orientation that is indiscernible in the long wavelength limit. Comparison of Eqs. (13) and (14) demonstrates that, at low frequencies, the scattered field is no longer dependent upon the angle of incidence (θ_i,ϕ_i) of the incident field onto the disk; as for the spherical models, only the angular position θ of the observer *relative to the direction of the incident wave* is found to be significant.

This does not necessarily apply to higher frequencies. In order to establish the significance of particle orientation throughout the range $0<k_0R<1$, consider the two experimental situations represented schematically in Fig. 7. A primary source produces a plane wave, which impinges upon a disk-shaped particle broad-side (a) and thin-side (b). The resulting scattered field is observed at angles of 180° and 90° relative to the direction of the incident wave. In order to remain faithful to the assumptions underlying the Born approximation, it is further assumed that both the primary source and the observer are located far away from the particle, at a distance $L\gg R$.

Applying Eq. (13) to the case of the disk hit broad-side (a) and observed at 180° gives

$$\Phi=k_0^2R^2\frac{t}{4}\frac{\sin(k_0t)}{(k_0t)}[\gamma_\kappa-\gamma_\rho]\tag{17}$$

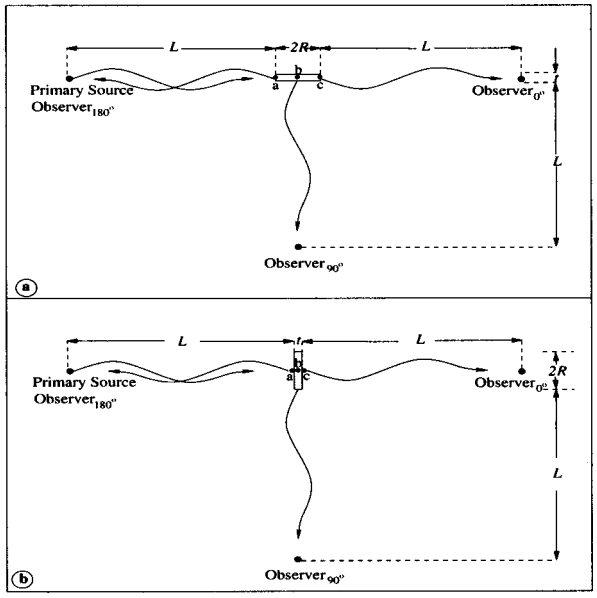


FIG. 7. Disk impinged upon broad-side (a) and thin-side (b).

with $\theta_i=0°$, $\theta_o=180°$, $\phi_i=\phi_o=90°$ in the system of coordinates defined in Fig. 6. For the observer at 90°,

$$\Phi=k_0^2R^2\frac{t}{2}\frac{J_1(k_0R)}{(k_0R)}\frac{\sin(k_0t/2)}{(k_0t/2)}[\gamma_\kappa] \tag{18}$$

with $\theta_i=0°$, $\theta_o=90°$, $\phi_i=\phi_o=90°$.

For the case of the disk hit thin-side (b) and observed at 180°, Eq. (13) gives

$$\Phi=k_0^2R^2\frac{t}{2}\frac{J_1(2k_0R)}{(2k_0R)}[\gamma_\kappa-\gamma_\rho] \tag{19}$$

with $\theta_i=270°$, $\theta_o=90°$, $\phi_i=\phi_o=270°$. Similarly, for the observer at 90°,

$$\Phi=k_0^2R^2\frac{t}{2}\frac{J_1(k_0R)}{(k_0R)}\frac{\sin(k_0t/2)}{(k_0t/2)}[\gamma_\kappa] \tag{20}$$

with $\theta_i=270°$, $\theta_o=0°$, $\phi_i=\phi_o=270°$.

The corresponding nondimensional scattering cross sections, for a red blood cell in saline modeled as a disk of radius R_1 and thickness t_1 (see Table I), measured at 180° and 90° relative to the direction of the incident wave, are plotted in Fig. 8 for the disk hit broad-side, and in Fig. 9 for the disk hit thin-side. Note that the scattering coefficient $\Phi^2/(\pi a^2)$ for the disk was arbitrarily put in nondimensional form using the maximal cross section of the sphere of equivalent volume; this is simply a normalization constant, and does not affect the relative magnitude of the results. Moreover, in order to permit comparison with the spherical models (also shown), the scattering cross section is now plotted as a function of frequency, since it is not possible to define a nondimensional size parameter for particles of different shapes. The frequency scale has been adjusted so that $k_0R_1\leqslant 1$.

Figures 8(a) and 9(a) show that disk orientation does not affect the scattered field for frequencies up to 20 MHz, beyond which the disk hit broad-side backscatters considerably more than the disk hit thin-side. Its backscattering cross section exceeds that of the disk hit thin-side by 27% at 30 MHz, 57% at 40 MHz, and as much as 100% at 50 MHz.

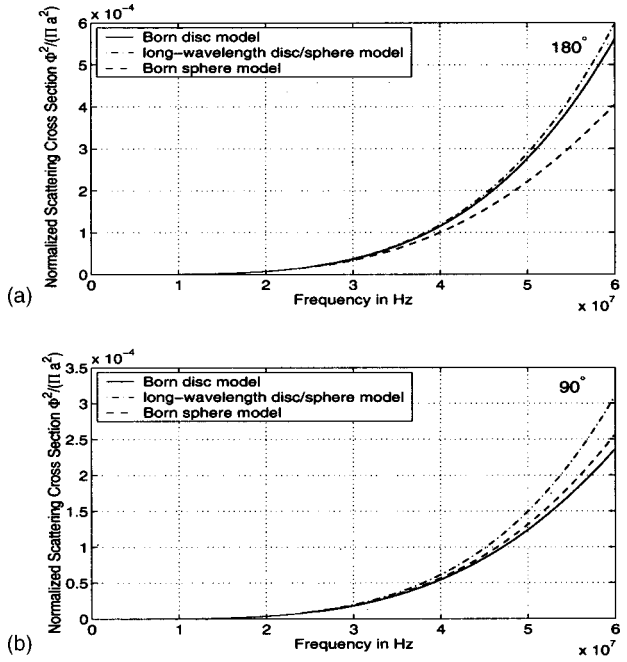


FIG. 8. Nondimensional scattering cross section for a red blood cell modeled as a disk hit *broad-side* and as a sphere, measured at (a) 180° and (b) 90° relative to the direction of the incident wave, as a function of frequency.

By contrast, direct comparison of Eqs. (18) and (20) and Figs. 8(b) and 9(b) indicates that the field observed at 90° relative to the direction of the incident wave is identical for the disk hit broad-side and for the disk hit thin-side throughout the frequency range. This is a direct consequence of the principle of reciprocity (Howe, 1998): the scattered field re-

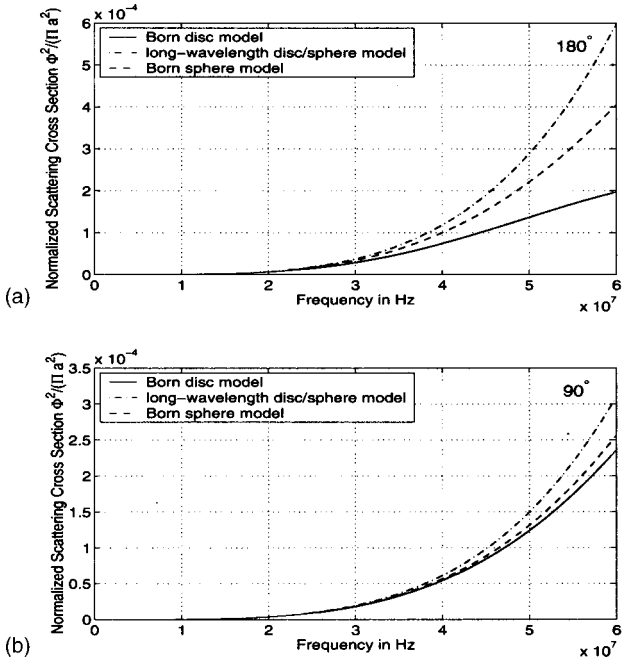


FIG. 9. Nondimensional scattering cross section for a red blood cell modeled as a disk hit *thin-side* and as a sphere, measured at (a) 180° and (b) 90° relative to the direction of the incident wave, as a function of frequency.

mains the same if the positions of the source and the observer are interchanged.

In physical terms, this observation may be explained by viewing the particle as a collection of elementary volumes, and the scattered wave as the superposition of the fields produced by each elementary source. Consider such an elementary volume of thickness t, located at an arbitrary position (x',y') in the coordinate system of the disk. In the case of the disk hit broad-side, every such volume will be impinged upon by the incident plane wave, produced by the primary source at time τ, at the same time instant $[\tau+L/c_0]$. Under the assumption that the observer is located very far from the particle, the scattered field produced by each elementary volume will require an additional time equal to $[\sqrt{R^2-x'^2}/c_1+(L+a-\sqrt{R^2-x'^2})/c_0]$ to reach the observer at 90°.

For the disk hit thin-side, an elementary volume located at (x',y') will be excited by the incident wave at time $[\tau+(L+a-\sqrt{R^2-x'^2}/c_0+\sqrt{R^2-x'^2}/c_1]$. The associated scattered wave will then reach the remote observer at time instant $[\tau+\sqrt{R^2-x'^2}/c_1+(L+a-\sqrt{R^2-x'^2})/c_0+L/c_0]$, i.e., at the same time as if the disk had been impinged upon broad-side. Thus, contributions from each elementary source combine at the 90° observer to form an identical scattered wave for both particle orientations.

Figures 8 and 9 also demonstrate that the long-wavelength and spherical models constitute good approximations of the disk model for frequencies up to 25 MHz, beyond which the observed discrepancies are considerable. The biggest error is induced for the disk hit thin-side and observed at 180° relative to the direction of the incident wave: the simplicity and symmetry of the spherical model clearly fail to mirror the implications of particle orientation at high frequencies.

V. THE SIGNIFICANCE OF PARTICLE SHAPE

The considerable effect that particle orientation has on the scattered field tends to suggest that an asymmetric particle, such as a red blood cell, should be modeled more as a disk of equivalent volume rather than as a sphere. However, biological scatterers seldom present themselves as perfect geometric shapes, and for modeling purposes one is left with the problem of choosing the aspect ratio of the simple shape that will best reflect the real behavior of the particle.

The red blood cell normally presents itself as a biconcave disk of average dimensions shown in Table I (Surgenor, 1975; Caro *et al.*, 1978). So far, this has been modeled as a flat disk of dimensions (R_1,t_1), disregarding the real thickness of the particle in favor of maintaining its physiological radius. The effect of approximating it instead as a disk of thickness t_2, which is equal to the average minimum thickness of the red cell, and of radius R_2, is now investigated. The directivity pattern of the normalized scattering cross section for both disks, impinged upon broad-side and thin-side, are plotted in Fig. 10 at three different frequencies.

The effect of scatterer shape and orientation only starts to become apparent at around 15 MHz. As one might expect, it is the field scattered by the disk of highest aspect ratio (disk 2) that is most strongly affected by particle orientation,

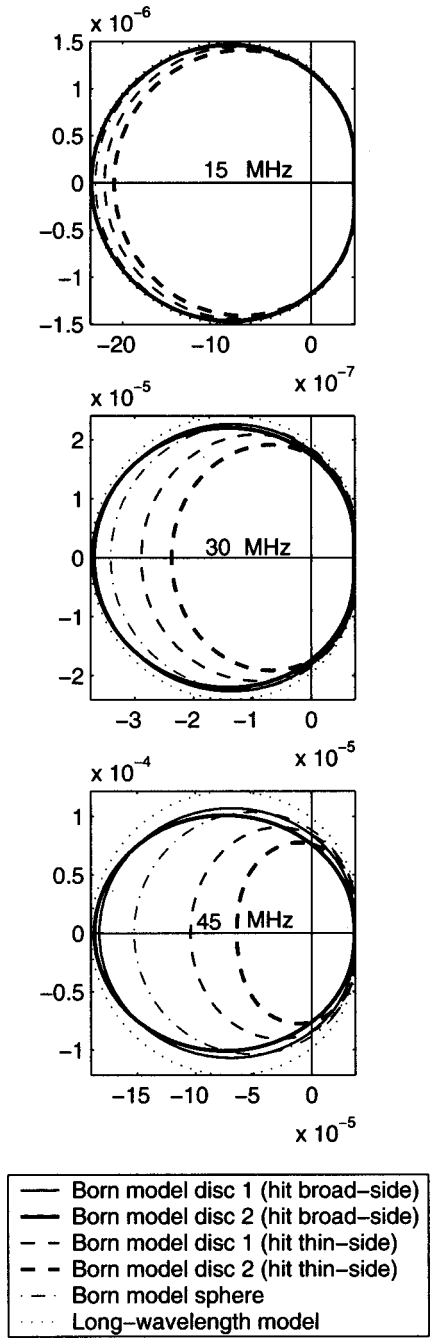


FIG. 10. Directivity pattern of the normalized scattering cross section $|\Phi|^2/(\pi a^2)$ for a red blood cell modeled as a Rayleigh scatterer (long wavelength), as a sphere, and as two disks of dimensions (R_1,t_1) and (R_2,t_2), hit broad-side and thin-side, at various frequencies $(0<k_0R<1)$. The particle, located at the origin, is impinged upon by an incident wave coming in from the left.

but only for the case of the disk hit thin-side. The field scattered by the disk hit broad-side is almost identical for both aspect ratios, even for very high frequencies; in this particular orientation the variation in thickness is hardly ''seen'' by the incident plane wave, because although $k_0R\approx1$, still $k_0t\ll1$.

Interestingly, the Born spherical model yields values of the scattering cross section that always lie in between those corresponding to the disk hit broad-side and the disk hit thin-side for backwards scattering. In the case of disk 1, the

spherical model provides a good approximation to the field averaged over the two particle orientations: this is consistent with the observations made by Roos and Twersky (Roos, 1983; Roos *et al.*, 1988; Twersky, 1964) described in Sec. I. The same cannot be said to apply to disk 2. In fact, it would appear that the higher the aspect ratio of the asymmetric particle, the more inaccurate the spherical model is in describing its scattered field.

Finally, it is also interesting to note that all models predict the same scattered field in the direction of the incident wave throughout the frequency range. Once more, this observation can be explained by viewing the particle as a collection of elementary volumes and by glancing at Fig. 7. The sooner an elementary volume is excited by the incident wave, the longer its scattered wave takes to reach the observer at 0°. Thus, independently of the shape of the particle, the field scattered by every elementary volume will reach the 0° observer at the very same instant as the incident field, and the total scattered field there will be identical for particles of equal volume.

For this reason, it is only measurements of scattering between angles of 90° and 180° relative to the direction of the incident wave that make it possible to observe the effects of particle shape and orientation. Such effects become increasingly detectable at high frequencies and the closer the observation angle is to 180° relative to the incident wave. More importantly, the choice of aspect ratio for modeling an asymmetric particle is critical, since this will determine the expected contrast in scattering cross section between different particle orientations. Changes in the larger of the two dimensions (the radius) will tend to affect scattering by the disk hit thin-side significantly, whilst modifications of the smaller dimension (the thickness) will have little effect on the disk hit broad-side. For the case of the red blood cell, disk 1 is a better model than disk 2.

VI. IMPLICATIONS FOR ULTRASONIC SCATTERING: THE INVERSE SCATTERING PROBLEM

In Secs. IV and V, it was demonstrated that particle shape and orientation have a significant effect on the directivity pattern of the scattered field. The question that naturally arises is whether particle orientation and particle shape can be deduced directly from measurements of the scattering cross section, and whether the observed effects are sufficiently significant to enable the distinction between biological particles of different shapes and sizes, assuming that their density and compressibility are known.

For reasons of cost and simplicity, it is desirable for the experimental arrangement used in this investigation to comprise no more than two ultrasonic transducers. This implies that the inverse scattering problem must be solved based on measurements from two scattering angles only, with one of the transducers acting both as an emitter and as a receiver. The choice of transducer frequency and of the angle between the two transducers is therefore critical in order to optimize the value of the information obtained at each observation point.

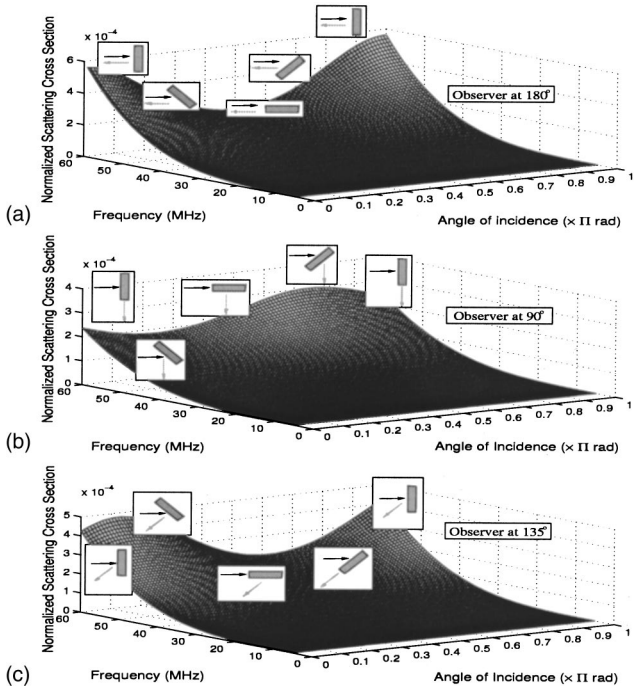


FIG. 11. Variation of the normalized scattering cross-section observed at (a) 180°, (b) 90°, and (c) 135° relative to the incident field, for a red blood cell modeled as a disk of dimensions (R_1, t_1), as a function of frequency and of the angle of incidence of the incident field $\theta_i(\phi_i = \phi_o = 90°)$.

The determination of particle orientation is difficult with the use of two transducers alone, since the problem is clearly three-dimensional. However, this may be possible by using an experimental setup designed so that one of the angular coordinates in Eq. (13) is fixed, say $\phi = \phi_i = \phi_o = 90°$. Figure 11 shows the predicted effect that the orientation of a disk-shaped red blood cell will have on the scattering cross section measured at three different observation angles relative to the incident wave. It can be seen that the variation in scattering cross section due to particle orientation is insignificant below 20 MHz, but increases with frequency. More important, the strongest variation is observed at 180° relative to the incident wave, whilst at 90° the observed variation is clearly less significant: this is in good agreement with the results presented in Sec. IV and suggests that the effect of particle orientation may go unnoticed if transducers at right angles to each other are used. On the other hand, the use of a transducer at 135° relative to the direction of the incident wave obviously provides a stronger contrast, which suggests that an apparatus comprising transducers at 45° to each other is more likely to permit the determination of particle orientation.

For biological scatterers, the particle dimensions are expected to vary and generally do so according to a normal distribution. It will only be possible to determine the orientation of a red blood cell if the variation in normalizing cross section due to its orientation is considerably more significant than that due to fluctuations in its size. There is notable disagreement in the haematology literature as to the exact values for the average dimensions and standard deviation of the radius and thickness of a red blood cell in a healthy adult;

these discrepencies stem mainly from the different conditions under which such measurements are obtained, which include cells aggregated into rouleaux, single flat cells in wet preparations, cells ''on edge'' in hanging drop preparations (Surgenor, 1975) or cells in dry films (Dacie and Lewis, 1984). Nevertheless, the standard deviation appears to never exceed $\pm 5\%$ of the mean value both for the radius and the thickness, whichever the measurement conditions. This is in good agreement with published values for the average mean cell volume (MCV) and standard deviation ($87\ \mu\text{m}^3 \pm 12\%$), which are considerably more reliable, since they are standardly measured using a Coulter counter (Rowan, 1983) and widely used as a diagnostic tool.

In order to assess the importance of cell size variation, the radius R_1 and thickness t_1 of the disk-shaped red blood cell are assumed to exhibit a variation corresponding to no more than $\pm 5\%$ of their mean values. Evaluation of the backscattering cross section based on the minimal and maximal dimensions of the cell shows that the variation in scattering cross section due to cell size fluctuations is of the order of $\pm 30\%$ about the mean value at 20 MHz, and decreases to approximately $\pm 20\%$ at 55 MHz. By contrast, the maximum and minimum values of the backscattering cross section due to particle orientation lie at $\pm 5\%$ of the mean value at 20 MHz, but increase to $\pm 40\%$ of the mean at 55 MHz. In fact, the variation in scattering cross section due to cell orientation begins to supersede the variation due to size fluctuation at about 47 MHz, but only becomes notably more significant beyond 55 MHz.

Due to the inherent size variability of biological particles, it therefore appears that the determination of red cell orientation requires the use of extremely high transducer frequencies. Perhaps of greater practical interest is the possibility of determining particle shape from measurements of the scattering cross section and by assuming that particle orientation is known *a priori*. In the case of red blood cells, it is possible to achieve this condition by submitting them to a shear flow, as discussed in Sec. I and illustrated in Fig. 2. This would cause disk-shaped particles to lie parallel to the plane of shear, whilst cells of different shape—spherical or other—would be left randomly oriented. Under these conditions, Fig. 10 shows that the directivity pattern of the scattering cross section makes it possible to distinguish between disk-shaped and spherical particles for frequencies superior to 20 MHz. The maximum contrast is observed if the disk is impinged upon thin-side; consequently, the emitting transducer should be placed parallel to the plane of shear. The optimal placement of the second transducer appears to be at right angles to the first, making use of the observation that the 90° scattering cross section varies little with particle shape or orientation. In this manner, the ratio between the scattering cross sections at 180° and 90° relative to the direction of the incident wave provides a measure of the distortion of the directivity pattern that is related to particle shape; this ratio is approximately equal to 1.5 for the sphere at 35 MHz, but is as high as 1.8 for the disc hit thin-side at the same frequency.

By devising an experimental setup that provides prior knowledge of particle orientation, it is therefore possible to deduce particle shape by measuring its scattering cross section at 180° and 90° relative to the direction of the incident wave. Assuming that the density and compressibility of the particle are also known, the value of the backscattering cross section can be used to estimate particle volume, whilst the ratio between the two measured scattering cross sections will indicate whether the cells are abnormally shaped (spherocytes, hemolyzed cells, etc.). Subject to experimental confirmation, the significance of particle shape and orientation may therefore form the basis of a novel detection technique.

VII. CONCLUSIONS

The Rayleigh solution fails to provide an accurate description of weak scattering by a fluid sphere for values of the nondimensional size parameter $k_0 a \geqslant 0.35$. On the other hand, application of the Born approximation produces a scattering model that agrees perfectly with the Rayleigh solution in the long-wavelength limit, yet closely matches the exact solution, even up to wavelengths of the order of the particle size. It can therefore be inferred that use of the Born approximation will yield simple, yet accurate, single-particle weak-scattering models.

Comparison between two such models for a disk and a sphere of equal volume indicates that an asymmetric particle can be safely modeled as a sphere for values of $k_0 a \leqslant 0.3$, beyond which particle shape and orientation truly become significant. In fact, the larger the aspect ratio of the particle, the more inaccurate is the description provided by the spherical model. For a red blood cell of volume $87\ \mu\text{m}^3$ modeled as a disk of dimensions $R_1 = 8\ \mu\text{m}$ and thickness $t_1 = 1.73\ \mu\text{m}$, the backscattering cross section for the disk impinged upon broad-side exceeds that of the disk hit thin-side by 27% at 30 MHz, 57% at 40 MHz, and as much as 100% at 50 MHz. However, this effect goes unnoticed if the scattered field is observed at 90° to the incident wave, because at that angle the directionally dependent dipole term goes to zero; this is a direct consequence of the principle of reciprocity, which predicts that the observed scattered field remains the same if the position of the source and the observer are interchanged.

It can therefore be concluded that particle shape and orientation have a significant effect in the right experimental context, even for weak-scatterers. However, in the case of red blood cells, it is not expected that particle orientation can be determined from inverse scattering measurements, due to the inherent size variability of biological scatterers. On the other hand, it should be possible to identify cell shape from the ratio of the scattering cross section at 180° and 90° to the incident wave, since this provides a measure of the distortion of the directivity pattern that is related to particle shape and orientation. This may form the basis of a novel cell-shape detection technique, using frequencies as low as 30 MHz.

ACKNOWLEDGMENT

The author wishes to thank the Greek National Scholarship Foundation (I.K.Y.) for kindly funding this research.

[1]In using this notation, it is implied that the actual pressure is in fact the real part of this expression (Morse and Ingard, 1968).

Allard, L., Cloutier, G., and Durand, L. G. (**1996**). ''Effect of the insonification angle on the Doppler backscattered power under red cell aggregation conditions,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 211–219.

Anderson, V. C. (**1950**). ''Sound scattering from a fluid sphere,'' J. Acoust. Soc. Am. **22**, 426–431.

Bessis, M. (**1974**). *Corpuscles: Atlas of Red Blood Cell Shapes* (Springer, Berlin).

Bowman, J. J., Senior, T. B. A., and Uslenghi, P. L. E. (eds.) (**1987**). *Electromagnetic and Acoustic Scattering by Simple Shapes*, 2nd ed. (Hemisphere, New York).

Burke, J. E. (**1966a**). ''Long-wavelength scattering by hard spheroids,'' J. Acoust. Soc. Am. **40**, 325–330.

Burke, J. E. (**1966b**). ''Low-frequency scattering by soft spheroids,'' J. Acoust. Soc. Am. **39**, 826–831.

Caro, C. J., Pedley, T. J., Schroter, R. C., and Seed, W. A. (**1978**). *The Mechanics of the Circulation* (Oxford University Press, Oxford).

Crighton, D. G., Dowling, A. P., Ffowcs Williams, J. E., Heckl, M., and Leppington, F. G. (**1992**). *Modern Methods in Analytical Acoustics* (Springer, Berlin).

Dacie, J. V., and Lewis, S. M. (**1984**). *Practical Haematology*, 6th ed. (Churchill Livingstone, London).

Dowling, A. P., and Ffowcs Williams, J. E. (**1983**). *Sound and Sources of Sound* (Wiley, New York).

Howe, M. S. (**1998**). *Acoustics of Fluid-Structure Interactions*, 1st ed., Cambridge Monographs on Mechanics (Cambridge University Press, Cambridge).

Ishimaru, A. (**1997**). *Wave Propagation and Scattering in Random Media*, 2nd ed., IEEE/OUP Series on Electromagnetic Wave Theory, Vol. 1 (IEEE Press/Oxford University Press, New York).

Kou, I. Y., and Shung, K. K. (**1994**). ''High frequency ultrasonic backscatter from erythrocyte suspension,'' IEEE Trans. Biomed. Eng. **41**, 29–34.

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

McLachlan, N. W. (**1961**). *Bessel Functions for Engineers*, 2nd ed. (Oxford University Press, New York).

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw–Hill, New York).

Quin, Z., Durand, L. G., and Cloutier, G. (**1998**). ''Kinetics of the 'black hole' phenomenon in ultrasound backscattering measurements with red blood cell aggregation,'' Ultrasound Med. Biol. **24**, 245–256.

Rayleigh, J. W. S. (**1872**). ''Investigation of the disturbance produced by a spherical obstacle on the waves of sound,'' Proc. London Math. Soc. **4**, 253–283.

Roos, M. S. (**1983**). ''Microparticle and Cell Characterization Using Acoustic Scattering,'' Ph.D. thesis, Yale University.

Roos, M. S. (**1988**). ''A technique for the study of acoustic scattering from microparticles,'' J. Acoust. Soc. Am. **83**, 770–776.

Roos, M. S., Apfel, R. E., and Wardlaw, S. C. (**1988**). ''Application of 30-MHz acoustic scattering to the study of human red blood cells,'' J. Acoust. Soc. Am. **83**, 1639–1644.

Rowan, R. M. (**1983**). *Blood Cell Volume Analysis* (Clark, London).

Roy, R. A. (**1987**). ''Quantitative particle characterization by scattered ultrasound,'' Ph.D. thesis, Yale University.

Shung, K. K., and Thieme, G. A. (eds.). (**1993**). *Ultrasonic Scattering in Biological Tissues* (CRC Press, Boca Raton, FL).

Surgenor, D. M. (ed.). (**1975**). *The Red Blood Cell*, 2nd ed. (Academic, New York), Vol. 1.

Twersky, V. (**1964**). ''Acoustic bulk parameters of random value distributions of small scatterers,'' J. Acoust. Soc. Am. **36**, 1314–1329.

Twersky, V. (**1988**). ''Low-frequency scattering by mixtures of correlated nonspherical particles,'' J. Acoust. Soc. Am. **84**, 409–415.

Waterman, P. C. (**1969**). ''New formulation of acoustic scattering,'' J. Acoust. Soc. Am. **45**, 1417–1429.

Floquet wave ultrasonic method for determination of single ply moduli in multidirectional composites

L. Wang and S. I. Rokhlin[a)]
Nondestructive Evaluation Program, The Ohio State University, Edison Joining Technology Center, Columbus, Ohio 43221

(Received 5 September 2001; accepted for publication 11 June 2002)

An inversion method based on Floquet wave velocity in a periodic medium has been introduced to determine the single ply elastic moduli of a multi-ply composite. The stability of this algorithm is demonstrated by numerical simulation. The applicability of the plane wave approximation to the velocity measurement in the double-through-transmission self-reference method has been analyzed using a time-domain beam model. It shows that the finite width of the transmitter affects only the amplitudes of the signals and has almost no effect on the time delay. Using this method, the ply moduli for a multiply composite have been experimentally determined. While the paper focuses on elastic constant reconstruction from phase velocity measurements by the self-reference double-through-transmission method, the reconstruction methodology is also applicable to assessment of data collected by other methods. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1498854]

PACS numbers: 43.35.Cg, 43.35.Mr [YHB]

I. INTRODUCTION

Nondestructive determination of elastic constants of anisotropic materials using ultrasonic phase velocity measurements has been an active field of study for the last several decades.[1–15] The elastic constants are obtained by inverting the Christoffel equation using a set of phase or group velocity data in different propagation directions. Rokhlin and Wang[12] have shown that seven of the nine elastic constants can be found from measurements in two symmetry planes for an orthotropic material and the solution is independent of the initial guesses and only slightly dependent on data scatter. The other two elastic constants can be found from measurements in nonsymmetry planes. However, these methods are only suitable for measurements of properties of homogeneous materials. The fiber-reinforced composites have various layups of unidirectional plies (laminas) to meet strength and stiffness requirements along different loading directions. They are assembled in different periodic stacking sequences of unit composite cells formed by laminas with different fiber directions (for example, 0°/−45°/90°/45°) to form multilayered structures as shown in Fig. 1 (the structure is unbounded in the $x-y$ plane). Each ply (lamina) is formed by fibers embedded in epoxy and oriented in one direction. Laminas in the form of a fabric with 0/90 fiber orientation are also available. In practice the stiffness of the composites is usually measured mechanically using samples cut in different directions. However by this method it is difficult to measure the transverse (xy,yz) properties and it is destructive. The ultrasonic method is the method of choice for nondestructive measurement of elastic properties; however in the case of dispersion the designer will be interested in the static limits instead of the elastic constants measurement at ultrasonic frequency. Since a lamina is the basic unit of a multiply (multidirectional) composite, it is important to develop techniques for determination of the lamina moduli by measurements on the composite after manufacturing. Measuring lamina properties is important for design purposes and for providing first-hand information on lamina changes due to processing. Also if lamina moduli are known, the static moduli of the composite laminate can be easily calculated.[16]

Ultrasonic characterization of elastic properties of multiply composites has been a challenge to the nondestructive evaluation community. One method of measuring the elastic constants is to use low frequency ultrasonic waves. In this case, the nonhomogeneous composite may be considered as an effective homogeneous medium. Therefore one can use the same procedure developed for a homogeneous layer to reconstruct the effective elastic constants from the wave velocity measurements. Examples of this have been given by Rose *et al.*,[12] Hosten,[13,14] and Lobkis *et al.*[15] The effective elastic constants are assumed to be equal to the static constants and the lamina properties are determined from them using inversion of a static micromechanical model for multiply composites.[13] However due to structural dispersion the ultrasonically measured effective elastic constants of a multiply composite are frequency dependent[17,18] and therefore cannot always be precisely related to the static elastic constants. By contrast, the unidirectional ply elastic constants are frequency independent in a much wider frequency range and can be considered equal to the static values (graphite/fibers have diameter 5–10 μm and thus laminas can be considered nondispersive).

Chu and Rokhlin[19] have proposed a method to determine lamina elastic moduli for [0/90] metal matrix composites without the effective media approach. However they neglect the effect of wave interference within a single lamina in the laminate on the time-delay measurement. Such an assumption is valid only for composites with thick laminas or

a)Author to whom correspondence should be addressed. Electronic mail: rokhlin.2@osu.edu

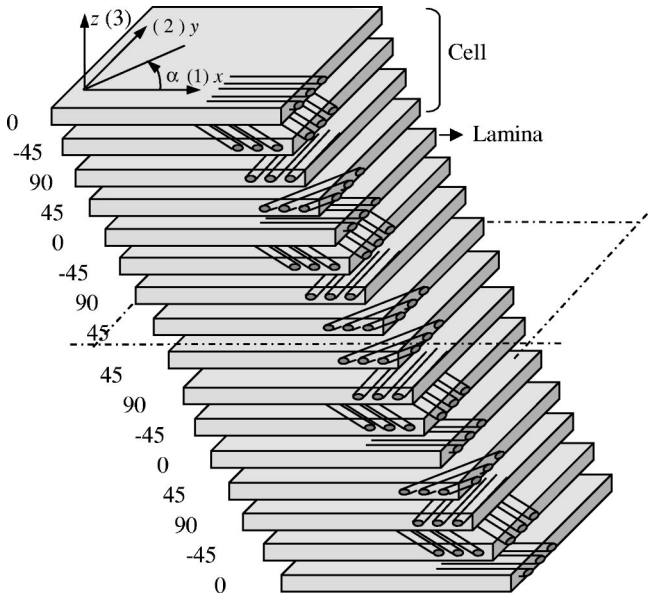


FIG. 1. Layout of $[0/-45/90/45]_{2s}$ multidirectional composites.

for composites with small impedance mismatch between 90° and 0° laminas. Therefore, the method cannot be used for graphite/epoxy multi-ply composites.

In this paper, a new Floquet-wave-based method of data processing in the self-reference double-through-transmission technique is developed for lamina property measurement of multi-ply composites. In Sec. II we introduce the problem and describe the application of the self-reference method for velocity measurement in multiply composites. In Sec. III we briefly review the way of computing the Floquet wave velocity. In Sec. IV we describe the Floquet-wave-based lamina moduli determination procedure and analyze the method's accuracy. Section V provides an experimental example. While the paper is focused on elastic constant reconstruction from Floquet wave phase velocity measurements by the self-reference double-through-transmission method, the reconstruction methodology is also applicable to assessment of data collected by other methods.

II. ULTRASONIC PHASE VELOCITY MEASUREMENT AND THE PROBLEM STATEMENT

A. Double-through-transmission measurements

To understand the difficulties with velocity measurements in multi-ply composites, let us first, as an example, briefly consider the ultrasonic double-through-transmission measurement method described by Rokhlin and co-workers.[9,11] The basic idea of this method is illustrated in Fig. 2. The sample is immersed in a fluid (water) and the ultrasonic pulse is transmitted through the sample, reflected from the plane reflector and returned to the transducer after the second through transmission. The pulse length in space is selected to be smaller than the sample thickness and thus there is no interference of the multiple reflections in the sample. Therefore the pulse propagation in the sample may be considered as that in an unbounded composite medium. First the time of flight t_n at normal incidence on the sample is measured. Next by continuously varying the incident

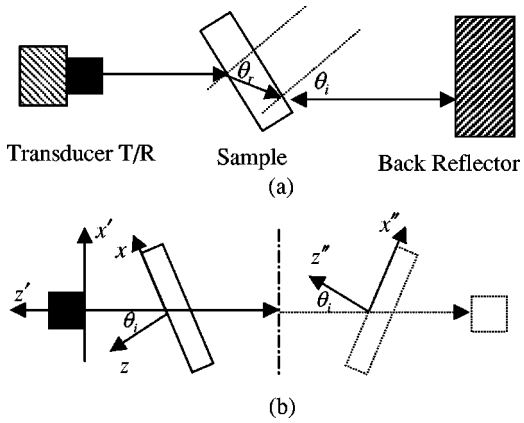


FIG. 2. (a) Schematic diagram of the self-reference bulk wave method. (b) Equivalent representation of the measurement used for modeling. The dashed lines indicate the mirror reflection of the back propagation path. Mirror plane is indicated by the broken vertical line.

angle, the time of flight t_{θ_i} versus incidence angle θ_i is measured. Due to the acoustic path length change in the sample, the time delay difference Δt_{θ_i} between t_{θ_i} and the reference time of flight t_n at normal incidence is obtained ($\Delta t_{\theta_i}=t_{\theta_i}-t_n$). Adopting the equation for Δt_{θ_i} derived in Ref. 11 to the self-reference double-through-transmission method[9] for arbitrary anisotropy from geometrical considerations, we get

$$\Delta t_{\theta_i}=\frac{2H}{1/V_0-1/V_n}-\frac{H\cos(\theta_r-\theta_i)}{V_0\cos\theta_r}+\frac{H}{V_p(\theta_r)\cos\theta_r},$$
$$\frac{\sin\theta_r}{V_p(\theta_r)}=\frac{\sin\theta_i}{V_0}, \tag{1}$$

where V_n is the velocity in the sample measured at normal incidence, and V_0 is the sound speed in the immersion fluid, H is the thickness of the sample, $V_p(\theta_r)$ is the phase velocity in the composite at angle θ_r, θ_r is the refraction angle. To improve precision in the Δt_{θ_i} determination for transverse wave one should account for phase correction.[20]

The phase velocity at refraction angle θ_r (corresponding to the incident angle θ_i shown in Fig. 2) is calculated for a generally anisotropic material using the phase velocity in the normal direction V_n and the time-delay change Δt_{θ_i} for the rotation angle θ_i of the sample:[9]

$$V_{\theta_i}(\theta_r)=\left(\frac{1}{V_n^2}+\frac{\Delta t_0-(\Delta t_0+\Delta t_{\theta_i})\cos\theta_i}{HV_0}+\frac{\Delta t_{\theta_i}(2\Delta t_0+\Delta t_{\theta_i})}{4H^2}\right)^{-1/2}, \tag{2}$$

with

$$\theta_r=\sin^{-1}\left(\frac{V_{\theta_i}\sin\theta_i}{V_0}\right)$$

and $\Delta t_0=2H(1/V_0-1/V_n)$. The phase velocity at normal incidence V_n is measured with high precision by overlapping multiple reflections from the front and back surface of the sample. Since all measurements are made relative to the ref-

erence acoustic path in the presence of the sample, the effect of geometric imperfection is small.[9]

The relation between the phase and group velocity determined from the time-of-flight measurement for the experimental configuration shown in Fig. 2 is addressed in Ref. 11. From the experimental phase or group velocities in different propagation directions, the elastic constants can be determined by inverting the Christoffel equation.[7–11] The method in its original form is applicable only to homogeneous media.

B. Application of Floquet waves

The multi-ply composites are extremely nonhomogeneous due to the strong anisotropy of the ply (lamina) and therefore there is significant wave velocity variability from ply to ply. The ply thickness is usually between 100–150 μm and is less than the pulse length, thus the ultrasonic signals interfere within the lamina and one cannot calculate the total time delay by simply adding the time delays for different plies; therefore the standard meaning of wave velocity is not applicable. The velocity and elastic property measurements in such materials require a conceptual step forward by accounting for the nonhomogeneity of the structure and the ply-wave interference.

Due to the periodicity of the composite structure the Floquet wave approach seems well suited for medium characterization. Floquet waves are formed in a periodic medium naturally accounting for the wave interference in the structure and their wave speed has a clear physical meaning. The Floquet wave theory provides a method to calculate the time delay for a signal propagating through a multi-ply composite and to determine the lamina elastic constants. The theoretical foundation of the method is briefly described in the following section followed by description of the elastic constant reconstruction and experimental examples in the subsequent parts.

III. FLOQUET WAVE DISPERSION AND VELOCITY CALCULATION

A. Floquet wave characteristic equation

The multilayered graphite epoxy composite forms a finite periodic structure with a periodic unit called the ''composite cell'' (Fig. 1). The cell is a layup from several unidirectional laminas and is thus a multilayered anisotropic unit.

A free wave propagating in a periodic medium is called a Floquet wave. To find the Floquet wave properties one applies the periodicity conditions[21,23]

$$\begin{bmatrix} \mathbf{u}_n \\ \boldsymbol{\sigma}_n \end{bmatrix} = e^{i\zeta h_c} \begin{bmatrix} \mathbf{u}_0 \\ \boldsymbol{\sigma}_0 \end{bmatrix}, \tag{3}$$

where ζ is the Floquet wave number component in the z direction, and h_c the thickness of the unit cell. Different methods[21–24] can be used to obtain the characteristic equation for computing the Floquet wave velocity in a periodic anisotropic medium. In our work, to describe the cell we introduce the layer stiffness matrix which relates stresses and displacements on both sides of the layer as defined by Eq. (A1) in the Appendix. The cell stiffness matrix $\mathbf{K}_c$ is obtained from the stiffness matrices of the layers forming the cell[25] using the recursive stiffness matrix algorithm [Eq. (A3)]. The Floquet wave characteristic equation in the form of stiffness matrix elements is obtained by combining the periodicity condition (3) with the cell stiffness matrix $\mathbf{K}_c$ (A4):[17,18]

$$(e^{i\zeta h_c}\mathbf{K}_c^{21} - e^{-i\zeta h_c}\mathbf{K}_c^{12} + \mathbf{K}_c^{22} - \mathbf{K}_c^{11})\mathbf{u}_0 = 0, \tag{4}$$

where $\mathbf{u}_0$ is the Floquet wave unit displacement vector defined at the top surface of the cell. The Floquet wave numbers ζ are found by searching for the roots of the determinant of the linear system (4) which has six solutions for $e^{i\zeta h_c}$. Due to the 2π periodicity of $e^{i\zeta h_c}$, the Floquet wave number ζ is not uniquely defined. In the wave number interval $-\pi < \zeta h_c < +\pi$, we can uniquely determine six Floquet wave numbers ζ, and the corresponding Floquet wave velocities $V_F = (\omega/\zeta)\cos\theta_F$, where θ_F is the angle between the Floquet wave propagation direction and the z axis (Fig. 1). As in homogeneous anisotropic media,[26] the Floquet wave propagation direction is determined from its energy flux vector.[27]

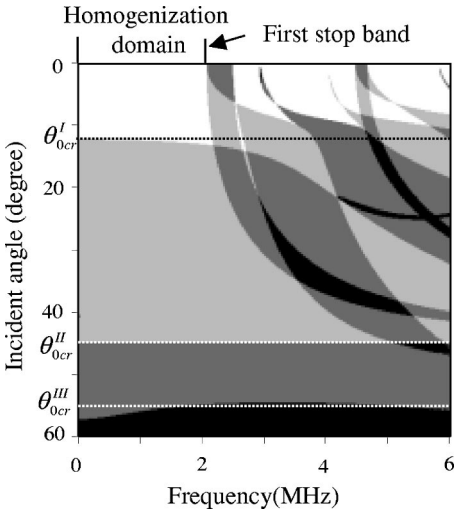


FIG. 3. Pass and stop bands of the three Floquet waves in k_x and ω domains for [0/90] periodic composite (lamina thickness is 0.194 mm). The propagating plane is oriented along 0° fiber direction (white, three propagating waves; gray, one or two propagating waves; black, stop band, no propagating waves permitted).

B. Stop and pass bands and measurement domain

The wave numbers obtained from Eq. (4) can be real and complex. The complex solutions correspond to stop bands and the real solutions to pass bands of the Floquet wave spectrum.[21] Figure 3 shows the stop and pass bands for the Floquet waves in the k_x and ω domains for a [0/90] composite with the incident plane oriented in the 0° direction. To relate this to ultrasonic experiments, we convert k_x to an equivalent angle of incidence onto the composite from water using $k_x = \omega \sin(\theta_i)/V_0$ (V_0 is the acoustic velocity in water 1.5 mm/μs). The gray level represents the number of propagating Floquet waves in the pass band. White indicates the area where all three Floquet waves have real wave numbers and black corresponds to the nonpropagating region (stop bands) for all three Floquet waves (they become nonhomogeneous with complex wave numbers). Two different levels

of gray between white and black indicate that one or two of the Floquet waves are propagating waves and have real wave numbers. At frequencies above the first stop band, shown by the vertical line in Fig. 3, the Floquet wave is very dispersive which leads to strong distortion of the transmitted signals; the transmitted signals even disappear within the stop bands.[28] This leads to difficulties in time delay measurements in composites at frequencies above the first stop band. At frequencies below the first stop band, the Floquet behavior is similar to a plane wave propagating in an anisotropic dispersive homogeneous medium. In this frequency domain, we may consider the periodic medium as an effective homogeneous anisotropic medium.[18] In Ref. 18, it is found that an approximate upper frequency f_h bound of the homogenization domain is given by

$$\frac{h_c}{\lambda_{\min}(\theta_i)}\cos\theta_i < 0.5 \quad \text{or} \quad f_h < \frac{0.5V_{\min}(\theta_i)}{h_c\cos\theta_i}, \tag{5}$$

where $\lambda_{\min}$ and $V_{\min}=\omega/k_{\max}$ are the minimum wavelength and velocity at a given incident propagation angle θ_i among all the bulk waves in each lamina of the cell. The estimated homogenization domain using Eq. (5) for a [0/90] composite is shown by the dashed lines in Fig. 3. Since the cell is about twice as thick for the multidirectional [0/45/90/−45] composite, the frequency range in Fig. 3 for this composite should be scaled down by a factor of 2.[18]

C. Symmetrical composite laminate

Multilayered composites usually are symmetrical as shown in Fig. 1 and thus are not periodic in the middle plane. The above discussion is valid for this case too. One can divide the symmetrical multilayered structure into two periodic media above and below the middle plane; for example, the $[0/-45/90/45]_{Ns}$ composite structure can be decomposed into two periodic structures: top $[0/-45/90/45]_N$ and bottom $[45/90/-45/0]_N$. Since the lamina is unidirectional and the z axis is one of its symmetry axes, the velocity of the three Floquet waves in the $+z$ direction equal those in the $-z$ direction. Thus the time of flight through the top part ($[0/-45/90/45]_N$ composite) is identical to that through the bottom part ($[45/90/-45/0]_N$ composite). Therefore, for our analysis we can consider the symmetrical composite as a periodic composite at frequencies below the first stop band.

IV. FLOQUET-WAVE-BASED LAMINA ELASTIC CONSTANTS DETERMINATION

A. Lamina elastic constants reconstruction

In this section, we propose a method for lamina properties reconstruction based on Floquet wave theory. The difference between this method and the methods used for homogeneous anisotropic media[9–11] is how the time delay calculation is done. For homogeneous media, the time delay is calculated from Eq. (1) using the phase velocity $V_p(\theta_r)$ determined from the Christoffel equation. To calculate the time delay for a periodic medium we use Eq. (1) replacing $V_p(\theta_r)$ by the Floquet wave phase velocity $V_F(\theta_r)$ obtained from Eq. (4). The coefficients of the Floquet wave characteristics [Eq. (4)] are obtained from the cell stiffness matrix $\mathbf{K}_c$ which is calculated from initial guesses of the lamina parameters for the given cell. For graphite/epoxy composites the elastic constants, density and thickness of all laminas can be assumed to be the same and thus the difference between each ply in the cell is the fiber orientation angle. The elastic constants of the lamina are assumed to be independent of frequency within the frequency range of the measurements. Thus the Floquet velocity depends only on the lamina elastic constants which need to be obtained and the cell lamina layout which is assumed to be known. The lamina properties determination using the Floquet wave approach follows almost the same computational procedure as that for a unidirectional composite layer.[10,11] The time delays are measured for different incident angles in one or several orientations of the incident plane. Then the least square optimization using the Levenberg–Marquardt algorithm is applied to minimize the error (cost function) between experimental time delays and those calculated using the Floquet theory. The unknown lamina moduli of the composite are found by minimizing the sum of the squares of the deviations between the experimental and calculated time delays considering the elastic constants as variables in a multidimensional space

$$\min_{C_{ijklm}\in R^n} \frac{1}{2}\sum_{i=1}^{m} (\Delta t_{\theta_i}^{\exp}(\alpha_i) - \Delta t_{\theta_i}^{F}(\alpha_i, C_{ij}))^2, \tag{6}$$

where $\Delta t_{\theta_i}^{\exp}(\alpha_i)$ is the measured time delay at incident angle θ_i and orientation angle α_i. $\Delta t_{\theta_i}^{F}(\alpha_i, C_{ij})$ is the corresponding computed time delay; n is the number of independent parameters (lamina moduli) to be determined (the space dimension) and m is the number of measurements. The time delays $\Delta t_{\theta_i}^{F}(\alpha_i, C_{ij})$ are calculated from Eq. (1) using initial guesses for the lamina moduli to obtain the velocity $V_F(\theta_r)$ from the Floquet wave equation [Eq. (4)].

Once the lamina properties are obtained, one can compute the frequency-dependent effective elastic constants of the composite using the homogenization method described in Refs. 17 and 18, which is applicable at frequencies below the first stop band. As zero-frequency limit one obtains the static

TABLE I. The results of reconstruction from simulated Floquet wave time-delay data in the incident plane $\alpha=0°$ for [0/45/90/−45] composite.

	C_{11} (GPa)	C_{22}	C_{12}	C_{23}	C_{55}
Original value	143.2	15.8	7.5	8.2	7
Initial guess	214.8	12.8	3.75	4.1	10.5
No scatter	0%	0%	0%	0%	0%
1% scatter	0.04%	0.01%	0.7%	0.03%	0.03%
5% scatter	0.5%	0.04%	2.27%	1.03%	3.05%

TABLE II. The results of reconstruction from Floquet wave time-delay data in the incident planes $\alpha=0°$ and 30° for [0/45/90/−45] composite.

	C_{11}	C_{22}	C_{12}	C_{23}	C_{55}
Original value	143.2	15.8	7.5	8.2	7
Initial guess	214.8	12.8	3.75	4.1	10.5
No scatter	0%	0%	0%	0%	0%
1% scatter	0.01%	0.01%	0.69%	0.02%	0.02%
5% scatter	0.41%	0.02%	1.2%	0.90%	2.1%

TABLE III. The results of reconstruction from Floquet wave time-delay data in the incident plane $\alpha=0°$ for [0/90] composite.

	C_{11}	C_{22}	C_{12}	C_{23}	C_{55}
Original value	143.2	15.8	7.5	8.2	7
Initial guess	214.8	12.8	3.75	4.1	10.5
No scatter	0%	0%	0%	0%	0%
1% scatter	1.49%	0.01%	0.37%	0.20%	0.0%
5% scatter	0.39%	0.02%	0.9%	0.30%	0.32%

effective elastic constants of the composite. These static elastic constants are useful for static stress analysis in composite structural design.

B. Accuracy of reconstruction and effect of data scatter: Numerical simulation

To check the stability and accuracy of the reconstruction, we performed computer simulations on a set of synthetic velocity data with random scatter, similar to the method described in Refs. 9–11. The results are given in Tables I and II for composite [0/45/90/−45] and Tables III, IV, and V for composite [0/90]. First for the set of elastic constants of the lamina (given in the first row of these tables), the time delay was calculated using Eqs. (1) and (4) for selected orientations of the incident plane and a discrete set of refraction angles at 1 MHz frequency for composite [0/45/90/−45] and 2 MHz for composite [0/90]. Next, random scatter was added to the time-delay data with data scatter levels 1% and 5%. The sets of time delays obtained serve as synthetic sets of experimental data. The initial guesses of the lamina elastic constants selected for reconstruction are given in the second row of the tables. These values are at least 50% different from the exact values used for synthetic sets determination (given in the first row). Using this data with scatter, the lamina elastic constants are obtained using the reconstruction procedure discussed above. The relative error between the reconstructed elastic constants and the exact elastic constants are given in the third row (0% data scatter level), fourth row (1% data scatter level), and fifth row (5% data scatter level).

Tables I and II give the results for [0/45/90/−45] composite using different orientation angles. Table I shows the results for plane of incidence at 0° orientation. The initial guess is 50% different from the exact values given in the first row. We assume that the lamina is transverse isotropic and one can see that all five elastic constants of the lamina are recovered. This is interesting since for a homogeneous anisotropic layer, one symmetry plane measurement is not sufficient to recover all elastic constants.[9–11] For a multidirectional composite, because the cell already includes laminas with different orientations, it becomes possible to determine all elastic constants from one plane of measurement only.

TABLE IV. The results of reconstruction from Floquet wave time-delay data in the incident planes $\alpha=30°$ for [0/90] composite.

	C_{11}	C_{22}	C_{12}	C_{23}	C_{55}
Original value	143.2	15.8	7.5	8.2	7
Initial guess	214.8	12.8	3.75	4.1	10.5
No scatter	0%	0%	0%	0%	0%
1% scatter	0.05%	0.01%	0.3%	0.03%	0.0%
5% scatter	0.11%	0.02%	0.42%	0.05%	0.00%

TABLE V. The results of reconstruction from Floquet wave time-delay data in the incident planes $\alpha=0°$ and 45° for [0/90] composite.

	C_{11}	C_{22}	C_{12}	C_{23}	C_{55}
Original value	143.2	15.8	7.5	8.2	7
Initial guess	214.8	12.8	3.75	4.1	10.5
No scatter	0%	0%	0%	0%	0%
1% scatter	0.04%	0.01%	0.07%	0.03%	0.0%
5% scatter	0.01%	0.02%	0.27%	0.03%	0.05%

The results for reconstruction from velocity data in two orientation angles (0°, 30°) are shown in Table II. Comparing with Table I, one can see the reconstruction is slightly less sensitive to scatter using data from two orientation angles of the incident plane. From these tables, one can see C_{12} is much more sensitive to noise than other elastic constants. This conclusion also agrees with those of Ref. 11 for a unidirectional composite.

Tables III, IV, and V give the results for [0/90] composites using data from different orientations of the incident plane. One can see the results are similar. To illustrate the quality of the inversion, Figure 4 shows the time delays obtained with 5% data scatter for [0/90] composites at 30° incident plane orientation (solid lines are the time delay computed with the lamina properties obtained and circles are the given set of data with scatter). All three Floquet waves are excited at this incident plane orientation.

C. Time domain beam model

The time delay calculated by Eq. (1) is based on the plane wave assumption. However experimentally, the transmitter and receiver have finite aperture and the incident wave has to be considered as a beam. Here in two-dimensional approximation we will investigate the effect of finite width of the transmitter/receiver on the time delay measurements using exact wave propagation theory[25] through the multi-ply composite layer in fluid.

The schematic diagram of the double-through-transmission method is given in Fig. 2. We consider the back reflector as a mirror with Fig. 2(b) showing the wave travel

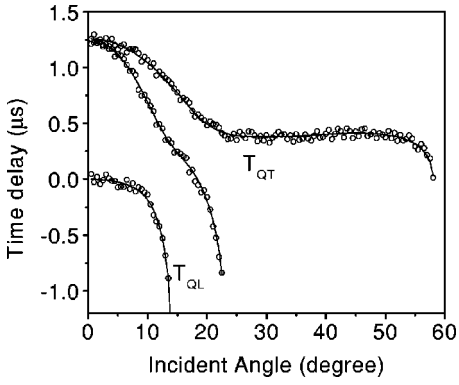


FIG. 4. Comparison of the simulated experimental time delays with 5% noise (circles) and that calculated based on reconstructed set of data (lines) for [0/90] composite at incident plane orientation $\alpha=30°$.

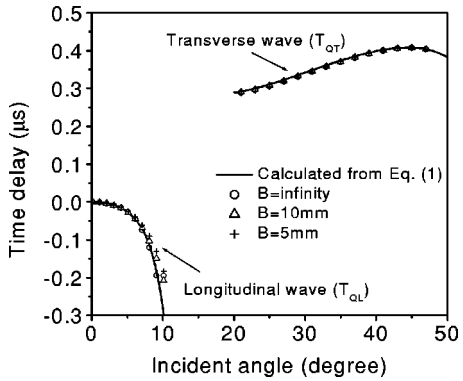


FIG. 5. Simulated experimental time delay for $[0/90]_{4s}$ composites at 2 MHz using Eq. (1) (solid lines) and time domain plane wave (B is infinite) and beam models with different beam width (open circle, $B=$infinity; triangle, $B=10$ mm; cross, $B=5$ mm).

path used in the model. Accounting for the phase delay in fluid as shown in Fig. 2(b), one can obtain the time-domain voltage output

$$V_{\text{out}}(t)=\int_{-\infty}^{+\infty}F(\omega)d\omega\int_{-\infty}^{+\infty}\Phi_t(k_x')T_t(k_x)\times T_b(k_x'')e^{i(-2k_z'L+k_zd+k_z''d)}\,dk_x', \tag{7}$$

where L is the distance between the surfaces of the transducer and back reflector, d is the thickness of the sample. $k_z=\sqrt{(\omega/V_0)^2-k_x^2}$. k_x',k_z' and k_x'',k_z'' are the wave numbers in the coordinate systems rotated from the axis z [Fig. 2(b)] by θ_i (primed) and $2\theta_i$ (double primed), respectively (z' is the normal to the transducer face). $F(\omega)$ is the frequency response of the transmitter/receiver; $\Phi_t(k_x')$ is the angular response of the transducer. They are experimentally determined by inverting the measured reflection signal from a homogeneous semispace such as a block of aluminum. $T_t(k_x)$ and $T_b(k_x'')$ are the transmission coefficients through the layer for the wave incident from fluid on the layer top or bottom, respectively. For calculation we use Gaussian functions to approximate the angular responses $\Phi_t(k_x')$ of the transducers

$$\Phi_t(k_x')=\frac{B}{\sqrt{2\pi}}e^{(k_x'B)^2/2}, \tag{8}$$

where B is the beam width of the transmitter or receiver. For plane wave approximation, the transducer response can be represented by a delta function $\Phi_t(k_x')=\delta(k_x')$.

Figure 5 compares the simulated time delay measurement using different beam sizes for a $[0/90]_{4s}$ composite at 0° incident plane orientation. The solid lines present the results calculated directly from Eq. (1) using the Floquet wave velocity. The calculation includes the phase correction[20] for the transmission coefficient from a fluid to a semi-infinite periodic medium. Points for different beam width B (open circles, triangles, and crosses) are the time delays determined directly from the time domain signals computed using the beam model [Eq. (7)] and the exact wave theory for propagating through the layered composite sample. The incident pulse has center frequency 2 MHz. One can see that the effect of beam size is small and thus a plane wave approximation is applicable, except for angles close to the critical angle (10°). At incident angles close to the critical angle, due to sharp phase shift the ultrasonic beam may spread and the received signal is distorted. In practice one should avoid phase velocity measurement when observing strong distortion of the ultrasonic pulse near the critical angle.

TABLE VI. Reconstructed lamina elastic constants in GPa from the measured Floquet wave time delay.

	Lamina moduli				
Composite	C_{11}	C_{22}	C_{12}	C_{23}	C_{55}
[0/90] graphite/epoxy	121	12.1	4.2	8.6	9.3
[0/45/90/−45] graphite/PEEK	116	11.6	7.7	7.9	7.2

V. EXPERIMENTAL RESULTS AND DISCUSSION

Two composite samples, $[0/45/90/-45]_{12s}$ (graphite/PEEK composite) and $[0/90]_{4s}$ (graphite/epoxy composite), have been used to demonstrate the proposed method experimentally (PEEK is a thermoplastic matrix material). The thickness of the [0/90] sample is 2.5 mm and the lamina thickness is 0.156 mm. The thickness of the [0/45/90/−45] sample is 14.6 mm and lamina thickness is 0.152 mm. The density for both samples is 1.6 g/cm^3.

The measurement approach and experimental setup were described in Ref. 10. In this study, ultrasonic signals with center frequency 1 MHz and 2 MHz are used for measurements on [0/45/90/−45] and [0/90] composites, respectively. The time-delay difference Δt_{θ_i} for longitudinal (T_{QL}) and quasishear (T_{QT}) waves are measured as a function of the incident angle θ_i from water with 0.5° step. The incident plane is oriented at $\alpha=0°$ relative to the fiber direction in the top lamina (Fig. 1). The multiple reflections in the sample at normal incidence are used to measure the velocity V_n in Eq. (1) and to determine the elastic constant C_{33} ($C_{33}=C_{22}=\rho V_n^2$, ρ is the sample density). In this section, the lamina is considered transversely isotropic (five independent elastic constants), however this limitation is not essential and can be easily lifted and the lamina considered to be orthotropic. The four other elastic constants (C_{11},C_{12},C_{23},C_{55}) are determined using the Floquet-wave-based method described above based on the measured time delays Δt_{θ_i} in only one plane of incidence (for an orthotropic lamina measurements should be performed in two incident planes). The reconstructed lamina properties for [0/90] and [0/45/90/−45] composites are given in the first and second rows of Table VI, respectively. To show the data quality, we plot in Fig. 6 the measured time delay difference Δt_{θ_i} (open circles) and that calculated (solid lines) using the reconstructed elastic constants given in Table VI [the deviation between the curves indicates the value of the cost function given by Eq. (6)]. As shown in Table VI the measured lamina properties for these two composites have some differences. This is due to the differences in the matrices and thickness of the samples for these composites (epoxy is for [0/90] composite with thick-

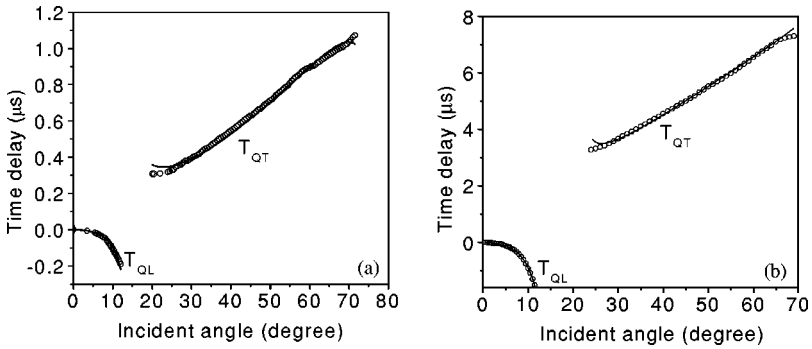


FIG. 6. Comparison of the experimental time delay (circles) and that calculated using the reconstructed set of data (lines) given in Table VI (incident plane orientation $\alpha=0°$). (a) [0/90] composite, (b) [0/45/90/−45] composite.

ness 2.5 mm and PEEK is for [0/45/90/−45] with thickness 14.6 mm); the sample thickness affects properties due to its effect on processing. The difference in moduli is consistent with static measurements indicating that the longitudinal and shear moduli of graphite/PEEK composites are slightly smaller than those of the graphite/epoxy composites.[16]

After the lamina properties have been obtained, the frequency-dependent effective elastic constants of the composite are calculated using the dynamic homogenization method.[17,18] At frequency 2 MHz, they are given for the [0/90] composite in the first row of Table VII. As an example, in Fig. 7, we plot the effective elastic constant C_{11} as a function of frequency for the [0/90] composite. As one can see the dynamic elastic constants at 2 MHz may be significantly different from those at 0.1 MHz (essentially static).

Because our measurements are performed at a frequency within the homogenization zone, the multidirectional composite can be replaced by an effective homogeneous medium[18] and from the measured time-delay data one can reconstruct the effective elastic constants of the composite by an inversion of the Christoffel equation.[9,11] The results obtained in this way are given in the second row in Table VII. The effective moduli are reasonably close to those shown in the first row which are computed from the lamina properties by the Floquet wave homogenization method (relative deviation about 5%). Ideally, in the homogenization domain both methods should provide identical results, thus discrepancies indicate homogenization error due to pulse bandwidth overlap with the first stop band. Also the homogenization-based method cannot recover C_{12} from 0° incident plane orientation measurements; to determine it will require experimental data in an additional incident plane.

As shown in Fig. 7, the effective elastic constants are frequency dependent. The static moduli can be obtained from lamina properties, performing the homogenization at very low frequency; the static effective moduli of the composite are given in the third row of Table VII. Significant differences exist between the static and dynamic effective elastic constants. Usually lamina moduli are obtained by static composite models from the ultrasonically measured composite effective moduli which are assumed to be equal to the composite static effective moduli. As one can see, in some cases using this approach one may introduce significant error because of the frequency dependency of the laminated composite effective moduli. The Floquet wave method allows one to obtain the lamina properties directly and thus to compute correctly the static elastic properties of multidirectional composites.

TABLE VII. Reconstructed effective homogeneous elastic moduli in GPa for [0/90] composite at 2 MHz and calculated from lamina static effective moduli.

[0/90] composite effective moduli				
C_{11}	C_{22}	C_{12}	C_{23}	C_{55}
Floquet wave (2 MHz)				
57.8	12.1	6.16	6.68	3.7
Plane wave (2 MHz)				
55.3	12.1		7.06	3.5
Floquet wave (static)				
66.4	12.1	4.59	6.35	3.0

VI. CONCLUSION

The Floquet wave theory has been applied to determination of the elastic constants of a single ply in multidirectional composites. Reconstruction of the lamina elastic constants from measured time-delay data has been done by nonlinear least-square minimization. For the transverse isotropic lamina, the method allowed us to measure all five lamina elastic constants from the velocity measurement in only one incident plane. The stability and accuracy of the reconstruction method has been evaluated by simulations on the synthetic data with random scatter. The results of the reconstruction were adequate, with deviation from the actual constants

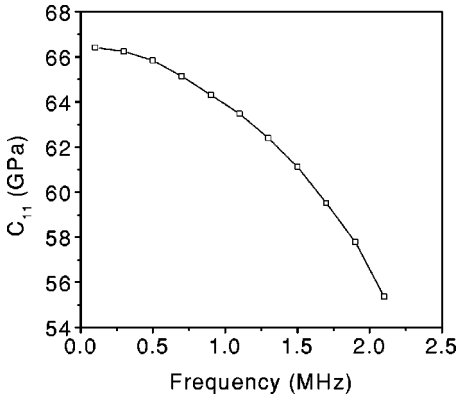


FIG. 7. Frequency dependent elastic constant C_{11} calculated from the measured lamina elastic constants using dynamic homogenization method (Refs. 17 and 18) for [0/90] composite.

set below 1%, even for 5% random scatter in the time-delay data. Applicability of the double-transmission self-reference velocity measurement method has been assessed using a two-dimensional-beam time-domain model. It shows that the finite width of the transmitter affects only the amplitude of the signal and has very little effect on the time delay. While the paper is focused on elastic constant reconstruction from phase velocity measurements by the self-reference double-through-transmission method, the reconstruction methodology is also applicable to assessment of data collected by other methods.

ACKNOWLEDGMENT

The authors are grateful to Bin Xie who assisted with obtaining experimental data for the [0/45/90/−45] composite.

APPENDIX: CALCULATION OF THE CELL STIFFNESS MATRIX

The stiffness matrix for mth layer is defined as

$$\begin{bmatrix} \boldsymbol{\sigma}_{m-1} \\ \boldsymbol{\sigma}_m \end{bmatrix}_m = \mathbf{K} \begin{bmatrix} \mathbf{u}_{m-1} \\ \mathbf{u}_m \end{bmatrix}_m, \tag{A1}$$

where $\boldsymbol{\sigma}_{m-1}$ and $\boldsymbol{\sigma}_m$ are the stresses on the top and bottom surfaces of the layer m; $\mathbf{u}_{m-1}$ and $\mathbf{u}_m$ are displacements and $\mathbf{K}_m$ is the 6×6 layer stiffness matrix represented as[17,25]

$$\mathbf{K}_m(6\times6) = \begin{bmatrix} \mathbf{K}_m^{11} & \mathbf{K}_m^{12} \\ \mathbf{K}_m^{21} & \mathbf{K}_m^{22} \end{bmatrix}. \tag{A2}$$

The cell stiffness matrix is obtained from the layer stiffness matrices $\mathbf{K}_m$ by the recursive algorithm[25]

$$\mathbf{K}_M = \begin{bmatrix} \mathbf{K}_{M-1}^{11} + \mathbf{K}_{M-1}^{12}(\mathbf{K}_m^{11} - \mathbf{K}_{M-1}^{22})^{-1}\mathbf{K}_M^{21} & -\mathbf{K}_{M-1}^{12}(\mathbf{K}_m^{11} - \mathbf{K}_{M-1}^{22})^{-1}\mathbf{K}_m^{12} \\ \mathbf{K}_m^{21}(\mathbf{K}_m^{11} - \mathbf{K}_{M-1}^{22})^{-1}\mathbf{K}_{M-1}^{21} & \mathbf{K}_m^{22} - \mathbf{K}_m^{21}(\mathbf{K}_m^{11} - \mathbf{K}_{M-1}^{22})^{-1}\mathbf{K}_m^{12} \end{bmatrix}, \tag{A3}$$

where $\mathbf{K}_M$ is the total stiffness matrix for the top m layers, $\mathbf{K}_{M-1}^{ij}$ is the total stiffness matrix for the top $m-1$ layers, $\mathbf{K}_m^{ij}$ are stiffness matrix elements for the mth layer.

For an n layered cell as shown in Fig. 1, using the recursive algorithm $n-1$ times, we obtain the cell stiffness matrix for the n-layered structure. We denote the cell stiffness matrix for the layered structure $\mathbf{K}_c$,

$$\begin{bmatrix} \boldsymbol{\sigma}_0 \\ \boldsymbol{\sigma}_n \end{bmatrix} = \mathbf{K}_c \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{K}_c^{11} & \mathbf{K}_c^{12} \\ \mathbf{K}_c^{21} & \mathbf{K}_c^{22} \end{bmatrix} \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_n \end{bmatrix}. \tag{A4}$$

[1] M. F. Markham, "Measurement of the elastic constants of fiber composites by ultrasonics," Composites **1**, 145–149 (1970).

[2] R. E. Smith, "Ultrasonic elastic constants of carbon fibers and their composites," J. Appl. Phys. **43**, 2555–2562 (1974).

[3] L. H. Pearson and W. J. Murri, "Measurement of ultrasonic wavespeeds in off-axis directions of composite materials," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1987), Vol. 6B, pp. 1093–1101.

[4] M. Grimsditch, "Effective elastic constants of superlattices," Phys. Rev. B **31**, 6818–6819 (1985).

[5] B. Hosten, M. Deschamps, and B. R. Tittman, "Inhomogeneous wave generation and propagation in lossy anisotropic solids. Application to the characterization of viscoelastic composite materials," J. Acoust. Soc. Am. **82**, 1769–1770 (1987).

[6] R. A. Kline, "Elastic properties of composites," in *Handbook of Elastic Properties of Solids, Liquid and Gases*, edited by M. Levy and L. Furr (Academic, San Diego, 2001), Vol. 2, Chap. 22, pp. 365–392.

[7] A. G. Every and W. Sachse, "Determination of the elastic constants of anisotropic solids from acoustic-wave group-velocity measurements," Phys. Rev. B **42**, 8196–8205 (1990).

[8] A. G. Every and W. Sachse, "Sensitivity of inversion algorithms for recovering elastic constants of anisotropic solids from longitudinal wavespeed data," Ultrasonics **30**, 43–48 (1992).

[9] Y. C. Chu and S. I. Rokhlin, "Comparative analysis of through-transmission ultrasonic bulk wave methods for phase velocity measurements in anisotropic materials," J. Acoust. Soc. Am. **95**, 3204–3212 (1994).

[10] Y. C. Chu, A. D. Degtyar, and S. I. Rokhlin, "On determination of orthotropic material moduli from ultrasonic velocity data in nonsymmetry planes," J. Acoust. Soc. Am. **95**, 3191–3203 (1994).

[11] S. I. Rokhlin and W. Wang, "Double through-transmission bulk wave method for ultrasonic phase velocity measurement and determination of elastic constants of composite materials," J. Acoust. Soc. Am. **91**, 3303–3312 (1992).

[12] W. R. Rose, S. I. Rokhlin, and L. Adler, "Evaluation of anisotropic properties of graphite-epoxy composite plate using Lamb waves," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1987), Vol. 6B, pp. 1111–1118.

[13] B. Hosten, "Stiffness matrix invariants to validate the characterization of composite materials with ultrasonic methods," Ultrasonics **30**, 365–371 (1992).

[14] B. Hosten, "Ultrasonic through-transmission method for measuring the complex stiffness moduli of composite materials," in *Handbook of Elastic Properties of Solids, Liquids and Gases*, edited by A. G. Every and W. Sachse (Academic, San Diego, 2001), Vol. 1, Chap. 3, pp. 67–86.

[15] O. I. Lobkis, D. E. Chementi, and H. Zhang, "In-plane elastic property characterization in composite plates," J. Acoust. Soc. Am. **107**, 1852–1858 (2000).

[16] I. M. Daniel and O. Ishai, *Engineering Mechanics of Composite Materials* (Oxford, New York, 1994).

[17] L. Wang and S. I. Rokhlin, "Analysis of ultrasonic wave propagation in multi-ply composites: Homogenization and effective anisotropic media," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 2000), Vol. 20B, pp. 1015–1023.

[18] L. Wang and S. I. Rokhlin, "Floquet wave homogenization of periodic anisotropic media," J. Acoust. Soc. Am. **112**, 38–45 (2002).

[19] Y. C. Chu and S. I. Rokhlin, "A method for determination of elastic constants of a unidirectional lamina from ultrasonic bulk velocity measurements on [0/90] cross-ply composites," J. Acoust. Soc. Am. **96**, 342–354 (1994).

[20] A. I. Lavrentyev and S. I. Rokhlin, "Phase correction for ultrasonic bulk wave measurements of elastic constants in anisotropic materials" in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 16B, pp. 1367–1374.

[21] A. M. Braga and G. Herrmann, "Floquet waves in anisotropic periodically layered composites," J. Acoust. Soc. Am. **91**, 1211–1227 (1992).

[22] A. H. Nayfeh, "The general problem of elastic wave propagation in multilayered anisotropic media," J. Acoust. Soc. Am. **89**, 1521–1533 (1991).

[23] A. H. Nayfeh, *Wave Propagation in Layered Anisotropic Media* (North-Holland, Amsterdam, 1995).
[24] C. Potel, J. Belleval, and Y. Gargouri, ''Floquet waves and classical plane-waves in an anisotropic periodically multilayered medium: Application to the validity domain of homogenization,'' J. Acoust. Soc. Am. **97**, 2815–2825 (1995).
[25] L. Wang and S. I. Rokhlin, ''Stable reformulation of transfer matrix method for wave propagation in layered anisotropic media,'' Ultrasonics **39**, 407–418 (2001).
[26] S. I. Rokhlin, T. K. Bolland, and L. Adler, ''Reflection and refraction of elastic waves on a plane interface between two generally anisotropic media,'' J. Acoust. Soc. Am. **79**, 906–918 (1986).
[27] C. Potel and D. Belleval, ''Energetic criterion for radiation of Floquet waves in infinite anisotropic periodically multilayered media,'' Acustica **87**, 340–351 (2001).
[28] S. I. Rokhlin, Q. Xie, Y. Liu, and L. Wang, ''Ultrasonic study of quasi-isotropic composites,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1999), Vol. 18B, pp. 1249–1256.

Theory of elastic wave propagation in anisotropic film on anisotropic substrate: TiN film on single-crystal Si

V. K. Tewary
National Institute of Standards Technology, Boulder, Colorado 80305

(Received 19 November 2001; revised 15 May 2002; accepted 14 June 2002)

The delta-function representation of the elastodynamic Green's function is used to derive an expression for the elastic wave forms on the surface of an anisotropic thin film on an anisotropic substrate due to a point or a line source located at the surface of the film. The dispersion relation for surface acoustic waves (SAWs) is obtained from the poles of the Green's function. A computationally efficient algorithm is formulated to obtain the elastic constants and the density of the film from the SAW dispersion data. The theory is used to analyze measured SAW dispersion relations in a titanium nitride film on silicon. The analysis yields values of the elastic constants and the density of the film. Excellent agreement is obtained between the theoretical and experimental dispersion results. Calculated wave forms for the surface wave due to a pulsed line source on the surface of the film are reported. [DOI: 10.1121/1.1499131]

PACS numbers: 43.35.Cg, 43.35.Pt, 43.20.Bi [SGK]

I. INTRODUCTION

We present a comprehensive formulation of the surface acoustic wave (SAW) propagation in anisotropic thin films on an anisotropic half-space substrate in terms of the elastodynamic Green's function. We calculate the elastodynamic surface Green's function for an anisotropic thin film coated on an anisotropic substrate. We apply the Green's function to derive an expression for the space–time-dependent displacement field or the elastic wave forms on the surface of the film due to a point or a line source on the surface of the film. From the poles of the Green's function in frequency space, we derive the dispersion relations for SAWs (surface acoustic waves). We also present a computationally efficient algorithm for determining material parameters such as the elastic constants and the density of the film using our representation of the Green's function and the observed dispersion relations for SAW in the film. We apply our formulation to calculate the displacement field due to a pulsed line source on the surface of a titanium nitride (TiN) film on single-crystal silicon (Si) and use our inversion algorithm to determine the elastic constants and the density of the film from the measured SAW dispersion curves.

Currently, propagation of SAWs in anisotropic thin films on anisotropic solids is of interest for elastic characterization of thin films in many industrial applications. SAW dispersion curves can be used for measurement of elastic constants of the film and other material parameters such as the density and the thickness. The characteristic features of SAW in thin films arise because the thickness of the film may be comparable to or less than the wavelength of the elastic waves. A mathematical model is needed for interpretation and inversion of SAW dispersion curves to obtain the material parameters of the solid. The advantage of a Green's function method is that it gives the wave forms as well as velocities of elastic waves. It also provides a general formula that can be applied to calculate the elastic wave forms due to any arbitrary but integrable force distribution.

Mourad *et al.*[1] have calculated the Green's function for SAWs in a thin film on a substrate using perturbation theory. Calculations of the elastodynamic Green's function for an anisotropic film of arbitrary thickness on an anisotropic substrate have not been reported in the literature. Our Green's function method is applicable to films of any thickness and anisotropy and anisotropic substrates. We also use our formulation of the Green's function to develop a computationally efficient algorithm for determining the material parameters of the film from the observed SAW dispersion curves.

For calculation of the elastodynamic Green's function, we use the delta function representation given earlier.[2] Various representations of the Green's function have been given by Wang and Achenbach,[3] Every and Kim,[4] and Ting.[5] The delta-function representation is convenient for anisotropic solids because the Green's function is expressed as a product of two delta functions. Hence, for a general 3D (three-dimensional) anisotropic infinite solid, numerical integration is required over only one variable. For bounded solids, the advantage of the product of two delta functions is lost but at least one integral can be done analytically. For time-dependent boundary-value problems, we use the Laplace transform of the Green's function to ensure causality. We obtain the SAW dispersion relations from the poles of the Green's function without any need for numerical integration. The SAW dispersion relations can of course be obtained directly from the secular equation without calculating the Green's functions.

The Green's function method has been applied to surface waves in homogeneous half-space tetragonal solids in Refs. 6 and 7. Many authors, such as, Barnett and Lothe,[8] Auld,[9] Nayfeh,[10] Li and Thomson,[11] Solie and Auld[12] have studied surface- and guided-wave propagation in anisotropic solids. Very interesting papers on anisotropic half-space solids have been published by Mourad and Deschamps[13] who have solved the Lamb's problem in these solids using the Cagniard de Hoop method and Every *et al.*[14] who have calculated their elastodynamic response to sudden surface loading. Ting has discussed elastic anisotropy in considerable

detail and described some useful models in a recent treatise.[5] Elastic wave propagation provides a useful technique for characterization of materials (see, for example, Kinra and Dalal[15] and the review article by Chimenti,[16] which gives other references).

Several features of wave propagation in layered solids have been studied by Barnett *et al.*[17] Its and Lee,[18] Pan and Datta,[19] Lih and Mal,[20] Niklasson, Datta, and Dunn,[21] and other authors. For a discussion of SAW in layered solids and many useful references, see, for example, Farnell and Adler,[22] and Nayfeh.[10] Computational methods for determination of elastic constants by using the measured wave propagation characteristics in homogeneous anisotropic solids have been described in many papers such as Refs. 23–25. Inversion of the measured SAW dispersion curves for determination of elastic constants of isotropic films has been addressed in Refs. 26–31 (see the quoted papers for other references).

In most of the inversion methods quoted above, one calculates the velocity of the elastic waves as function of the frequency and obtains the materials parameters of the film by a least squares fitting between the calculated and the observed dispersion curve. We have used a direct method that is computationally more efficient and more suitable for anisotropic solids. We substitute the observed dispersion relations for SAW in the expressions for the poles of the Green's function and solve the resulting set of nonlinear equations for the material parameters of the film.

We apply our theory and the inversion algorithm to interpret the experimental results of Hurley and Richards[32] on the dispersion of SAWs in a polycrystalline textured TiN film on single-crystal Si. Currently there is a strong interest in TiN coatings on Si because of the hardness properties of TiN. Excellent agreement is obtained between the theoretical and the experimental dispersion curves. However, the limited measurements reported in Ref. 32 do not yield a unique set of elastic constants. The estimated values of the elastic constants and the density of TiN are consistent with those available in the literature.

We have also calculated the wave forms on the surface of the film due to a pulsed line source. However, our main interest in this paper is in the calculation and measurement of SAW dispersion curves and estimating the elastic constants and the density of the film by comparing the theoretical and experimental values of SAW velocities. A direct comparison between theoretical and experimental wave forms is difficult because the wave forms depend upon the space–time behavior of the applied force, which is difficult to quantify experimentally. On the other hand, the wave velocities are independent of the nature or the geometry (line or point) of the applied force and depend only upon the material properties of the solid.

II. ELASTODYNAMIC GREEN'S FUNCTION FOR AN ANISOTROPIC LAYERED SOLID

We consider an anisotropic thin film on an anisotropic substrate. The film and the substrate are indicated as solids *A* and *B*, respectively, in Fig. 1 that also shows the Cartesian frame of reference. We assume that *A* and *B* are elastically

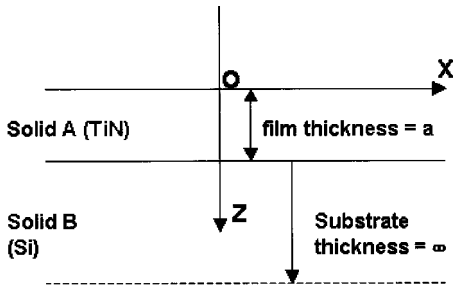


FIG. 1. Cartesian frame of reference. The point *O* on the surface is the origin of the coordinates. The *Y* axis is normal to the plane of the paper.

orthorhombic with their *c* axes parallel to the *Z* axis. The solid *B* extends to infinity in the $+Z$ direction. Both *A* and *B* extend to infinity in the $X-Y$ plane.

We denote the space variables by position vectors $\mathbf{x}$ and $\mathbf{x}'$. Their Cartesian components are denoted by indices 1, 2, and 3, corresponding to *X, Y,* and *Z* coordinates, respectively. We denote the time variable by *t* and its canonically conjugate variable by ω. For positive values, ω can be identified as a frequency. Summation over repeated indices is implied unless stated otherwise.

We calculate the response of the composite solid to a point force applied at the free surface of solid *A* at $\mathbf{x}=0$. We assume that there are no body forces in solids *A* or *B*. We first consider solid *A*. We seek the solution of the Christoffel equation

$$L_{ij}^{A}u_{j}^{A}(\mathbf{x},t)=0, \tag{1}$$

where $\mathbf{u}$ is the displacement field,

$$L_{ij}^{A}=c_{ikjl}^{A}\partial^{2}/\partial x_{k}\partial x_{l}-\rho^{A}\partial^{2}/\partial t^{2}; \tag{2}$$

the indices *i, j, k,* and *l* assume the values 1, 2, or 3 corresponding to the Cartesian coordinates *X, Y,* or *Z*, ρ^{A} is the density of *A*, and $\mathbf{c}^{A}$ is the fourth-rank elastic-constant tensor. The corresponding quantities for solid *B* will be identified by the superscript *B*. An equation similar to Eq. (1) is written for *B*.

It is convenient to introduce a 2D vector $\boldsymbol{\chi}$ in the vector subspace of $\mathbf{x}$ such that $\mathbf{x}$ can be written as $(\boldsymbol{\chi},x_3)$. Obviously $\chi_1=x_1$, $\chi_2=x_2$.

We specify boundary conditions (i) and (ii) for $t>0$ as given below.

(i) At $x_3=0$ (the top surface), $\mathbf{x}=(\boldsymbol{\chi},0)$

$$T_{i}^{A}(\boldsymbol{\chi},x_3=0,t)=\eta_{i}\delta(\boldsymbol{\chi})\delta(t), \tag{3}$$

where, for superscripts *A* or *B*, $\mathbf{T}$ is the traction vector in the *z* direction, defined by

$$T_{i}(\mathbf{x},t)=\tau_{i3}(\mathbf{x},t), \tag{4}$$

and τ is the stress tensor, given by

$$\tau_{im}(\mathbf{x},t)=c_{imjk}\partial u_{j}(\mathbf{x},t)/\partial x_{k}. \tag{5}$$

The 3D vector $\boldsymbol{\eta}$ denotes the strength of the load applied at the origin and δ denotes the Dirac delta function. The Dirac delta function for a vector argument is defined as a product of the delta function of its components, as given below

$$\delta(\boldsymbol{\chi})=\delta(\chi_1)\,\delta(\chi_2). \tag{6}$$

(ii) At $x_3=a$ (the interface), $\mathbf{x}=(\boldsymbol{\chi},a)$,

$$T_i^A(\boldsymbol{\chi},x_3=a,t)=T_i^B(\boldsymbol{\chi},x_3=a,t), \tag{7}$$

$$u_i^A(\boldsymbol{\chi},x_3=a,t)=u_i^B(\boldsymbol{\chi},x_3=a,t). \tag{8}$$

Equation (3) specifies a free surface except at the origin, and Eqs. (7) and (8) are the continuity conditions for a perfectly bonded interface. In addition, we require $\tau_{i3}^B(\mathbf{x})\to 0$ as $x_3\to\infty$. We also require the causality condition that $\mathbf{u}=0$ for $t<0$.

As shown in Appendix A, we can write a solution of Eq. (1) using the delta-function representation[2] of the Green's function in slowness space as follows:

$$\begin{aligned}\mathbf{u^A}(\mathbf{x},t)=&\int \mathbf{G_h^A}(\mathbf{q})\,\delta[\mathbf{q}\cdot\mathbf{x}-(t-t')]f_0(\boldsymbol{\xi},t')\,d\mathbf{q}\,dt'\\&+\int \mathbf{G_h^A}(q^*)\,\delta[\mathbf{q}^*\cdot\mathbf{x}+q_3a\\&-(t-t')]f_a(\boldsymbol{\xi},t')\,d\mathbf{q}^*\,dt',\end{aligned} \tag{9}$$

where $\mathbf{q}$ and $\mathbf{q}^*$ are the slowness vectors, $\mathbf{q}=(q_1,q_2,q_3)$, $\mathbf{q}^*=(q_1,q_2,-q_3)$, $\boldsymbol{\xi}=(\xi_1,\xi_2)$ is a 2D vector in the vector subspace of $\mathbf{q}$ such that $\xi_1=q_1$ and $\xi_2=q_2$, $d\mathbf{q}=dq_1\,dq_2\,dq_3$ and $\mathbf{f}_0$ and $\mathbf{f}_a$ are arbitrary but integrable functions of ξ and t. As in the case of the position vector, we can write a slowness vector in terms of $\boldsymbol{\xi}$ and q_3 as $\mathbf{q}=(\boldsymbol{\xi},q_3)$.

In analogy with the virtual force method for calculating Green's function in real space,[33] we will refer to $\mathbf{f_0}$ and $\mathbf{f_a}$ as the virtual forces in the q space. In view of the boundary conditions given by Eqs. (3)–(8), $\mathbf{f_0}$ and $\mathbf{f_a}$ are independent of q_3. The integration in Eq. (9) is to be performed over the entire space of $\mathbf{q}$ and t'. As shown in Appendix A, the actual range of q_i would depend upon the boundary conditions. In order to ensure causality, the displacement must be 0 for $t'>t$. Hence the integration over t' is between $t'=0$ and $t'=t$. For brevity the integration limits and separate integration symbols for the variables are not written in Eq. (9).

We will determine the virtual forces by imposing the boundary conditions. Taking the Laplace transform of Eq. (9), we obtain

$$\begin{aligned}\mathbf{u^A}(\mathbf{x},s)=&\int \mathbf{G_h^A}(\mathbf{q})\exp(-s\mathbf{q}\cdot\mathbf{x})\mathbf{f}_0(\boldsymbol{\xi},s)\,d\mathbf{q}+\int \mathbf{G_h^A}(q^*)\\&\times\exp[-s\{\mathbf{q}^*\cdot\mathbf{x}+q_3a\}]\mathbf{f}_a(\boldsymbol{\xi},s)\,d\mathbf{q}^*,\end{aligned} \tag{10}$$

where s is the Laplace variable with $\mathrm{Re}(s)>0$. From Eqs. (5) and (10),

$$\begin{aligned}\mathbf{T}^A(\mathbf{x},s)/s=&\int \mathbf{S^A}(\mathbf{q})\mathbf{G_h^A}(\mathbf{q})\exp(-s\mathbf{q}\cdot\mathbf{x})\mathbf{f}_0(\boldsymbol{\xi},z)\,d\mathbf{q}\\&+\int \mathbf{S^A}(\mathbf{q}^*)\mathbf{G_h^A}(\mathbf{q}^*)\\&\times\exp[-s\{\mathbf{q}^*\cdot\mathbf{x}+q_3a\}]\mathbf{f_a}(\boldsymbol{\xi},s)\,d\mathbf{q},\end{aligned} \tag{11}$$

where

$$[\mathbf{S^A}(\mathbf{q})]_{ij}=-c^A_{i3jk}q_k. \tag{12}$$

We now calculate the displacement and the traction fields at $x_3=0$ and at $x_3=a$. As described in Appendix B, we carry out the integration over q_3 and obtain the following from Eqs. (10) and (11):

$$\begin{aligned}\mathbf{u^A}(\boldsymbol{\chi},x_3=0,s)=&\int[\mathbf{g}^{A0}(\boldsymbol{\xi},s)\mathbf{f}_0(\boldsymbol{\xi},s)\\&+\boldsymbol{\gamma}^{Aa}(\boldsymbol{\xi},s)\mathbf{f}_a(\boldsymbol{\xi},s)]\exp(-s\boldsymbol{\xi}\cdot\boldsymbol{\chi})\,d\boldsymbol{\xi},\end{aligned} \tag{13}$$

$$\begin{aligned}\mathbf{u^A}(\boldsymbol{\chi},x_3=a,s)=&\int[\mathbf{g}^{Aa}(\boldsymbol{\xi},s)\mathbf{f}_0(\xi,s)\\&+\boldsymbol{\gamma}^{A0}(\boldsymbol{\xi},s)\mathbf{f}_a(\xi,s)]\exp(-s\boldsymbol{\xi}\cdot\boldsymbol{\chi})\,d\boldsymbol{\xi},\end{aligned} \tag{14}$$

$$\begin{aligned}\mathbf{T^A}(\boldsymbol{\chi},x_3=0,s)/s=&\int[\boldsymbol{\psi}^{A0}(\boldsymbol{\xi},s)\mathbf{f}_0(\boldsymbol{\xi},s)\\&+\boldsymbol{\sigma}^{Aa}(\boldsymbol{\xi},s)\mathbf{f}_a(\boldsymbol{\xi},s)]\exp(-s\boldsymbol{\xi}\cdot\boldsymbol{\chi})\,d\boldsymbol{\xi},\end{aligned} \tag{15}$$

$$\begin{aligned}\mathbf{T^A}(\boldsymbol{\chi},x_3=a,s)/s=&\int[\boldsymbol{\psi}^{Aa}(\boldsymbol{\xi},s)\mathbf{f}_0(\boldsymbol{\xi},s)\\&+\boldsymbol{\sigma}^{A0}(\boldsymbol{\xi},s)\mathbf{f}_a(\boldsymbol{\xi},s)]\exp(-s\boldsymbol{\xi}\cdot\boldsymbol{\chi})\,d\boldsymbol{\xi},\end{aligned} \tag{16}$$

where

$$\mathbf{g}^{A0}(\boldsymbol{\xi},s)=\sum_i \mathbf{M^A}(\boldsymbol{\xi},q_3=Q_i)W_i, \tag{17}$$

$$\mathbf{g}^{Aa}(\boldsymbol{\xi},s)=\sum_i \mathbf{M^A}(\boldsymbol{\xi},q_3=Q_i)\exp[-sQ_ia]W_i. \tag{18}$$

$$\boldsymbol{\gamma}^{A0}(\boldsymbol{\xi},s)=\sum_i \mathbf{M^A}(\boldsymbol{\xi},q_3=-Q_i)W_i, \tag{19}$$

$$\boldsymbol{\gamma}^{Aa}(\boldsymbol{\xi},s)=\sum_i \mathbf{M^A}(\boldsymbol{\xi},q_3=-Q_i)\exp[-sQ_ia]W_i, \tag{20}$$

$$\boldsymbol{\psi}^{A0}(\boldsymbol{\xi},s)=\sum_i \mathbf{S^A}(\boldsymbol{\xi},q_3=Q_i)\mathbf{M^A}(\xi,q_3=Q_i)W_i, \tag{21}$$

$$\boldsymbol{\psi}^{Aa}(\boldsymbol{\xi},s)=\sum_i \mathbf{S^A}(\xi,q_3=Q_i)\mathbf{M}(\boldsymbol{\xi},q_3=Q_i)\exp[-sQ_ia]W_i, \tag{22}$$

$$\boldsymbol{\sigma}^{A0}(\boldsymbol{\xi},s)=\sum_i \mathbf{S^A}(\boldsymbol{\xi},q_3=-Q_i)\mathbf{M}(\boldsymbol{\xi},q_3=-Q_i)W_i, \tag{23}$$

$$\begin{aligned}\boldsymbol{\sigma}^{Aa}(\boldsymbol{\xi},s)=&\sum_i \mathbf{S^A}(\boldsymbol{\xi},q_3=-Q_i)\mathbf{M}(\boldsymbol{\xi},q_3=-Q_i)\\&\times\exp[-sQ_ia]W_i,\end{aligned} \tag{24}$$

$$W_i = 1 \Big/ \left[2CQ_i \prod_{j \neq i} (Q_i^2 - Q_j^2) \right], \tag{25}$$

and C, Q_i, and $\mathbf{M}$ have been defined in Appendix C.

Now we write the displacement and the traction fields for solid B. Solid B has no body forces and no free surface. The solution for solid B needs to satisfy only the continuity conditions at the interface and, of course, the condition that the stress must vanish at infinity. Only one virtual force is needed for solid B. Proceeding as for solid A, we write the Laplace transforms of the displacement and the traction fields for solid B as follows:

$$\mathbf{u^B}(\mathbf{x},s) = \int \mathbf{G}_h^{\mathbf{B}}(\mathbf{q}) \exp[-s\{\mathbf{q}\cdot\mathbf{x} - q_3 a)\}] \mathbf{f}_b(\boldsymbol{\xi},s) d\mathbf{q}, \tag{26}$$

$$\mathbf{T}^B(\mathbf{x},s)/s = \int \mathbf{S^B}(\mathbf{q}) \mathbf{G_h^B}(\mathbf{q}) \exp[-s\{\mathbf{q}\cdot\mathbf{x} - q_3 a\}] \times \mathbf{f}_b(\boldsymbol{\xi},s) d\mathbf{q}, \tag{27}$$

where $\mathbf{f}_b(\boldsymbol{\xi},s)$ is the Laplace transform of the virtual force in solid B. The displacement and the traction fields in solid B at the interface are given by

$$\mathbf{u^B}(\boldsymbol{\chi}, x_3 = a, s) = \int \mathbf{g}^{B0}(\boldsymbol{\xi},s) \mathbf{f}_b(\boldsymbol{\xi},s) \exp(-s\boldsymbol{\xi}\cdot\boldsymbol{\chi}) d\boldsymbol{\xi} \tag{28}$$

and

$$\mathbf{T^B}(\boldsymbol{\chi}, x_3 = a, s)/s = \int \boldsymbol{\psi}^{B0}(\boldsymbol{\xi},s) \mathbf{f}_b(\boldsymbol{\xi},s) \exp(-s\boldsymbol{\xi}\cdot\boldsymbol{\chi}) d\boldsymbol{\xi}, \tag{29}$$

where $\mathbf{g}^{B0}$ and $\boldsymbol{\Psi}^{B0}$ are defined analogous to Eqs. (17) and (21).

As shown in Appendix A, the boundary condition at the top surface ($x_3 = 0$) is satisfied if

$$\mathbf{T^A}(\boldsymbol{\chi}, x_3 = 0, s) = s\boldsymbol{\eta} \int \exp(-s\boldsymbol{\xi}\cdot\boldsymbol{\chi}) d\boldsymbol{\xi}. \tag{30}$$

From Eqs. (7), (8), (14)–(16), and (28)–(30), we obtain the following algebraic equations for determination of the virtual forces:

$$\boldsymbol{\psi}^{A0}(\boldsymbol{\xi},s)\mathbf{f}_0(\boldsymbol{\xi},s) + \boldsymbol{\sigma}^{Aa}(\boldsymbol{\xi},s)\mathbf{f}_a(\boldsymbol{\xi},s) = \boldsymbol{\eta}, \tag{31}$$

$$\boldsymbol{\psi}^{Aa}(\boldsymbol{\xi},s)\mathbf{f}_0(\boldsymbol{\xi},s) + \boldsymbol{\sigma}^{A0}(\boldsymbol{\xi},s)\mathbf{f}_a(\boldsymbol{\xi},s) - \boldsymbol{\psi}^{B0}(\boldsymbol{\xi},s)\mathbf{f}_b(\boldsymbol{\xi},s) = 0, \tag{32}$$

$$\mathbf{g}^{Aa}(\boldsymbol{\xi},s)\mathbf{f}_0(\boldsymbol{\xi},s) + \boldsymbol{\gamma}^{A0}(\boldsymbol{\xi},s)\mathbf{f}_a(\boldsymbol{\xi},s) - \mathbf{g}^{B0}(\boldsymbol{\xi},s)\mathbf{f}_b(\boldsymbol{\xi},s) = 0. \tag{33}$$

The solution of the above equations in matrix form is given by

$$\mathbf{f} = \mathbf{V}^{-1}\mathbf{J}, \tag{34}$$

where $\mathbf{f}$ is a column vector containing the unknown virtual forces, $\mathbf{J}$ is a column vector containing the terms on the RHS (right-hand side) of Eqs. (31)–(33), and $\mathbf{V}$ is a square matrix consisting of the coefficients of the unknown forces on the LHS. After some algebraic manipulations, we obtain the following expressions for the virtual forces:

$$\mathbf{f}_0(\boldsymbol{\xi},s) = [\boldsymbol{\psi}^{A0} - \boldsymbol{\sigma}^{Aa}\mathbf{C}_1^{-1}\mathbf{C}_0]^{-1}\boldsymbol{\eta}, \tag{35}$$

$$\mathbf{f}_a(\boldsymbol{\xi},s) = [\boldsymbol{\sigma}^{Aa} - \boldsymbol{\psi}^{A0}\mathbf{C}_0^{-1}\mathbf{C}_1]^{-1}\boldsymbol{\eta}, \tag{36}$$

and

$$\mathbf{f}_b(\boldsymbol{\xi},s) = (\mathbf{g}^{B0})^{-1}[\mathbf{g}^{Aa}\mathbf{f}_0 + \boldsymbol{\gamma}^{A0}\mathbf{f}_a], \tag{37}$$

where

$$\mathbf{C}_0 = \boldsymbol{\psi}^{Aa} - \boldsymbol{\psi}^{B0}(\mathbf{g}^{B0})^{-1}\mathbf{g}^{Aa}, \tag{38}$$

$$\mathbf{C}_1 = \boldsymbol{\sigma}^{A0} - \boldsymbol{\psi}^{B0}(\mathbf{g}^{B0})^{-1}\boldsymbol{\gamma}^{A0}. \tag{39}$$

For brevity, we have not explicitly shown the functional dependence of the quantities on $\boldsymbol{\xi}$ and s on the RHS of above equations.

The inverse Laplace transform of Eq. (10) with the virtual forces given by Eq. (34) is the desired solution of Eq. (1). This solution gives the space–time dependence of the propagation of a delta-function pulse in solid A due to a pulsed point source located at the surface. Equation (13) gives the displacement field on the surface associated with the propagating surface wave. The coefficient of $\boldsymbol{\eta}$ in the solution gives the elastodynamic surface Green's function for the layered solid. The Green's function is useful for solving the Christoffel equation for any integrable spatial distribution of forces at the surface having any temporal dependence.

In Sec. IV we give the calculated wave forms for a pulsed line force on a TiN film on silicon. It is possible to measure the wave forms experimentally and compare with the theoretical result to obtain the elastic constants. However, it would require a precise knowledge of the space–time dependence of the applied force. Alternatively, the elastic constants and other material parameters of the film can be determined by comparing the calculated and measured dispersion of surface waves. This is described in the next section.

III. DISPERSION OF SURFACE WAVES AND DETERMINATION OF ELASTIC CONSTANTS

Experimentally, one can measure the SAW dispersion—the phase velocity of the surface waves as a function of frequency. These measurements can then be inverted to derive the material parameters such as the elastic constants, density, and the thickness of the film. In order to calculate the dispersion of the surface waves, we first express the solution given by Eq. (10) in frequency space. The causality condition is not required so the integration over t' in Eq. (9) extends from $-\infty$ to $+\infty$. The solution for surface waves in frequency space[7] is then given by Eq. (13) by replacing s by $\iota\omega$. This solution would correspond to an applied monochromatic line force that depends upon time as $\exp(\iota\omega t)$.

For a given ω, the integrand over $\boldsymbol{\xi}$ in Eq. (13) has poles at the zeros of the determinant of $\mathbf{V}$. The location of the poles in the complex ω space gives the allowed values of $\boldsymbol{\xi}$. The dispersion of the surface waves is, therefore, given by the solution of the following determinantal equation:

$$|\mathbf{V}(\boldsymbol{\xi},\omega)| = 0, \tag{40}$$

which gives the inverse velocity ξ of the wave as a function of the frequency ω.

Equation (40) is a $3n\times3n$ nonlinear determinantal equation where n is the dimensionality of the solid. For a 3D case, $n=3$ and for a 2D case, $n=2$. It may be computationally more convenient to solve the reduced $n\times n$ determinantal equation

$$|\mathbf{V}_R(\boldsymbol{\xi},\omega)|=0, \tag{41}$$

where

$$\mathbf{V}_R(\boldsymbol{\xi},\omega)=\boldsymbol{\psi}^{A0}-\boldsymbol{\sigma}^{Aa}\mathbf{C}_1^{-1}\mathbf{C}_0. \tag{42}$$

The usual procedure (see, for example, Ref. 29) of obtaining the elastic constants from the measured dispersion is to use a least-squares fit between the observed dispersion curves and theoretical curves obtained by Adler's method.[22] It involves a large number of forward calculations through the following steps: (i) Starting with some guess values of the elastic constants, calculate the velocity $v=\mathrm{Re}(1/\xi)$ as a function of ω by solving Eq. (40) for ξ for different values of ω; (ii) compare the theoretical and measured dispersion curve; and (iii) vary the elastic constants until the best fit is obtained between the two curves. One difficulty in this procedure lies in the identification of the right branch of the theoretical dispersion curves. Since the determinant V is $3n\times3n$, Eq. (40) has $3n$ branches of solutions. Only one of them will correspond to the experimental curve. The difficulty is further aggravated by the crossing over of the branches in the ξ-space in some cases.

In our method of inversion, we do not need to do the forward calculations. We use the measured values of v for specific values of ω directly in Eq. (40), and solve the resulting nonlinear equation for the material parameters. This method is computationally faster and more efficient because we effectively save computational steps (i) and (ii) described above, and avoid the problem of identification of a specific branch of the solution. If the experimental dispersion curve is not very smooth, it is useful to preprocess the data. First fit a smooth regression curve to the experimental data and then use the smoothed values in Eq. (40) for solving it for the elastic constants.

IV. ANALYSIS AND INVERSION OF MEASURED SAW DISPERSION FOR TiN ON Si AND CALCULATION OF WAVE FORMS

We now apply our theory to calculate the dispersion of surface waves in an anisotropic TiN film on a single-crystal Si substrate and obtain the elastic constants and density of TiN using the inversion method described above. Using these parameters, we calculate the wave forms on the free surface. For inversion, we use the experimental[32] SAW dispersion curve for a polycrystalline TiN film on a single-crystal Si substrate. The nominal film thickness was 306 nm, with an estimated uncertainty of $\pm10\%$. The substrate thickness, which we assume to be infinite in our model, was approximately 0.5 mm. The measured dispersion curve is shown by the triangles in Fig. 2. The measurements were made using an optical line source and correspond to SAWs propagating in the [110] direction in Si, which we have taken to be the X axis. The interface between the film and the substrate was the (0,0,1) plane of Si, which is normal to the Z axis in our frame of reference shown in Fig. 1.

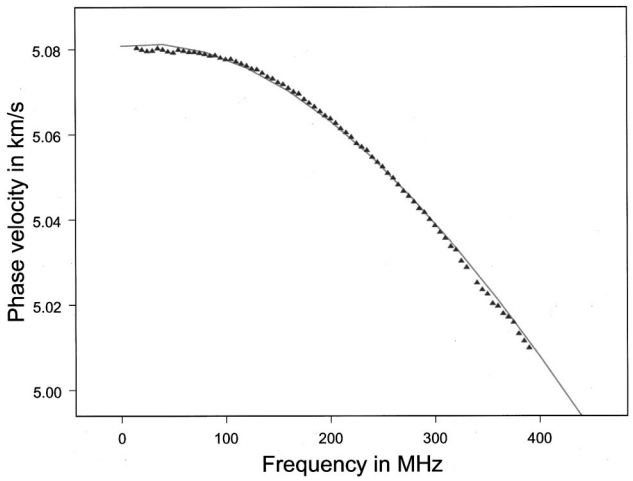


FIG. 2. Theoretical and experimental dispersion curves for titanium nitride film for surface waves propagating in the [110] direction of silicon. Solid line, theory; triangles, measured values (Ref. 32).

The preferred orientation (texture) of the polycrystalline film was [111] with respect to the crystallographic axes of the face-centered-cubic TiN.[34] This direction is normal to the (1,0,0) plane of Si and is taken to be the Z axis in our frame of reference in Fig. 1. Accordingly, we model the film as an orthorhombic solid with its c axis parallel to the Z axis in Fig. 1. Orthorhombic is the most general structure with rectangular unit cell geometry that has nine independent elastic constants. However, since our calculations and the measurements are restricted to only one direction, we can determine only four elastic constants, c_{11}, c_{13}, c_{33}, and c_{55}. In order to determine the remaining elastic constants and the symmetry of the film, measurements need to be made in more directions on the $X-Y$ plane. The Christoffel equation for orthorhombic solids is given in Appendix C.

The theory given in Sec. II is applicable to a general 3D problem of a point source. However, the experimental results are available only for a line source.[32] In this case the analysis is considerably simplified because the problem becomes 2D and Eq. (41) reduces to a 2×2 determinant. In order to model a line source in the Y direction, and waves propagating in the X direction, we set $q_2=0$ in our formulation. Then $\boldsymbol{\xi}$ and $\boldsymbol{\chi}$ become scalars and equal to q_1 and x_1, respectively. The Green's function and the Christoffel matrices are block diagonalized and reduce to 2×2 matrices. The virtual forces and the displacements are now 2D vectors. We use Eqs. (C2) and (C7) to calculate the matrices in Eqs. (17)–(25), (38), and (39). Finally, we calculate the 2×2 matrix $\mathbf{V_R}$ from Eq. (41). As is apparent from Eqs. (C1) and (C2), only four out of nine elastic constants, that is, c_{11}, c_{13}, c_{33}, and c_{55} contribute to the dispersion in the X direction. Hence only these elastic constants can be determined in the present case. In addition to the elastic constants, we also treat the density of the film as unknown and determine it by the same procedure.

The matrix elements of $\mathbf{V_R}$ are functions of ω and q_1. We use the experimental values of ω and q_1 (inverse velocity in the X direction) in Eq. (41). This gives us a nonlinear

TABLE I. Density and elastic constants of TiN films as reported in the literature (ρ, density in kg/m^3; c_{ij}^*, elastic constants in GPa of cubic TiN with respect to the crystallographic axes; c_{ij}, elastic constants transformed to the axes shown in Fig. 1 and after lateral averaging).

	Our values	Set I (Ref. 36)	Set II (Ref. 37)	Set III (Ref. 38)	Set IV (Ref. 39)
ρ	5408		5460	5390	5400
c_{11}^*		498	507	625	
c_{12}^*		106	96	165	
c_{44}^*		168	163	163	
c_{11}	455	470	465	558	473
c_{13}	148	125	124	210	119
c_{55}	159	187	191	208	182
c_{33}	446	461	450	536	469

equation in the elastic constants and the density. Each pair of values (ω,q_1) gives one equation. In order to determine the four elastic constants and the density, we need at least five equations, that is, five pairs of (ω,q_1) values. In practice, we choose 5–10 pairs of (ω,q_1) values and solve the corresponding set of 5–10 nonlinear equations by a minimization procedure. The solution takes a few seconds on a 1.7 GHz desktop computer. After determining the values of the elastic constants and the density, we substitute them in Eq. (41) and calculate q_1 for the entire range of ω.

The values of the four elastic constants and the density of the film as determined using the above procedure are given in Table I. Our values are with respect to the frame of reference in Fig. 1 in which the [110] crystallographic direction of silicon has been assumed to be the X axis. We used the following values of the density and the elastic constants (quoted in Ref. 35) of silicon (with respect to the crystallographic axes):

$$c_{11}^*=165.7 \text{ GPa}, \quad c_{12}^*=63.9 \text{ GPa},$$

$$c_{44}^*=79.6 \text{ GPa}, \quad \rho=2331 \text{ kg/m}^3.$$

The asterisks over c's denote that these values are with respect to the crystallographic axes of silicon and not the frame of reference of Fig. 1. The solid line in Fig. 2 shows the calculated dispersion curve. We see that the fit between our theoretical and experimental values is very good, the difference being less than 0.1 percent. The fact that we got such a good fit is not surprising as such, because of a large number (5) of adjustable parameters.

For the purpose of comparison, we quote some published values of elastic constants of cubic TiN, denoted by c_{ij}^* in Table I. The star denotes that these values are with respect to the crystallographic axes of cubic TiN. The elastic constants reported in Set IV are with respect to the same axes as in Fig. 1. A direct comparison between our values and those in sets I, II, and III of Table I is not possible because we do not know the orientation in the XY plane of the grains of the TiN film used in experiments. We only know that the texture of the film is in the [111] direction with respect to its cubic crystallographic axes. In order to represent the elastic constants of sets I, II, and III in our frame of reference, we make a plausible assumption that all the grains are oriented in their crystallographic [111] direction which is taken to be the Z axis in our frame of reference. We also assume that the grains are randomly oriented in the XY plane. This would make the film laterally isotropic on the XY plane. Accordingly, we transform the elastic constants of sets I, II, and III, from the crystallographic axes to the frame of reference in Fig. 1 and average over all orientations in the XY plane. The resulting sets of values, denoted by c_{ij}, are also shown in Table I.

We see from Table I that our values of c_{ij} are generally consistent with those obtained by other authors. The differences may be attributed to processing conditions of the film such as residual stresses, nonuniform thickness, and the density of the film that can vary with the nitrogen content of the film. An important source of error in our estimated values of the elastic constants is the uncertainty in the value of the thickness of the film. The results are very sensitive to the thickness of the film and any error in the measured value of the thickness will affect the estimated values of the elastic constants. It is possible to treat the thickness of the film also as an adjustable parameter and determine its value by our inversion procedure. However, it would increase the number of adjustable parameters and hence the uncertainty in each parameter. For measuring the elastic constants, it is desirable to measure independently the thickness, and, if possible, the density of the film.

As in other methods of inversion,[26,29] our inversion procedure does not give a unique set of elastic constants of the film since, in general, a system of nonlinear equations has several sets of solutions. Physical considerations are needed to choose a particular set of elastic constants. Because of the computational efficiency of our inversion algorithm, we were able to find several sets of parameters, which give about as good a fit as in Fig. 2 with our chosen set. In these sets, the values of c_{11} and c_{13} do not vary by more than 0.5% but c_{33} and c_{55} can vary as much as 10% and 50%, respectively. We have chosen the set reported in Table I, on considerations such as the sign of the anisotropy parameter, relative magnitudes of the elastic constants, value of density, etc.

The lack of uniqueness in the derived values of the set of elastic can possibly be reduced by making measurements (i) at higher frequencies and (ii) in different directions, which are not symmetry related, (iii) of wave forms using a well-characterized point source, and (iv) using independently measured values of the density. A more detailed discussion of uncertainties in the values of the elastic constants obtained by our inversion process will be discussed elsewhere. The velocities at low frequencies are not very sensitive to the properties of the film and are mainly determined by the substrate. In our calculations, we have assumed the thickness of the substrate to be infinite.

Finally, we calculate the wave forms at the surface of the same solid, that is, TiN film on Si. We consider a pulsed line force that has a delta-function dependence on time as represented by Eq. (3). We assume the source to be a unit force applied normal to the film, that is, $\boldsymbol{\eta}=(0,0,1)$. We take the inverse Laplace transform of Eq. (13) by integrating numerically over the Bromwich contour and over positive values of $q_1=\xi_1$ as described in Appendix A. The values of the elastic constants and the density are as given in column 2 of Table I.

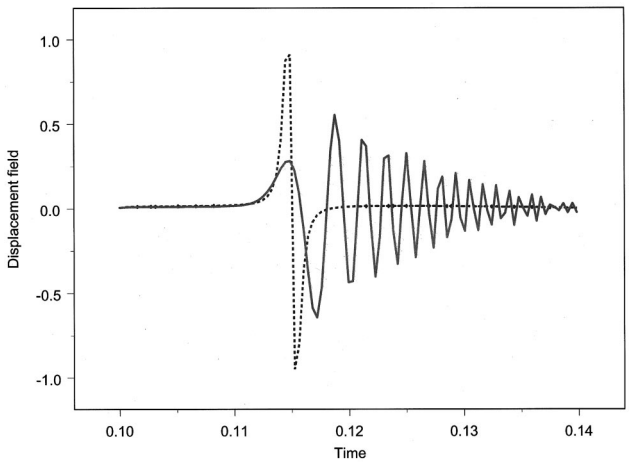


FIG. 3. Z component of the displacement field in arbitrary units at the surface of TiN film on Si at $x=0.5844$ mm in the [110] direction due to a pulsed $\delta(t)$ line source at $x=0$. Dashed line, a (film thickness)$=0$ (no film); solid line, $a=306$ nm. Time in microseconds.

Figures 3 and 4 show the Z (vertical) component of the displacement field in arbitrary units at $x_1=0.5844$ mm in the [110] and [100] directions, respectively, as function of time in microseconds for a film of thickness 306 micrometers. The curves for the horizontally polarized wave form are similar and not shown. The curve for $a=0$ corresponds to that for half-space silicon without any film. In both the figures, a comparison between the solid and the dashed curves shows the effect of dispersion on the displacement field. In a half-space solid without a film, there is no dispersion in the propagation of surface waves. The figures show that the effect of the dispersion is to separate out the frequencies in the wave packet that results into the oscillatory nature of the wave form. The figures also show that, as expected from the dispersion data of Fig. 2, the higher frequencies arrive later than the lower frequencies. The shape of the wave form is in qualitative agreement with the experimental results[32] and those calculated by Mourad *et al.*[1] for thin films by using the perturbation theory.

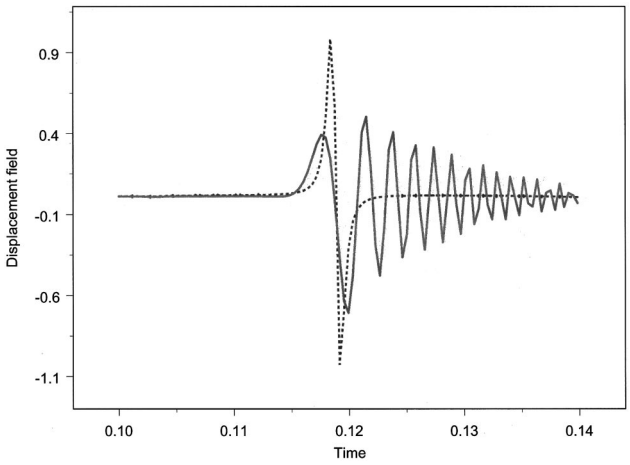


FIG. 4. Z component of the displacement field in arbitrary units at the surface of TiN film on Si at $x=0.5844$ mm in the [100] direction due to a pulsed $\delta(t)$ line source at $x=0$. Dashed line, a (film thickness)$=0$ (no film); solid line, $a=306$ nm. Time in microseconds.

The position of the first peak in the solid curve in both Figs. 3 and 4 corresponds to the first arrival of the low frequency wave which is a characteristic of the substrate. The SAW velocity in silicon in the [110] direction is 5.08 km/s which travels 0.5844 mm in 0.115 microseconds. This is the position of the first peak in Fig. 3. Similarly, the position of the first peak in Fig. 4 is at 0.118 microseconds that corresponds to the SAW velocity 4.92 km/s in silicon in the [100] direction. In Figs. 3 and 4, the arrival of the longitudinal wave is not evident because its amplitude is too small. The slight difference between the position of the first peak in the solid and the dotted curves is due to sampling of points near the singularity in the dotted curve.

A quantitative comparison with the experimental results is not possible because the actual shape of the wave form depends upon the exact space–time dependence of the applied force which is not known. We can, in principle, assume the source to be Gaussian in space and time and convolve the Green's function with the assumed Gaussian curve. However, a precise numerical comparison between the theoretical and experimental curves will still not be possible because of the uncertainty in the width and the height of the Gaussian. The quantitative space–time dependence of the source depends upon the power and duration of the laser beam as well as the thermal characteristics of the sample. We expect that the assumption of a Gaussian source will not change the qualitative nature of our results because the delta function pulse is a limiting form of the Gaussian.

V. CONCLUSIONS

The main results and conclusions of this paper are summarized below.

(1) A theoretical formulation has been developed for calculation of wave forms and velocities, and dispersion relations for surface acoustic waves in anisotropic films on anisotropic substrates. The formulation is based upon the delta-function representation of the elastodynamic Green's function.

(2) A computationally efficient algorithm has been proposed for determining the values of elastic constants and the density of the film from the measured SAW dispersion curves. The algorithm consists of using the measured dispersion data directly in the expression for the poles of the Green's function, which gives a set of nonlinear equations for determination of the material parameters of the film.

(3) The theory has been applied to interpret measurements of SAW dispersion in a polycrystalline TiN film textured in the ⟨111⟩ direction on single-crystal silicon. The inversion algorithm has been applied to the experimental dispersion curve in order to determine the values of c_{11}, c_{13}, c_{55}, c_{33}, and the density of the TiN film. The results are consistent with the elastic constants of TiN films reported in the literature.

(4) The wave forms are calculated on the surface of the film for a pulsed line source with delta function time dependence. The effect of the film is to introduce dispersion in the wave propagation that separates the frequencies in the pulse and results into oscillatory wave forms.

(5) For better determination of the elastic constants, dispersion should be made in more directions and higher frequencies, and the density and the thickness of the film be measured independently.

Note added in proof: For more details of the experimental results given in Ref. 32, please see: D. C. Hurley, V. K. Tewary, and A. J. Richards, ''Surface acoustic wave methods to determine the anisotropic elastic properties of thin films,'' Meas. Sci. Technol. **12**, 1486–1494 (2001).

ACKNOWLEDGMENTS

The author thanks Dr. Donna Hurley for giving her detailed experimental results on SAW dispersion in TiN on Si and many helpful discussions. The author also thanks Dr. George Alers and Dr. John Wendoloski for their useful comments and suggestions.

APPENDIX A: CALCULATION OF THE SURFACE GREEN'S FUNCTION

We derive a solution of Eq. (1) in terms of the surface Green's function in slowness space using the delta-function representation.[2,3] We omit the superscripts for brevity. As given in Ref. 2, we can write the homogeneous solution of Eq. (1) as follows:

$$\mathbf{u}(\mathbf{x},t)=\int \mathbf{G}_h(\mathbf{q})\mathbf{F}(\mathbf{q},t')\,\delta[\mathbf{q}\cdot\mathbf{x}-(t-t')]d\mathbf{q}\,dt', \quad \text{(A1)}$$

where $\mathbf{F}(\mathbf{q},t)$ is an arbitrary but integrable vector function,

$$\mathbf{G}_h(\mathbf{q})=\lim_{\varepsilon\to+0}\ \mathrm{Im}[\mathbf{\Lambda}(\mathbf{q})-(1-\iota\varepsilon)\rho\mathbf{I}]^{-1}, \quad \text{(A2)}$$

$$\Lambda_{ij}(\mathbf{q})=c_{ikjl}q_kq_l, \quad \text{(A3)}$$

is the Christoffel matrix in slowness space, $\mathbf{I}$ is the identity matrix, and $\mathbf{q}$ is a vector in the slowness space. The slowness vector has the dimensions of inverse velocity so $\mathbf{q}\cdot\mathbf{x}$ has the dimensions of time. The matrix $\mathbf{G_h(q)}$ is the imaginary part of the Green's function in slowness space. The subscript h on $\mathbf{G}$ denotes that it corresponds to the homogeneous solution of the Christoffel equation.

The integration in Eq. (A1) is over the entire vector space of $\mathbf{q}$ and t' subject to the condition that $t'<t$ in order to ensure causality. For a homogeneous infinite solid $-\infty<q_i<\infty$. For a bounded solid, the range of q_i depends upon the boundary conditions. We choose $\mathbf{F}$ such that $\mathbf{u}$ in Eq. (A1) satisfies the boundary condition at the surface given by Eq. (3).

Using the following representation of the delta function of a real variable y,

$$\delta(y)=(1/\pi)\lim_{\varepsilon\to+0}\mathrm{Im}[1/(y-\iota\varepsilon)], \quad \text{(A4)}$$

it can be shown[2] that

$$\mathbf{G}_h(\mathbf{q})=\mathbf{M}(\mathbf{q})\,\delta[D(\mathbf{q})] \quad \text{(A5)}$$

and

$$[\mathbf{\Lambda}(\mathbf{q})-\rho\mathbf{I}]\mathbf{G}_h(\mathbf{q})=0, \quad \text{(A6)}$$

where $\mathbf{M}(\mathbf{q})$ is the matrix of cofactors of $[\mathbf{\Lambda}(\mathbf{q})-\rho\mathbf{I}]$, and $D(\mathbf{q})$ is its determinant. In order to prove Eq. (A6), we use the property that the product of a matrix with its cofactor matrix is its determinant times the identity matrix, and the product of a variable with its delta function is 0. The constant factors on the RHS of Eq. (A5) are absorbed in $\mathbf{F}(\mathbf{q},\mathbf{t}')$ since it is an arbitrary function.

Any derivative or antiderivative of the delta function can be used in Eq. (A1). The choice depends upon the space–time dependence of the applied load. As shown below, the delta function in Eq. (A1) corresponds to a point force having a step-function dependence on time or to a line force having a $\delta(t)$ dependence.[2]

The surface Green's function has to satisfy the boundary condition given by Eq. (3). At the surface, $x_3=0$. The Laplace transform of the traction vector defined by Eq. (4) and given by Eq. (15) at the surface is written in the form

$$\mathbf{T}(\boldsymbol{\chi},s)/s=\int \mathbf{\Psi}(\boldsymbol{\xi},s)\mathbf{F}(\boldsymbol{\xi},s)\exp(-s\boldsymbol{\xi}\cdot\boldsymbol{\chi})d\boldsymbol{\xi}, \quad \text{(A7)}$$

where

$$\mathbf{\Psi}(\boldsymbol{\xi},s)=\int \mathbf{S}(\mathbf{q})\mathbf{G_h}(\mathbf{q})dq_3, \quad \text{(A8)}$$

where $\boldsymbol{\chi}$, $\boldsymbol{\xi}$, and $\mathbf{S}$ have been defined, respectively, in Eqs. (3), (9), and (12).

In order to satisfy the surface boundary condition, we choose $\mathbf{F}$ such that

$$\mathbf{\Psi}(\boldsymbol{\xi},s)\mathbf{F}(\boldsymbol{\xi},s)=\boldsymbol{\eta}. \quad \text{(A9)}$$

This gives from Eq. (A7)

$$\mathbf{T}(\boldsymbol{\chi},s)=\boldsymbol{\eta}\int s\exp(-s\boldsymbol{\xi}\cdot\boldsymbol{\chi})d\boldsymbol{\xi}. \quad \text{(A10)}$$

The inverse Laplace transform of the above equation is

$$\mathbf{T}(\boldsymbol{\chi},t)=\boldsymbol{\eta}\int \delta'(t-\boldsymbol{\xi}\cdot\boldsymbol{\chi})d\boldsymbol{\xi}, \quad \text{(A11)}$$

where the prime denotes the first derivative of the delta function with respect to t. In deriving Eq. (A11), we have used the fact that the inverse Laplace transform of $\exp(-sp)$ for any real p is $\delta(t-p)$.

In case of a point source on the surface, the limits of integration over ξ_1, ξ_2 are taken to be $-\infty$ to ∞. We use the following Fourier representation of the delta function for any real p:

$$\delta(p)=(1/2\pi)\int_{-\infty}^{+\infty}\exp(\iota\omega p)d\omega. \quad \text{(A12)}$$

From Eqs. (A11) and (A12),

$$\mathbf{T}(\chi,t)=\boldsymbol{\eta}(\iota/2\pi)\int d\boldsymbol{\xi}\left[\int_0^{+\infty}\omega\exp\iota\omega(t-\boldsymbol{\xi}\cdot\boldsymbol{\chi})d\omega - \int_0^{+\infty}\omega\exp-\iota\omega(t-\boldsymbol{\xi}\cdot\boldsymbol{\chi})\right]d\omega. \quad \text{(A13)}$$

In Eq. (A13), we have separated the integral over the entire range of ω in two integrals, each over positive values of ω. Since $\boldsymbol{\xi}\cdot\boldsymbol{\chi}=\xi_1\chi_1+\xi_2\chi_2$, we can carry out the integral

over ξ_1 and ξ_2 independently between $-\infty$ to ∞ using Eq. (A12). This gives

$$\mathbf{T}(\chi,t)=\boldsymbol{\eta}(2\pi\iota)\int_0^{+\infty}\omega[\exp\iota\omega t-\exp-\iota\omega t]\times\delta(\omega\chi_1)\delta(\omega\chi_2)d\omega=-4\pi^2H(t)\delta(\boldsymbol{\chi})\eta, \tag{A14}$$

where $H(t)$ is the Heaviside step function and the delta function for vector arguments is defined in Eq. (4).

In case of a line force, we need to take the inverse Laplace transform of Eq. (A10) in the following manner. The integral in Eq. (A10) reduces to a 1D integral,

$$\mathbf{T}(\chi_1,s)=\boldsymbol{\eta}s\int_0^{\infty}\exp(-s\xi_1\chi_1)d\xi_1. \tag{A15}$$

In this case, the limits of integral over ξ are taken to be from $\xi_1=0$ to $+\infty$ for $\chi_1>0$. We write the above integral in the upper half of the Bromwich contour in the complex s plane in the following form:

$$\mathbf{T}(\chi_1,s)=\boldsymbol{\eta}s\ \mathrm{Lim}_{\varepsilon\to+0}\int_0^{\infty}[\exp-s\xi_1(\chi_1+\iota\varepsilon)-\exp-s\xi_1(\chi_1-\iota\varepsilon)]d\xi_1. \tag{A16}$$

The limit has to be evaluated after the integration has been carried out. We now integrate over ξ_1 subject to the following constraint at the upper limit of the Bromwich contour and for $\chi_1>0$:

$$\chi_1\,\mathrm{Re}(s)>\varepsilon\,\mathrm{Im}(s)>0. \tag{A17}$$

Using Eq. (A4) and the constraint (A17), Eq. (A16) reduces to

$$\mathbf{T}(\chi_1,s)=-2\pi\boldsymbol{\eta}\delta(\chi_1). \tag{A18}$$

Finally, the inverse Laplace transform of Eq. (A18) gives

$$\mathbf{T}(\chi_1,t)=-2\pi\boldsymbol{\eta}\delta(\chi_1)\delta(t), \tag{A19}$$

as required by the boundary condition given in Eq. (3). In actual numerical calculations, we take $s=\eta+\iota\omega$ and integrate numerically over ω from $-W$ to W and over $q_1=\xi_1$ from 0 to Q, where $W\to+\infty$, $Q\to+\infty$, and $\varepsilon\to0$ with the condition $\eta\chi_1\gg W\varepsilon$. The value of Q is chosen such that the exponentials in Eq. (A16) are vanishingly small at $\xi_1=Q$. A test of numerical convergence is when the results are not sensitive to the values of η, W, ε, and Q.

APPENDIX B: INTEGRATION OF GREEN'S FUNCTION IN SLOWNESS SPACE

In Eqs. (10) and (11), we need to evaluate an integral of the type

$$N(\boldsymbol{\xi})=\int P(\boldsymbol{\xi},q_3)\delta[D(\boldsymbol{\xi},q_3)]dq_3, \tag{B1}$$

where $P(\boldsymbol{\xi},q_3)$ is an integrable function of $\boldsymbol{\xi}$ and q_3. The function $D(\mathbf{q})$ is, in general, a sixth degree polynomial in q_3. We can therefore write

$$D(\mathbf{q})=C\prod_{\alpha}(q_3-Q_\alpha), \tag{B2}$$

where C is the coefficient of the largest power of q_3, that is q_3^6, in $D(\boldsymbol{\xi},q_3)$ and Q_α ($\alpha=1-6$) are six roots of the polynomial equation

$$D(\boldsymbol{\xi},q_3)=0. \tag{B3}$$

For an orthorhombic solid, $D(\boldsymbol{\xi},q_3)$ will be a cubic polynomial in q_3^2. The set of six roots Q_α will therefore consist of three pairs of roots of opposite signs. If all the roots are real, we order the roots and choose three positive roots Q_i such that $Q_{i+3}=-Q_i$ ($i=1-3$). From Eq. (B2), after doing the partial fractions in $1/(q_3-Q_\alpha)$, we obtain

$$\delta[D(\boldsymbol{\xi},q_3)]=\sum_i\delta[q_3-Q_i]\Big/\left[2CQ_i\prod_{j\neq i}(Q_i^2-Q_j^2)\right]. \tag{B4}$$

Note that unlike the standard definition of the delta function of a product, it is not needed to take the modulus of the weight factor, which is the coefficient of the delta function in the RHS of Eq. (B4). This is because of the manner in which the partial fractions have been ordered in Eq. (B4).

From Eqs. (B1) and (B4),

$$N(\boldsymbol{\xi})=\sum_i P(\boldsymbol{\xi},q_3=Q_i)\Big/\left[2CQ_i\prod_{j\neq i}(Q_i^2-Q_j^2)\right]. \tag{B5}$$

For some values of $\boldsymbol{\xi}$, Q_i may be complex. For an elastically stable solid, $D(\boldsymbol{\xi},q_3)$ must be real. Since the delta function is defined only for real arguments, Eq. (B1) is valid but the delta function in Eq. (B4) needs to be properly defined. We show that even for complex roots of q_3, an equation similar to Eq. (B4) can be formally written.

In Eq. (B1), we transform the integration variable from q_3 to the real variable $D(\mathbf{q})$ and express $q_3=\lambda(\boldsymbol{\xi},D)$ as function of the determinant D where λ may be complex. This gives

$$N(\boldsymbol{\xi})=\int P(\boldsymbol{\xi},D)\delta[D]dD/(dD/dq_3)\quad\text{for}\quad q_3=\lambda(\boldsymbol{\xi},D). \tag{B6}$$

Since Q_α is a root of Eq. (B3), we have

$$\lambda_\alpha(\boldsymbol{\xi},D=0)=Q_\alpha. \tag{B7}$$

Integrating over D in Eq. (B6) using the properties of the delta function gives Eq. (B5). Hence we can formally use Eq. (B4) even for complex Q_j. This does not mean that the delta function is defined for complex arguments. For example, if Q_j is complex, the delta function in Eq. (B4) cannot be represented by Eq. (A4). For complex Q_j, Eq. (B4) means only that the result of the integration over q_3 is to replace q_3 in the integrand by Q_i and multiply by the weight factor as in Eq. (B4). However, we need to choose and order the roots appropriately in Eq. (B4). For example, in the upper half of the Bromwich contour, since $\mathrm{Im}(s)>0$, we choose the three roots that have negative imaginary parts to ensure that the stress given by Eq. (27) vanishes as $x_3\to\infty$.

APPENDIX C: CHRISTOFFEL EQUATION FOR A GENERAL ORTHORHOMBIC SOLID

We assume that the orthorhombic axes are parallel to the Cartesian axes. A general orthorhombic solid has nine independent elastic constants. The elements of the Christoffel matrix in slowness space[7] are given below

$$\begin{aligned}
&\Lambda_{11}(\mathbf{q})=c_{11}q_1^2+c_{66}q_2^2+c_{55}q_3^2,\\
&\Lambda_{22}(\mathbf{q})=c_{66}q_1^2+c_{22}q_2^2+c_{44}q_3^2,\\
&\Lambda_{33}(\mathbf{q})=c_{55}q_1^2+c_{44}q_2^2+c_{33}q_3^2,\\
&\Lambda_{12}(\mathbf{q})=\Lambda_{21}(\mathbf{q})=(c_{12}+c_{66})q_1q_2,\\
&\Lambda_{13}(\mathbf{q})=\Lambda_{31}(\mathbf{q})=(c_{13}+c_{55})q_1q_3,\\
&\Lambda_{23}(\mathbf{q})=\Lambda_{32}(\mathbf{q})=(c_{23}+c_{44})q_2q_3.
\end{aligned} \tag{C1}$$

In Sec. IV our interest is in calculating the Green's function for a line force parallel to the Y axis. In this case, $q_2=0$. The matrix Λ is therefore, block diagonalized. The relevant block matrix is 2×2 containing the elements (1,1), (1,3), (3,1), and (3,3). The elements of the 2×2 matrix $\mathbf{M}(\mathbf{q})$ as defined in Eq. (A5) are given below

$$\begin{aligned}
&M_{11}(\mathbf{q})=\Lambda_{33}(\mathbf{q})-\rho,\quad M_{33}(\mathbf{q})=\Lambda_{11}(\mathbf{q})-\rho,\\
&M_{13}(\mathbf{q})=M_{31}(\mathbf{q})=-\Lambda_{13}(\mathbf{q}).
\end{aligned} \tag{C2}$$

The determinant of $\Lambda(\mathbf{q})$ is a 4-degree polynomial in q_3. In the present case, it is a quadratic polynomial in q_3^2 as written below

$$D(\mathbf{q})=C(q_3^4+2d_2q_3^2+d_0)=C\prod_{m=1,2}(q_3^2-Q_m^2), \tag{C3}$$

where, as defined in Eq. (B2),

$$C=c_{33}c_{55}, \tag{C4}$$

$$2Cd_2=c_{55}(c_{55}q_1^2-\rho)+c_{33}(c_{11}q_1^2-\rho)-(c_{13}+c_{55})q_1^2, \tag{C5}$$

$$Cd_0=(c_{55}q_1^2-\rho)(c_{11}q_1^2-\rho), \tag{C6}$$

and the other coefficients are obtained in terms of the elastic constants by expanding the determinant. The roots Q_m of Eq. (B3) are

$$Q_{1,2}^2=-d_2\pm(d_2^2-d_0)^{1/2}. \tag{C7}$$

Equation (C1) reduces to the equation for solids of higher symmetry by imposing the following well-known relations amongst the elastic constants:

Tetragonal (crystallographic axes parallel to the Cartesian axes with c-axis parallel to the Z axis)

$$c_{44}=c_{55},\quad c_{22}=c_{11},\quad c_{13}=c_{23}. \tag{C8}$$

Hexagonal [in addition to Eq. (C8)]

$$c_{66}=0.5(c_{11}-c_{12}). \tag{C9}$$

Cubic [in addition to Eq. (C8)]

$$c_{66}=c_{44},\quad c_{33}=c_{11},\quad c_{12}=c_{23}. \tag{C10}$$

Isotropic [in addition to Eq. (C10)]

$$c_{44}=0.5(c_{11}-c_{12}). \tag{C11}$$

[1] A. Mourad, C. Desmet, W. Lauriks, H. Coufal, and J. Thoen, "The Green's function for surface acoustic waves: Comparison between theory and experiment," J. Acoust. Soc. Am. **100**, 1538–1541 (1996).
[2] V. K. Tewary, "Computationally efficient representation for elastodynamic and elastostatic Green's functions for anisotropic solids," Phys. Rev. B **51**, 15695–15702 (1995).
[3] C. Y. Wang and J. D. Achenbach, "A new method to obtain 3-D Green's functions for anisotropic solids," Wave Motion **18**, 273–289 (1993).
[4] A. G. Every and K. Y. Kim, "Time domain dynamic response functions of elastically anisotropic solids," J. Acoust. Soc. Am. **95**, 2505–2516 (1994).
[5] T. C. T. Ting, *Anisotropic Elasticity: Theory and Applications* (Oxford University Press, Oxford, 1996).
[6] V. K. Tewary and C. M. Fortunko, "Surface waves in three-dimensional half-space tetragonal solids," J. Acoust. Soc. Am. **100**, 86–88 (1996).
[7] V. K. Tewary, M. Mahapatra, and C. M. Fortunko, "Green's function for anisotropic half-space solids in frequency space and calculation of mechanical impedance," J. Acoust. Soc. Am. **100**, 2960–2963 (1996).
[8] D. M. Barnett and J. Lothe, "Free surface (Rayleigh) waves in anisotropic elastic half-spaces: the surface impedance methods," Proc. R. Soc. London, Ser. A **402**, 135–152 (1985).
[9] B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990), Vol. I.
[10] A. H. Nayfeh, *Wave Propagation in Layered Anisotropic Media with Applications to Composites* (Elsevier, Amsterdam, 1995).
[11] Y. Li and R. B. Thomson, "Influence of anisotropy on the dispersion characteristics of guided plate modes," J. Acoust. Soc. Am. **87**, 1911–1931 (1990).
[12] L. P. Solie and B. A. Auld, "Elastic waves in free anisotropic plates," J. Acoust. Soc. Am. **54**, 50–64 (1973).
[13] A. Mourad and M. Deschamps, "Lamb's problem for an anisotropic half-space studied by the Cagniard de Hoop method," J. Acoust. Soc. Am. **97**, 3194–3197 (1995).
[14] A. G. Every, K. Y. Kim, and A. A. Maznev, "The elastodynamic response of a semi-infinite anisotropic solid to sudden surface loading," J. Acoust. Soc. Am. **102**, 1346–1355 (1997).
[15] V. K. Kinra and V. Dayal, "A new technique for ultrasonic-nondestructive evaluation of thin specimens," Exp. Mech. **28**, 288–297 (1988).
[16] D. E. Chimenti, "Guided waves in plates and their use in materials characterization," Appl. Mech. Rev. **50**, 247–284 (1997).
[17] D. M. Barnett, J. Lothe, S. D. Gavazza, and M. J. P. Musgrave, "Considerations of the existence of interfacial (Stoneley) waves in bonded anisotropic elastic half-spaces," Proc. R. Soc. London, Ser. A **402**, 153–166 (1985).
[18] E. N. Its and J. S. Lee, "Propagation of surface waves across an anisotropic layer," ASME Trans. J. Appl. Mech. **61**, 596–604 (1994).
[19] E. Pan and S. K. Datta, "Ultrasonic waves in multilayered superconducting plates," J. Appl. Phys. **86**, 543–551 (1999).
[20] S. S. Lih and A. K. Mal, "On the accuracy of approximate plate theories for wave fields calculations in composite plates," Wave Motion **21**, 17–34 (1995).
[21] A. J. Niklasson, S. K. Datta, and M. L. Dunn, "On approximate guided waves in plates with thin anisotropic coatings by means of effective boundary conditions," J. Acoust. Soc. Am. **108**, 924–933 (2000).
[22] G. W. Farnell and E. L. Adler, in *Elastic Wave Propagation in Thin Layers*, edited by W. P. Mason and R. N. Thurston [(Academic, New York, 1972), Phys. Acoust. **9**, 35–127 (1972)].
[23] A. G. Every and K. Y. Kim, "Determination of elastic constants of anisotropic solids from elastodynamic Green's functions," Ultrasonics **34**, 471–472 (1996).
[24] K. Y. Kim, A. G. Every, and W. Sachse, "Determination of the elastic constants of anisotropic solids from group velocities measured in symmetry directions," Int. J. Mod. Phys. B **10**, 235–246 (1996).
[25] C. M. Flannery, E. Chilla, S. Semenov, and H.-J. Fröhlich, "Elastic properties of GaAs obtained by inversion of laser-generated surface acoustic wave measurements," IEEE Ultrasonics Symposium, 1999, p. 501–504.
[26] S. Makarov, E. Chilla, and H.-J. Fröhlich, "Determination of elastic constants of thin films from phase velocity dispersion of different surface acoustic wave modes," J. Appl. Phys. **78**, 5028–5034 (1995).
[27] D. Schneider, O. Ollendorf, and T. Schwarz, "Nondestructive evaluation of the mechanical behavior of TiN-coated steels by laser-induced ultra-

sonic surface waves,'' Appl. Phys. A: Mater. Sci. Process. **61**, 277–284 (1995).
[28] D. Schneider, T. Schwarz, H-J. Scheibe, M. Panzner, ''Non-destructive evaluation of diamond and diamond like carbon films by laser induced surface acoustic waves,'' Thin Solid Films **295**, 107–116 (1997).
[29] D. Schneider, T. Schwarz, and B. Schultrich, ''Determination of elastic modulus and thickness of surface layers by ultrasonic surface waves,'' Thin Solid Films **219**, 92–102 (1992).
[30] A. Neubrand and P. Hess, ''Laser generation and detection of surface acoustic waves: elastic properties of surface layers,'' J. Appl. Phys. **71**, 227–238 (1992).
[31] T. T. Wu and Y. H. Liu, ''Inverse determination of thickness and elastic properties of a bonding layer using laser-generated surface waves,'' Ultrasonics **37**, 23–30 (1999).
[32] D. C. Hurley and A. J. Richards ''Laser-ultrasonic methods for thin-film property measurements using high-frequency surface acoustic waves,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti [(AIP, Melville, NY, 2001), AIP Conf. Proc. **20**, 263–270 (2001)].
[33] V. K. Tewary, R. H. Wagoner, and J. P. Hirth, ''Elastic Green's function for a composite solid with a planar interface,'' J. Mater. Res. **4**, 113–123 (1989).
[34] A. Bendavid, P. J. Martin, R. P. Netterfield, and T. J. Kinder, Surf. Coat. Technol. **70**, 97–106 (1994).
[35] J. P. Hirth and J. Lothe, *Theory of Dislocations*, 2nd ed. (Krieger, Malabar, FL, 1992).
[36] R. Y. Fillit and A. J. Perry, Surf. Coat. Technol. **36**, 647–659 (1988).
[37] W. J. Meng and G. L. Eesley, Thin Solid Films **271**, 108–116 (1995).
[38] J. O. Kim and J. D. Achenbach, Thin Solid Films **214**, 25–34 (1992).
[39] W. Pang, A. G. Every, J. D. Comins, P. R. Stoddart, and X. Zhang, J. Appl. Phys. **86**, 311–317 (1999).

Wave propagation in a solid cylinder of arbitrary cross-section immersed in fluid

M. Venkatesan and P. Ponnusamy
School of Mathematics, Anna University, Chennai-600 025, India

(Received 22 October 2001; revised 14 June 2002; accepted 18 June 2002)

The problem of wave propagation in a solid cylinder of arbitrary cross-section immersed in fluid is studied using the Fourier expansion collocation method. The frequency equations are obtained for longitudinal and flexural vibrations and are studied numerically for elliptic and cardioidal cylinders and are presented in the tabular form and also in the graphical form. The general theory can be used to study any kind of cylinder with proper geometric relations. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1499130]

PACS numbers: 43.35.Cg, 43.40.Ey, 43.40.Cw [ANN]

I. INTRODUCTION

A through knowledge of various wave propagation characteristics, as a function of material and geometrical parameters, is necessary for a wide range of applications, from geophysical prospecting in cased holes, nondestructive evaluation of oil and gas pipelines, to the insulated fiber optic cables for data transmission.

Gazis[1] has studied the most general form of harmonic waves in a hollow cylinder of circular cross-section of infinite length. He has presented the frequency equation in Part I and numerical results in Part II in detailed form. Nagaya[2,3] has discussed wave propagation in infinite bar of arbitrary cross-section applicable to a bar of general cross-section, based on the three-dimensional theory of elasticity. The boundary conditions along the free surface of arbitrary cross-section are satisfied by means of Fourier expansion collocation method. Paul and Venkatesan[4] have obtained the frequency equation of the vibration of a solid piezoelectric cylinder of arbitrary cross-section using the Fourier expansion collocation method and analyzed it for an elliptic cylinder. Sinha *et al.*[5] have discussed the axisymmetric wave propagation in circular cylindrical shell immersed in fluid, in two parts. In Part I, the theoretical analysis of the propagating modes is discussed and in Part II, the axisymmetric modes excluding torsional modes are obtained theoretically and experimentally and are compared. Berlinear and Solecki[6] have studied the wave propagation in fluid loaded transversely isotropic cylinder. In that paper, Part I consists of the analytical formulation of the frequency equation of the coupled system consisting of the cylinder with inner and outer fluid and Part II gives the numerical results.

The objective of this paper is to study the wave propagation in a solid cylinder of arbitrary cross-section immersed in inviscid fluid. The free vibration of the rod immersed in fluid is studied using the Fourier expansion collocation method and the frequency equation is obtained. The frequency equations of longitudinal and flexural modes are analyzed numerically for an elliptic cylinder and the computed frequencies are tabulated in tables for different aspect ratios (a/b) and plotted the dispersion curves for $(a/b)=1.3$ and 1.7. Similarly the dispersion curves, for cardioidal cross-section, are also drawn.

II. FORMULATION OF THE PROBLEM AND BOUNDARY CONDITIONS

The three-dimensional wave equations in cylindrical polar coordinates r, θ, and z are

$$\begin{aligned}
&(\lambda+2\mu)(u_{,rr}+r^{-1}u_{,r}-r^{-2}u)+r^{-1}(\lambda+\mu)v_{,r\theta}\\
&\quad+(\lambda+\mu)w_{,rz}+r^{-2}\mu u_{,\theta\theta}-r^{-2}(\lambda+3\mu)u_{,\theta}+\mu u_{,zz}\\
&\quad=\rho u_{,tt},\\
&\mu(v_{,rr}+r^{-1}v_{,r}-r^{-2}v)+r^{-2}(\lambda+2\mu)u_{,\theta\theta}\\
&\quad+\mu v_{,zz}+r^{-1}(\lambda+\mu)u_{,r\theta}+r^{-2}(\lambda+\mu)u_{,\theta}\\
&\quad+r^{-1}(\lambda+\mu)w_{,\theta z}=\rho v_{,tt},\qquad(1)\\
&(\lambda+2\mu)w_{,zz}+(\lambda+\mu)u_{,rz}+r^{-1}(\lambda+\mu)u_{,z}+r^{-1}\\
&\quad\times(\lambda+\mu)v_{,\theta z}+\mu(w_{,rr}+r^{-1}w_{,r}+r^{-2}w_{,\theta\theta})=\rho w_{,tt},
\end{aligned}$$

where u, v and w are the displacements along radial, circumferential and axial directions, ρ is the mass density and λ and μ are the Lame' constants.

To obtain the wave propagation of arbitrary cross-sectional rod, we seek the solution of Eq. (1) in the form

$$\begin{aligned}
u(r,\theta,z,t)&=\sum_{n=0}^{\infty}\varepsilon_n(\phi_{n,r}+r^{-1}\psi_{n,\theta})e^{i(kz+\omega t)},\\
v(r,\theta,z,t)&=\sum_{n=0}^{\infty}\varepsilon_n(r^{-1}\phi_{n,\theta}-\psi_{n,r})e^{i(kz+\omega t)},\qquad(2)\\
w(r,\theta,z,t)&=(i/a)\sum_{n=0}^{\infty}\varepsilon_n W_n e^{i(kz+\omega t)},
\end{aligned}$$

where $\varepsilon_n=1/2$ for $n=0$, $\varepsilon_n=1$ for $n\geqslant 1$, $i=\sqrt{-1}$, k is the wave number, ω is the frequency, $\phi_n(r,\theta)$, $\psi_n(r,\theta)$ and $W_n(r,\theta)$ are the displacement potentials and a is the geometrical parameter of the cylinder.

By introducing dimensionless quantities $\varsigma=ka$, $\bar{\lambda}=(\lambda/\mu)$, $\Omega^2=\rho\,\omega^2a^2/\mu$, $\bar{z}=z/a$, $T=t\sqrt{\mu/\rho}/a$, $x=r/a$ and substituting Eq. (2) in Eq. (1), we obtain

$$[(2+\overline{\lambda})\nabla^2+\Omega^2-\varsigma^2]\phi_n-\varsigma(1+\overline{\lambda})W_n=0$$
$$(1+\overline{\lambda})\nabla^2\phi_n+[\nabla^2+\Omega^2-\varsigma^2(2+\overline{\lambda})]W_n=0 \tag{3a}$$

and

$$[\nabla^2+\Omega^2-\varsigma^2]\psi_n=0, \tag{3b}$$

where $\nabla^2\equiv\partial^2/\partial x^2+x^{-1}\partial/\partial x+x^{-2}\partial^2/\partial\theta^2$.

Eliminating W_n from Eq. (3a), we obtain

$$(A\nabla^4+B\nabla^2+C)\phi_n=0, \tag{4}$$

where $A=2+\overline{\lambda}$, $B=(3+\overline{\lambda})\Omega^2-\varsigma^2(3+2\overline{\lambda})$ and $C=[\Omega^2-\varsigma^2(2+\overline{\lambda})][\Omega^2-\varsigma^2]$.

The solution of Eq. (4) for the symmetric mode is

$$\phi_n=\sum_{j=1}^{2} A_{jn}J_n(\alpha_j ax)\cos n\theta, \tag{5a}$$

$$W_n=\sum_{j=1}^{2} d_j A_{jn}J_n(\alpha_j ax)\cos n\theta. \tag{5b}$$

Similarly the solution for the antisymmetric mode is given by

$$\overline{\phi}_n=\sum_{j=1}^{2} \overline{A}_{jn}J_n(\alpha_j ax)\sin n\theta, \tag{6a}$$

$$\overline{W}_n=\sum_{j=1}^{2} d_j\overline{A}_{jn}J_n(\alpha_j ax)\sin n\theta, \tag{6b}$$

where J_n is the Bessel function of first kind, $(\alpha_j a)^2$ are the roots of the algebraic equation $A(\alpha a)^4-B(\alpha a)^2+C=0$ and the constant

$$d_j=(1+\overline{\lambda})[\Omega^2-\varsigma^2-(2+\overline{\lambda})(\alpha_j a)^2]/\varsigma. \tag{7}$$

Solving Eq. (3b), we obtain

$$\psi_n=A_{3n}J_n(\beta ax)\sin n\theta \tag{8a}$$

for symmetric mode and

$$\overline{\psi}_n=\overline{A}_{3n}J_n(\beta ax)\cos n\theta \tag{8b}$$

for the antisymmetric mode, where $(\beta a)^2=\Omega^2-\varsigma^2$. If $(\alpha a)^2<0$ and $(\beta a)^2<0$, then the Bessel function of first kind J_n is to be replaced by the modified Bessel function of first kind I_n.

In cylindrical coordinates, the acoustic pressure and radial displacement equation of motion for an inviscid fluid are of the form,[7]

$$p^f=-B^f(u^f_{,r}+r^{-1}(u^f+v^f_{,\theta})+w^f_{,z}) \tag{9}$$

and

$$c^{-2}u^f_{,tt}=\Delta_{,r}, \tag{10}$$

respectively, where (u^f,v^f,w^f) is the displacement vector, B^f is the adiabatic bulk modulus, $c=\sqrt{B^f/\rho^f}$ is the acoustic phase velocity of the fluid in which ρ^f is the density of the fluid and

$$\Delta=(u^f_{,r}+r^{-1}(u^f+v^f_{,\theta})+w^f_{,z}). \tag{11}$$

Substituting $u^f=\phi^f_{,r}$, $v^f=r^{-1}\phi^f_{,\theta}$ and $w^f=\phi^f_{,z}$, and seeking the solution of Eq. (10) in the form

$$\phi^f(r,\theta,z,t)=\sum_{n=0}^{\infty}\varepsilon_n[\phi^f_n\cos n\theta+\overline{\phi^f_n}\sin n\theta]e^{i(kz+\omega t)}, \tag{12}$$

we get

$$\phi^f_n=A_{4n}H^1_n(\delta ax), \tag{13}$$

where $(\delta a)^2=\Omega^2/\overline{\rho}\overline{B^f}-\varsigma^2$ in which $\overline{\rho}=\rho/\rho^f$, $\overline{B^f}=B^f/\mu$, H^1_n is the Hankel function of the first kind and $\overline{\phi^f_n}$ is as same as ϕ^f_n. If $(\delta a)^2<0$, then the Hankel function of first kind is to be replaced by K_n, where K_n is the modified Bessel function of second kind. By substituting the expressions of the displacement vector in terms of ϕ^f and Eq. (13) in Eq. (9), we could express the acoustic pressure as

$$p^f=\sum_{n=0}^{\infty}\varepsilon_n A_{4n}\Omega^2\overline{\rho}H^1_n(\delta ax)\cos n\theta e^{i(\overline{sz}+\Omega T)}. \tag{14}$$

In this problem, the vibration of a solid cylinder of arbitrary cross-section immersed in fluid is considered. Since the boundary is irregular, it is difficult to satisfy the boundary conditions directly. Hence, in the same lines of Nagaya[2,3] the Fourier expansion collocation method is applied. Thus the boundary conditions are obtained as

$$(\sigma_{qq}+p^f)_l=(\sigma_{qs})_l=(\sigma_{zq})_l=(u-u^f)_l=0. \tag{15}$$

where q is the coordinate normal to the boundary and s is the tangential coordinate to the boundary, σ_{qq} is the normal stress, σ_{qs}, σ_{zq} are the shearing stresses and $(\ )_l$ is the value at the lth segment of the boundary. The first and last conditions are due to the continuity of the stresses and displacements of the solid and fluid on the curved surface. If the angle γ_l between the normal to the segment and the reference axis is assumed to be constant, then the transformed expression for the stresses are given by

$$\begin{aligned}
\sigma_{qq}&=\lambda(u_{,r}+r^{-1}(u+v_{,\theta})+w_{,z})+2\mu[u_{,r}\cos^2(\theta-\gamma_l)\\
&\quad+r^{-1}(u+v_{,\theta})\sin^2(\theta-\gamma_l)\\
&\quad+0.5(r^{-1}[u-u_{,\theta}]-v_{,r})\sin 2(\theta-\gamma_l)],\\
\sigma_{qs}&=\mu[(u_{,r}-r^{-1}(v_{,\theta}+u))\sin 2(\theta-\gamma_l)\\
&\quad+(r^{-1}(u_{,\theta}-v)+v_{,r})\cos 2(\theta-\gamma_l)],\\
\sigma_{zq}&=\mu[(u_{,z}+w_{,r})\cos(\theta-\gamma_l)-(v_{,z}+r^{-1}w_{,\theta})\\
&\quad\times\sin(\theta-\gamma_l)].
\end{aligned} \tag{16}$$

The boundary conditions in Eq. (15) are transformed as

$$\begin{aligned}
&[(S_{xx})_l+(\overline{S_{xx}})_l]e^{i(\overline{sz}+(\Omega)T)}=0,\\
&[(S_{xy})_l+(\overline{S_{xy}})_l]e^{i(\overline{sz}+(\Omega)T)}=0,\\
&[(S_{xz})_l+(\overline{S_{xz}})_l]e^{i(\overline{sz}+(\Omega)T)}=0,\\
&[(S_r)_l+(\overline{S_r})_l]e^{i(\overline{sz}+(\Omega)T)}=0,
\end{aligned} \tag{17}$$

where

$$S_{xx}=0.5(A_{10}e_0^1+A_{20}e_0^2+A_{40}e_0^4)+\sum_{n=1}^{\infty}(A_{1n}e_n^1+A_{2n}e_n^2+A_{3n}e_n^3+A_{4n}e_n^4),$$

$$S_{xy}=0.5(A_{10}f_0^1+A_{20}f_0^2)+\sum_{n=1}^{\infty}(A_{1n}f_n^1+A_{2n}f_n^2+A_{3n}f_n^3),$$

$$S_{xz}=0.5(A_{10}g_0^1+A_{20}g_0^2)+\sum_{n=1}^{\infty}(A_{1n}g_n^1+A_{2n}g_n^2+A_{3n}g_n^3),$$

$$S_r=0.5(A_{10}h_0^1+A_{20}h_0^2+A_{40}h_0^4)+\sum_{n=1}^{\infty}(A_{1n}h_n^1+A_{2n}h_n^2+A_{3n}h_n^3+A_{4n}h_n^4),$$

$$\overline{S}_{xx}=0.5\overline{A}_{30}\overline{e}_0^3+\sum_{n=1}^{\infty}(\overline{A}_{1n}\overline{e}_n^1+\overline{A}_{2n}\overline{e}_n^2+\overline{A}_{3n}\overline{e}_n^3+\overline{A}_{4n}\overline{e}_n^4),$$

$$\overline{S}_{xy}=0.5\overline{A}_{30}\overline{f}_0^3+\sum_{n=1}^{\infty}(\overline{A}_{1n}\overline{f}_n^1+\overline{A}_{2n}\overline{f}_n^2+\overline{A}_{3n}\overline{f}_n^3),$$

$$\overline{S}_{xz}=0.5\overline{A}_{30}\overline{g}_0^3+\sum_{n=1}^{\infty}(\overline{A}_{1n}\overline{g}_n^1+\overline{A}_{2n}\overline{g}_n^2+\overline{A}_{3n}\overline{g}_n^3),$$

$$\overline{S}_r=0.5\overline{A}_{30}\overline{h}_0^3+\sum_{n=1}^{\infty}(\overline{A}_{1n}\overline{h}_n^1+\overline{A}_{2n}\overline{h}_n^2+\overline{A}_{3n}\overline{h}_n^3+\overline{A}_{4n}\overline{h}_n^4).$$

The equations for $e_n^1 \sim h_n^4$ are given in Appendix A.

For symmetric mode, the boundary conditions are

$$\begin{aligned}
&\sum_{n=0}^{\infty}\varepsilon_n\left[\sum_{j=1}^{2}A_{jn}e_n^j+A_{3n}e_n^3+A_{4n}e_n^4\right]=0,\\
&\sum_{n=0}^{\infty}\varepsilon_n\left[\sum_{j=1}^{2}A_{jn}f_n^j+A_{3n}f_n^3\right]=0,\\
&\sum_{n=0}^{\infty}\varepsilon_n\left[\sum_{j=1}^{2}A_{jn}g_n^j+A_{3n}g_n^3\right]=0,\\
&\sum_{n=0}^{\infty}\varepsilon_n\left[\sum_{j=1}^{2}A_{jn}h_n^j+A_{3n}h_n^3+A_{4n}h_n^4\right]=0,
\end{aligned}\tag{18a}$$

and for antisymmetric mode,

$$\begin{aligned}
&\sum_{n=0}^{\infty}\varepsilon_n\left[\sum_{j=1}^{2}\overline{A}_{jn}\overline{e}_n^j+\overline{A}_{3n}\overline{e}_n^3+\overline{A}_{4n}\overline{e}_n^4\right]=0,\\
&\sum_{n=0}^{\infty}\varepsilon_n\left[\sum_{j=1}^{2}\overline{A}_{jn}\overline{f}_n^j+\overline{A}_{3n}\overline{f}_n^3\right]=0,\\
&\sum_{n=0}^{\infty}\varepsilon_n\left[\sum_{j=1}^{2}\overline{A}_{jn}\overline{g}_n^j+\overline{A}_{3n}\overline{g}_n^3\right]=0,\\
&\sum_{n=0}^{\infty}\varepsilon_n\left[\sum_{j=1}^{2}\overline{A}_{jn}\overline{h}_n^j+\overline{A}_{3n}\overline{h}_n^3+\overline{A}_{4n}\overline{h}_n^4\right]=0.
\end{aligned}\tag{18b}$$

Performing the Fourier series expansion to Eq. (18) along the boundary, the boundary conditions are expanded in the form of double Fourier series. For symmetric mode

$$\begin{aligned}
&\sum_{m=0}^{\infty}\varepsilon_m\left[E_{m0}^1A_{10}+E_{m0}^2A_{20}+E_{m0}^4A_{40}+\sum_{n=1}^{\infty}(E_{mn}^1A_{1n}+E_{mn}^2A_{2n}+E_{mn}^3A_{3n}+E_{mn}^4A_{4n})\right]\cos m\theta=0,\\
&\sum_{m=0}^{\infty}\varepsilon_m\left[F_{m0}^1A_{10}+F_{m0}^2A_{20}+\sum_{n=1}^{\infty}(F_{mn}^1A_{1n}+F_{mn}^2A_{2n}+F_{mn}^3A_{3n})\right]\sin m\theta=0,\\
&\sum_{m=0}^{\infty}\varepsilon_m\left[G_{m0}^1A_{10}+G_{m0}^2A_{20}+\sum_{n=1}^{\infty}(G_{mn}^1A_{1n}+G_{mn}^2A_{2n}+G_{mn}^3A_{3n})\right]\cos m\theta=0,\\
&\sum_{m=0}^{\infty}\varepsilon_m\left[H_{m0}^1A_{10}+H_{m0}^2A_{20}+H_{m0}^4A_{40}+\sum_{n=1}^{\infty}(H_{mn}^1A_{1n}+H_{mn}^2A_{2n}+H_{mn}^3A_{3n}+H_{mn}^4A_{4n})\right]\cos m\theta=0,
\end{aligned}\tag{19}$$

and for antisymmetric mode,

$$\begin{aligned}
&\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(\overline{E}_{mn}^1\overline{A}_{1n}+\overline{E}_{mn}^2\overline{A}_{2n}+\overline{E}_{mn}^3\overline{A}_{3n}+\overline{E}_{mn}^4\overline{A}_{4n})\times\sin m\theta=0,\\
&\sum_{m=0}^{\infty}\varepsilon_m\sum_{n=0}^{\infty}(\overline{F}_{mn}^1\overline{A}_{1n}+\overline{F}_{mn}^2\overline{A}_{2n}+\overline{F}_{mn}^3\overline{A}_{3n})\cos m\theta=0,\\
&\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(\overline{G}_{mn}^1\overline{A}_{1n}+\overline{G}_{mn}^2\overline{A}_{2n}+\overline{G}_{mn}^3\overline{A}_{3n})\sin m\theta=0,\\
&\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(\overline{H}_{mn}^1\overline{A}_{1n}+\overline{H}_{mn}^2\overline{A}_{2n}+\overline{H}_{mn}^3\overline{A}_{3n}+\overline{H}_{mn}^4\overline{A}_{4n})\times\sin m\theta=0,
\end{aligned}\tag{20}$$

where

$$\begin{aligned}
E_{mn}^j&=(2\varepsilon_n/\pi)\sum_{i=1}^{I}\int_{\theta_{i-1}}^{\theta_i}e_n^j(R_i,\theta)\cos m\theta d\theta,\\
F_{mn}^j&=(2\varepsilon_n/\pi)\sum_{i=1}^{I}\int_{\theta_{i-1}}^{\theta_i}f_n^j(R_i,\theta)\sin m\theta d\theta,\\
G_{mn}^j&=(2\varepsilon_n/\pi)\sum_{i=1}^{I}\int_{\theta_{i-1}}^{\theta_i}g_n^j(R_i,\theta)\cos m\theta\, d\theta,\\
H_{mn}^j&=(2\varepsilon_n/\pi)\sum_{i=1}^{I}\int_{\theta_{i-1}}^{\theta_i}h_n^j(R_i,\theta)\cos m\theta\, d\theta,
\end{aligned}$$

TABLE I. Non-dimensional wave numbers (s) of longitudinal modes of elliptic cylinder.

a/b		1.0				1.1		2.0	
		Present method		Exact method		Present method		Present method	
Ω	Free cylinder without fluid	$R(\varsigma)$	$I(\varsigma)$	$R(\varsigma)$	$I(\varsigma)$	$R(\varsigma)$	$I(\varsigma)$	$R(\varsigma)$	$I(\varsigma)$
0.1	0.054 01	6.3459×10^{-2}	3.8295×10^{-6}	6.3459×10^{-2}	3.8295×10^{-6}	6.3268×10^{-2}	3.6511×10^{-6}	6.2781×10^{-2}	2.8182×10^{-6}
0.3	0.165 01	1.9051×10^{-1}	7.6009×10^{-5}	1.9051×10^{-1}	7.6011×10^{-5}	1.8994×10^{-1}	7.1335×10^{-5}	1.8855×10^{-1}	5.0620×10^{-5}
0.5	0.275 01	3.1789×10^{-1}	2.6626×10^{-4}	3.1789×10^{-1}	2.6628×10^{-4}	3.1696×10^{-1}	2.4785×10^{-4}	3.1493×10^{-1}	1.7039×10^{-4}
0.7	0.385 04	4.4583×10^{-1}	5.8554×10^{-4}	4.4583×10^{-1}	5.8560×10^{-4}	4.4456×10^{-1}	5.4329×10^{-4}	4.4238×10^{-1}	3.7345×10^{-4}
1.0	0.550 82	6.3936×10^{-1}	1.3423×10^{-3}	6.3937×10^{-1}	1.3426×10^{-3}	6.3770×10^{-1}	1.2461×10^{-3}	6.3697×10^{-1}	8.9095×10^{-4}

$$\overline{E}^{j}_{mn}=(2\varepsilon_n/\pi)\sum_{i=1}^{I}\int_{\theta_{i-1}}^{\theta_i}\overline{e}^{j}_{n}(R_i,\theta)\sin m\theta\, d\theta,$$

$$\overline{F}^{j}_{mn}=(2\varepsilon_n/\pi)\sum_{i=1}^{I}\int_{\theta_{i-1}}^{\theta_i}\overline{f}^{j}_{n}(R_i,\theta)\cos m\theta\, d\theta,$$

$$\overline{G}^{j}_{mn}=(2\varepsilon_n/\pi)\sum_{i=1}^{I}\int_{\theta_{i-1}}^{\theta_i}\overline{g}^{j}_{n}(R_i,\theta)\sin m\theta\, d\theta,$$

$$\overline{H}^{j}_{mn}=(2\varepsilon_n/\pi)\sum_{i=1}^{I}\int_{\theta_{i-1}}^{\theta_i}\overline{h}^{j}_{n}(R_i,\theta)\sin m\theta\, d\theta,$$

where $j=1,\ 2,\ 3$ and 4, $\varepsilon_m=1/2$ for $m=0$, and $\varepsilon_m=1$ for $m\geqslant 1$, I is the number of segments and R_i is the coordinate r at the boundary. The frequency equations are obtained from Eq. (19) and Eq. (20) for symmetric and antisymmetric modes by equating the determinants of the coefficients of the amplitudes A_{jn} and $\overline{A}_{jn}$ to zero.

III. NUMERICAL ANALYSIS

The numerical analysis of the frequency equation is carried out for elliptic and cardioidal cross-sectional cylinders immersed in fluid. The material properties used for the computation are as follows: for solid the Poisson ratio $\nu=0.3$, density $\rho=7849$ kg/m^3 and the Young's modulus $E=2.139\times10^{11}$ N/m^2 and for fluid the density $\rho^f=1000$ kg/m^3 and phase velocity $c=1500$ m/s.

The geometrical relations for the elliptic cross section are taken from Eq. (11) of Ref. 2 and they are used directly for the calculation. The dimensionless wave numbers which are complex in nature are computed for $0<\Omega\leqslant 1.0$ for different aspect ratios $a/b=1.0$, 1.1, 1.3, 1.7 and 2.0 using secant method (applicable for complex roots[9]). In this problem, three kinds of basic independent modes of wave propagation are considered, which are namely longitudinal and two flexural modes (symmetric and antisymmetric).

The frequency equation for the solid circular cylinder obtained by the exact method is given in Eq. (B1). This equation is solved numerically for (i) vibration of an isotropic solid circular cylinder without fluid and (ii) vibration of an isotropic solid circular cylinder immersed in fluid. The results obtained by this method are used to compare the results of the present method. The wave numbers computed for both cases are listed in the second and third columns of Table I. The wave numbers, which are real, are obtained for the solid without fluid and the complex wave numbers are obtained for the solid immersed in fluid. In tables $R(\varsigma)$ represents the real part of the wave number and $I(\varsigma)$ represents imaginary part of the wave number. Comparing the results of the free cylinder without fluid with the $R(\varsigma)$ obtained for the cylinder immersed in fluid, since the values of $I(\varsigma)$ are found to be very small, it is observed that the values of cylinder immersed in fluid have increased slightly from the results of the free cylinder without fluid. Also the exact results for a circular cylinder immersed in fluid are computed for the flexural vibration and are given in column 2 of Table II.

A. Longitudinal mode

In case of longitudinal mode, the cross-section vibrates along the axis of the cylinder, so that the vibrational displacements in the cross-section are symmetrical about both the major and the minor axes. Therefore the terms n and m are chosen as 0,2,4,6,... in Eq. (19) for the numerical calculation. The elliptical curve in the range $\theta=0$ and $\theta=\pi$ is divided into 20 segments, such that the distance between any 2 segments is negligible and the integration for a segment is performed numerically by use of the five point Gauss formula. The wave numbers are computed by fixing the frequencies for the longitudinal mode and are presented in

TABLE II. Nondimensional wave numbers (s) of flexural modes of elliptic cylinder.

a/b	1.0				1.1		2.0	
	Present method		Exact method		Present method		Present method	
Ω	$R(\varsigma)$	$I(\varsigma)$	$R(\varsigma)$	$I(\varsigma)$	$R(\varsigma)$	$I(\varsigma)$	$R(\varsigma)$	$I(\varsigma)$
0.1	6.6676×10^{-1}	4.8212×10^{-1}	6.6676×10^{-1}	4.8212×10^{-1}	5.8512×10^{-1}	4.3814×10^{-1}	2.7626×10^{-1}	2.0917×10^{-1}
0.3	8.6109×10^{-1}	3.7146×10^{-1}	8.6109×10^{-1}	3.7146×10^{-1}	8.0111×10^{-1}	3.2021×10^{-1}	6.4493×10^{-1}	9.4513×10^{-2}
0.5	1.1974	2.7168×10^{-1}	1.1974	2.7168×10^{-1}	1.1582	2.2588×10^{-1}	1.0741	5.7008×10^{-2}
0.7	1.5872	2.0733×10^{-1}	1.5872	2.0733×10^{-1}	1.5602	1.6916×10^{-1}	1.5065	4.0584×10^{-2}
1.0	2.2059	1.4931×10^{-1}	2.2059	1.4931×10^{-1}	2.1886	1.2035×10^{-1}	2.1550	2.8284×10^{-2}

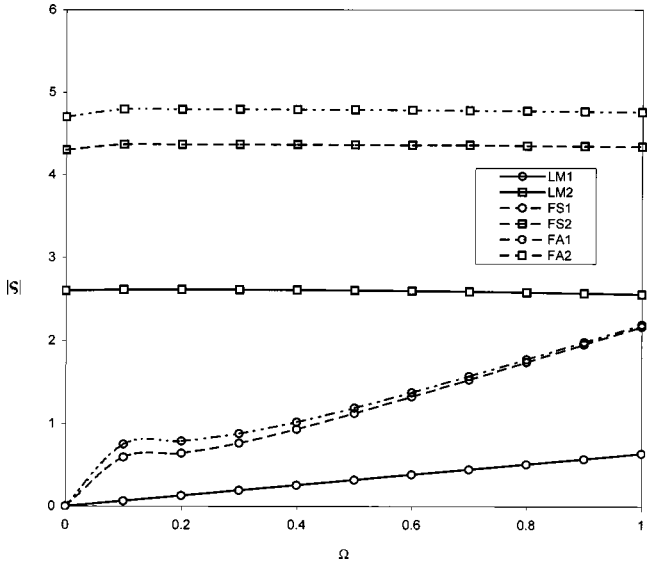


FIG. 1. Nondimensional frequency Ω versus dimensionless wave number $|s|$ of an elliptic cylinder for $(a/b)=1.3$.

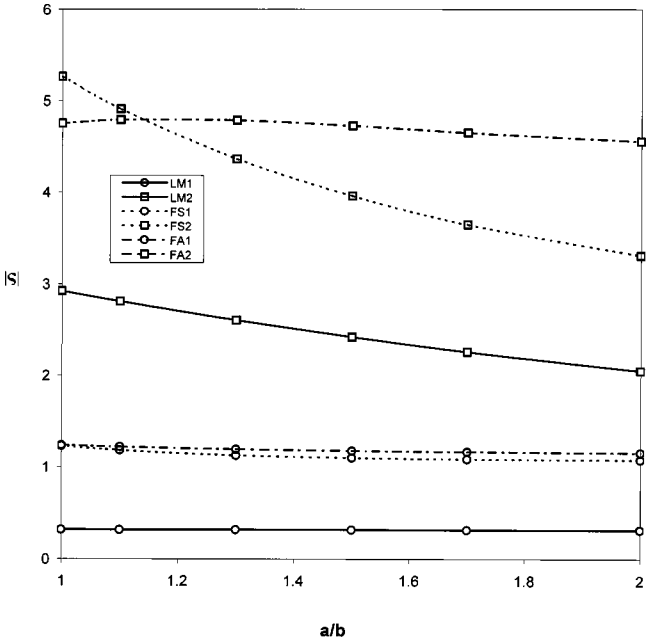


FIG. 2. Aspect ratios (a/b) versus dimensionless wave number $|s|$ of an elliptic cylinder for nondimensional frequency $\Omega=0.5$.

Table I. The convergence of the Fourier series is good and up to three terms $(n,m=0,2,4)$ have been included for computation. It can be noted that the results given in Table I for $a/b=1.0$ by the present method almost coincide with the exact results. The fundamental modes and the second modes of longitudinal motion given in Table I exhibit the proper physical behavior. Also the results for the aspect ratios $a/b=1.1$ and 2.0 are presented in Table I.

B. Flexural mode

In this case two kinds of flexural modes are considered. In the first case, the vibration displacements are antisymmetrical about the major axis and symmetrical about the minor axis. Hence the frequency equation may be obtained from Eq. (20) by choosing $n,m=1,3,5,7,...$ The present results are in very good agreement with exact results. Also the results of the elliptic cylinders for $a/b=1.1$ and 2.0 are given in Table II.

The results of longitudinal and two flexural modes are plotted in Fig. 1 for the dimensionless frequencies Ω versus absolute value of wave numbers $|s|$. The notation used in Fig. 1 like LM, FS and FA represent the longitudinal mode, the flexural (symmetric) mode and the flexural (antisymmetric) mode respectively, and 1 refers the first mode, and 2 refers the second mode. In Fig. 1, the first two modes of longitudinal and flexural (both symmetric and antisymmetric) vibrations are presented for $a/b=1.3$. It is observed that the first mode of all vibrations disperse well compared to the second mode. Also the plot of aspect ratios (a/b) vs $|s|$ is given in Fig. 2. The notations used in Fig. 2 have the same meaning as in Fig. 1.

Also the same problem is solved for a cardioidal cross-sectional cylinder immersed in fluid using the geometrical relations given in Eqs. (24) and (26) of Ref. 8, which are functions depending on a parameter s. This parameter s represents a circle when $s=0$ and represents a cardioid when $s=1/2$. First the parameter s is taken as 0 and verified the results with the results of the elliptic cylinder given in Table I for $a/b=1.0$. Since a cardioid is symmetric about only one axis, the longitudinal and flexural modes (symmetric) are carried out by choosing $n,m=0,1,2,3,...$ in Eq. (19) and the flexural modes (antisymmetric) are obtained by choosing n, $m=1,2,3,...$ in Eq. (20). The dimensionless frequencies Ω versus absolute values of the wave number $|s|$ are plotted in both the cases in Fig. 3 for $s=0.5$. In Fig. 3, LM denotes the longitudinal mode and FM denotes the flexural mode and 1 refers the first mode and 2 refers the second mode.

Numerical results can be obtained directly for the same frequency equation by substituting geometric values of the boundary of any cross-section analytically or numerically with satisfactory convergence.

IV. CONCLUSIONS

In this paper, the wave propagation in a solid cylinder of arbitrary cross-section immersed in fluid is studied. The fre-

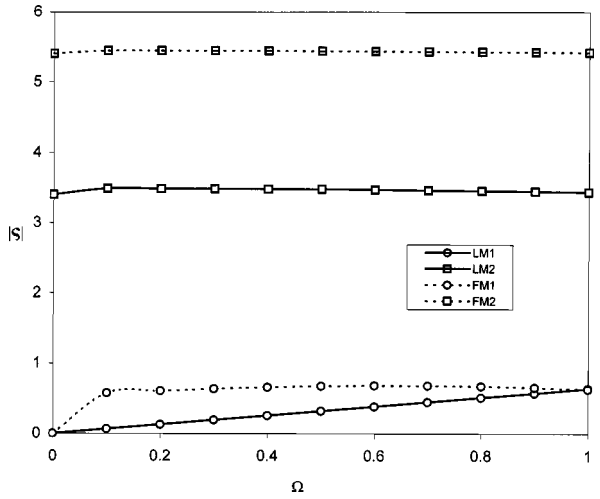


FIG. 3. Nondimensional frequency Ω versus dimensionless wave number $|s|$ of cardioidal cross-section.

quency equation is obtained using the Fourier expansion collocation method. The longitudinal and flexural vibrations are considered and the numerical results are calculated for elliptical cross-section and are tabulated. The dispersion curves are plotted for elliptic cylinder with $(a/b)=1.3$ and 1.7. Similarly the dispersion curves are drawn for cardioidal cross-section also. The results of any other cross-section may be obtained by using the corresponding geometrical relations in the frequency equation given in Eqs. (19) and (20).

ACKNOWLEDGMENTS

The authors wish to thank heartily Professor Peter B. Nagy, Department of Aerospace Engineering and Engineering Mechanics, University of Cincinnati, Ohio, and the reviewers for their valuable comments for the improvement of this work and one of the authors, Thiru. P. Ponnusamy, wishes to acknowledge the University Grants Commission, Hyderabad, for funding to undertake this research work, and the Directorate of Collegiate Education, Tamil Nadu, for the permission rendered toward the same. His gratitude also extends to Anna University, Chennai for providing with the facilities to take up this work.

APPENDIX A

The equations for $e_n^1\tilde{h}_n^4$ referred in Eq. (18) are as follows:

$$
\begin{aligned}
e_n^j &= 2\{n(n-1)J_n(\alpha_j ax)+(\alpha_j ax)J_{n+1}(\alpha_j ax)\}\\
&\quad\times\cos 2(\theta-\gamma_l)\cos n\theta - x^2\{(\alpha_j a)^2(2+\overline{\lambda})\\
&\quad\times\cos^2(\theta-\gamma_l)+\overline{\lambda}d_j\varsigma\}\cos n\theta + 2n\{(n-1)J_n(\alpha_j ax)\\
&\quad-(\alpha_j ax)J_{n+1}(\alpha_j ax)\}\sin n\theta\,\sin 2(\theta-\gamma_l),\\
\overline{e_n^j} &= 2\{n(n-1)J_n(\alpha_j ax)+(\alpha_j ax)J_{n+1}(\alpha_j ax)\}\\
&\quad\times\cos 2(\theta-\gamma_l)\sin n\theta - x^2\{(\alpha_j a)^2(2+\overline{\lambda})\\
&\quad\times\cos^2(\theta-\gamma_l)+\overline{\lambda}d_j\varsigma\}\sin n\theta - 2n\{(n-1)J_n(\alpha_j ax)\\
&\quad-(\alpha_j ax)J_{n+1}(\alpha_j ax)\}\cos n\theta\,\sin 2(\theta-\gamma_l),\\
e_n^3 &= 2n\{(n-1)J_n(\beta ax)-(\beta ax)J_{n+1}(\beta ax)\}\cos 2(\theta-\gamma_l)\\
&\quad\times\cos n\theta + 2\{[n(n-1)-(\beta ax)^2]J_n(\beta ax)\\
&\quad+(\beta ax)J_{n+1}(\beta ax)\}\sin 2(\theta-\gamma_l)\sin n\theta,\\
\overline{e_n^3} &= 2n\{(n-1)J_n(\beta ax)-(\beta ax)J_{n+1}(\beta ax)\}\cos 2(\theta-\gamma_l)\\
&\quad\times\sin n\theta - 2\{[n(n-1)-(\beta ax)^2]J_n(\beta ax)\\
&\quad+(\beta ax)J_{n+1}(\beta ax)\}\sin 2(\theta-\gamma_l)\cos n\theta,\\
e_n^4 &= \Omega^2\overline{\rho}H_n^1(\delta ax)\cos n\theta,\\
\overline{e_n^4} &= \Omega^2\overline{\rho}H_n^1(\delta ax)\sin n\theta,\\
f_n^j &= 2\{[n(n-1)-(\alpha_j ax)^2]J_n(\alpha_j ax)\\
&\quad+(\alpha_j ax)J_{n+1}(\alpha_j ax)\}\sin 2(\theta-\gamma_l)\cos n\theta\\
&\quad+2n\{(\alpha_j ax)J_{n+1}(\alpha_j ax)-(n-1)J_n(\alpha_j ax)\}\\
&\quad\times\cos 2(\theta-\gamma_l)\sin n\theta,\\
\overline{f_n^j} &= 2\{[n(n-1)-(\alpha_j ax)^2]J_n(\alpha_j ax)\\
&\quad+(\alpha_j ax)J_{n+1}(\alpha_j ax)\}\sin 2(\theta-\gamma_l)\sin n\theta\\
&\quad-2n\{(\alpha_j ax)J_{n+1}(\alpha_j ax)\\
&\quad-(n-1)J_n(\alpha_j ax)\}\cos 2(\theta-\gamma_l)\cos n\theta,\\
f_n^3 &= 2n\{(n-1)J_n(\beta ax)-(\beta ax)J_{n+1}(\beta ax)\}\sin 2(\theta-\gamma_l)\\
&\quad\times\cos n\theta - 2\{[n(n-1)-(\beta ax)^2]J_n(\beta ax)\\
&\quad+(\beta ax)J_{n+1}(\beta ax)\}\cos 2(\theta-\gamma_l)\sin n\theta,\\
\overline{f_n^3} &= 2n\{(n-1)J_n(\beta ax)-(\beta ax)J_{n+1}(\beta ax)\}\sin 2(\theta-\gamma_l)\\
&\quad\times\sin n\theta + 2\{[n(n-1)-(\beta ax)^2]J_n(\beta ax)\\
&\quad+(\beta ax)J_{n+1}(\beta ax)\}\cos 2(\theta-\gamma_l)\cos n\theta,\\
g_n^j &= (\varsigma+d_j)\{nJ_n(\alpha_j ax)\cos(\overline{n-1}\ \theta+\gamma_l)\\
&\quad+(\alpha_j ax)J_{n+1}(\alpha_j ax)\cos(\theta-\gamma_l)\cos n\theta\},\\
\overline{g_n^j} &= (\varsigma+d_j)\{nJ_n(\alpha_j ax)\sin(\overline{n-1}\ \theta+\gamma_l)\\
&\quad-(\alpha_j ax)J_{n+1}(\alpha_j ax)\cos(\theta-\gamma_l)\sin n\theta\},\\
g_n^3 &= \varsigma\{nJ_n(\beta ax)\cos(\overline{n-1}\ \theta+\gamma_l)\\
&\quad-(\beta ax)J_{n+1}(\beta ax)\sin n\theta\,\sin(\theta-\gamma_l)\},\\
\overline{g_n^3} &= \varsigma\{nJ_n(\beta ax)\sin(\overline{n-1}\ \theta+\gamma_l)\\
&\quad+(\beta ax)J_{n+1}(\beta ax)\cos n\theta\,\sin(\theta-\gamma_l)\},\\
h_n^j &= \{nJ_n(\alpha_j ax)-(\alpha_j ax)J_{n+1}(\alpha_j ax)\}\cos n\theta,\\
\overline{h_n^j} &= \{nJ_n(\alpha_j ax)-(\alpha_j ax)J_{n+1}(\alpha_j ax)\}\sin n\theta,\\
h_n^3 &= nJ_n(\beta ax)\cos n\theta,\\
\overline{h_n^3} &= nJ_n(\beta ax)\sin n\theta,\\
h_n^4 &= \Omega^2\overline{\rho}\{nH_n^1(\delta ax)-(\delta ax)H_{n+1}^1(\delta ax)\}\cos n\theta,\\
\overline{h_n^4} &= \Omega^2\overline{\rho}\{nH_n^1(\delta ax)-(\delta ax)H_{n+1}^1(\delta ax)\}\sin n\theta.
\end{aligned}
\tag{A1}
$$

APPENDIX B: SOLID CIRCULAR CYLINDER IMMERSED IN FLUID

For a typical case of a solid circular cylindrical rod immersed in fluid, the frequency equation is obtained analytically[6] and is given as follows:

$$|L|=0, \tag{B1}$$

where L is the matrix with elements l_{pq}, $p=1,2,3$, $q=1,2,3$ and are given by

$$
\begin{aligned}
l_{1q} &= \{2n(n-1)-(2+\overline{\lambda})(\alpha_q a)^2-\overline{\lambda\varsigma d_q}\}J_n(\alpha_q a)\\
&\quad+2(\alpha_q a)J_{n+1}(\alpha_q a),\quad q=1,2,\\
l_{13} &= n\{(n-1)J_n(\beta a)-(\beta a)J_{n+1}(\beta a)\},\\
l_{14} &= (\delta a)H_{n+1}^1(\delta a)-nH_n^1(\delta a),
\end{aligned}
$$

$$
\begin{aligned}
l_{2q} &= 2n\{(\alpha_q a)J_{n+1}(\alpha_q a) - (n-1)J_{n+1}(\alpha_q a)\}, \\
&\quad q = 1,2, \\
l_{23} &= \{[(\beta a)^2 - 2n(n-1)]\}J_n(\beta a) \\
&\quad - 2(\alpha_3 a)J_{n+1}(\beta a), \\
l_{24} &= 0, \\
l_{3q} &= (\varsigma + d_q)\{nJ_n(\alpha_q a) - (\alpha_q a)J_{n+1}(\alpha_q a)\}, \quad q = 1,2, \\
l_{33} &= n\varsigma J_n(\beta a), \\
l_{34} &= 0, \\
l_{4q} &= nJ_n(\alpha_q a) - (\alpha_q a)J_{n+1}(\alpha_q a), \quad q = 1,2, \\
l_{43} &= 0, \\
l_{44} &= \Omega^2 \overline{\rho} H_n^1(\delta a).
\end{aligned}
\tag{B2}
$$

[1] D. C. Gazis, "Three-dimensional investigation of the propagation of waves in hollow circular cylinders, I. Analytical formulation; II. Numerical results," J. Acoust. Soc. Am. **31**, 568–578 (1959).

[2] K. Nagaya, "Stress wave propagation in a bar of arbitrary cross-section," ASME J. Appl. Mech. **49**, 157–164 (1982).

[3] K. Nagaya, "Dispersion of elastic waves in bars with polygonal cross-section," J. Acoust. Soc. Am. **70**, 763–770 (1981).

[4] H. S. Paul and M. Venkatesan, "Wave propagation in a piezoelectric solid cylinder of arbitrary cross-section," J. Acoust. Soc. Am. **82**, 2013–2020 (1987).

[5] K. Sinha, J. Plona, Sergio Kostek, and Shu-Kong Chang, "Axisymmetric wave propagation in fluid-loaded cylindrical shells. I: Theory; II. Theory versus experiment," J. Acoust. Soc. Am. **92**, 1132–1155 (1992).

[6] J. Berliner and R. Solecki, "Wave propagation in fluid-loaded, transversely isotropic cylinders. Part. I. Analytical formulation; Part II. Numerical results," J. Acoust. Soc. Am. **99**, 1841–1853 (1996).

[7] J. D. Achenbach, *Wave Propagation in Elastic Solids* (Elsevier, New York, 1973).

[8] K. Nagaya, "Direct method on determination of eigen frequencies of arbitrarily shaped plates," ASME J. Vib., Acoust., Stress, Reliab. Des. **105**, 132–136 (1983).

[9] H. M. Antia, *Numerical Methods for Scientists and Engineers* (Hindustan Book Agency, New Delhi, 2002).

A full 3D plane-wave-expansion model for 1-3 piezoelectric composite structures

Mikaël Wilm, Sylvain Ballandras, Vincent Laude, and Thomas Pastureaud
Laboratoire de Physique et Métrologie des Oscillateurs/CNRS UPR 3203, associé à l'Université de Franche-Comté, 32 avenue de l'Observatoire, 25044 Besançon Cedex, France

(Received 29 November 2001; accepted for publication 31 May 2002)

The plane-wave-expansion (PWE) approach dedicated to the simulation of periodic devices has been extended to 1-3 connectivity piezoelectric composite structures. The case of simple but actual piezoelectric composite structures is addressed, taking piezoelectricity, acoustic losses, and electrical excitation conditions rigorously into account. The material distribution is represented by using a bidimensional Fourier series and the electromechanical response is simulated using a Bloch–Floquet expansion together with the Fahmy–Adler formulation of the Christoffel problem. Application of the model to 1-3 connectivity piezoelectric composites is reported and compared to previously published analyses of this problem. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1496081]

PACS numbers: 43.38.Ar, 43.38.Fx, 43.20.Ks, 43.40.At [SLE]

I. INTRODUCTION

Piezoelectric composite transducers have been developed for medical ultrasound imaging and nondestructive evaluation to overcome the limitations of standard 1D probes. The main advantages of piezoelectric composite structures are the optimization of their electromechanical coupling factor, their low specific acoustic impedance, and their capability to be shaped on curved surfaces (for more about piezoelectric composite, see, e.g., Refs. 1–3.

Plane-wave-expansion (PWE) models[4,5] have been developed to address the description of structures exhibiting periodic in-plane or bulk nonhomogeneity. Such an approach represents an alternative to finite element computations, easy to implement and providing complementary information about the capability of any structure to guide elastic waves. Unfortunately, most of the proposed developments do not take into account piezoelectricity, acoustic losses, or simply the finite thickness of actual devices. However, the possibility to simulate, for instance, a semi-infinite substrate using a PWE approach was recently demonstrated.[6] In the present work, the PWE model for periodic structures is extended to piezoelectric composite materials. These are generally composed of a 1D (2-2 connectivity) or 2D (1-3 connectivity) array of piezoelectric elements mixed with a polymer material. Addressing the problem of periodic arrays using the proposed method allows one to compute the general properties of the device by only treating one period of the structure. The material distribution is represented using Fourier series and the electromechanical vibration using a Bloch–Floquet expansion. The Fahmy–Adler formulation of the Christoffel problem[7] has been adapted to obtain the modal distribution along the thickness of the piezoelectric composite plate.

Assessment calculations are performed for academic problems, illustrating the efficiency of the proposed approach. For instance, the identification of plate modes in a homogeneous piezoelectric material (namely the Z cut plate of quartz) has been checked. The application of the model to 2-2 and 1-3 connectivity piezoelectric composites is reported and compared to previous analyses of the same problem generally based on simpler modeling approaches or on finite element computations. The possibility to compute the harmonic admittance is emphasized, giving access to complementary data such as piezoelectric coupling, propagation losses, and vibration shapes. As a conclusion, the future extensions of the model are discussed in order to build up a comprehensive tool that is able to accurately simulate piezoelectric composite structures radiating in different media.

II. PRINCIPLES OF THE MODEL

A. Basic definitions

Figure 1 shows the general geometry of the considered structures, respectively 2-2 and 1-3 connectivity piezocomposite structures and their elementary cells. The thickness of the plate is along x_3 and wave propagation occurs along x_1 and/or x_2 depending on the addressed problem. The plate is assumed periodic along x_1 for 2D problems, and two periodicities p_1 and p_2 in the (x_1,x_2) plane are considered for 3D problems.

According to the Floquet theorem, all fields $h(\boldsymbol{r},t)$, such as displacements or stresses, propagating in such periodic structures can be expanded as infinite series

$$h(\boldsymbol{r},t)=\sum_{\boldsymbol{G}} h_{\boldsymbol{K}+\boldsymbol{G}} e^{j(\omega t-\boldsymbol{K}\cdot\boldsymbol{r}-\boldsymbol{G}\cdot\boldsymbol{r})}, \tag{1}$$

where $\boldsymbol{r}=(x_1,x_2,x_3)^T$ and $\boldsymbol{G}=(2\pi m/p_1,2\pi n/p_2,0)^T$. The upperscript ^{T} stands for transposition. $\boldsymbol{G}$ are the vectors of the reciprocal lattice and $\boldsymbol{K}$ is the wave vector. For 2D problems, K_2 and G_2 are fixed to zero. The periodicity of the structure is used to expand the material constants as Fourier series

$$\alpha(\boldsymbol{r})=\sum_{\boldsymbol{G}} \alpha_{\boldsymbol{G}} e^{-j\boldsymbol{G}\cdot\boldsymbol{r}}, \tag{2}$$

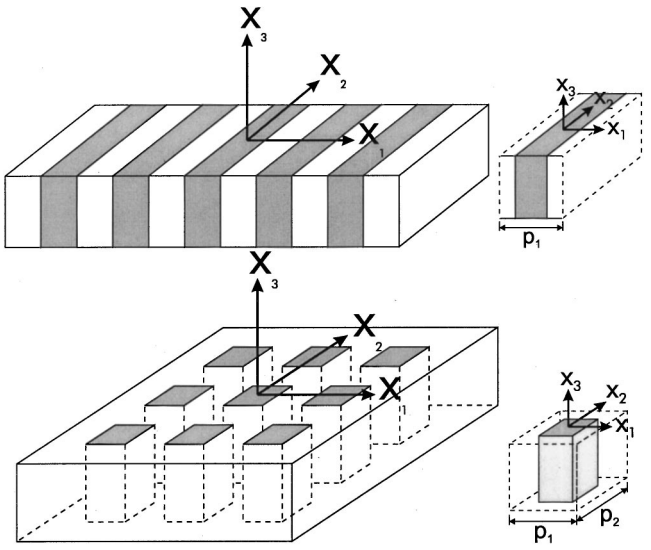


FIG. 1. General geometry of the addressed problem: respectively 2-2 and 1-3 connectivity composites and their elementary cells.

with $\alpha=\{\rho,c_{ijkl},e_{ijk},\epsilon_{ij}\}$. The terms $\alpha_{\boldsymbol{G}}$ are easily calculated, especially for different cross section shapes of 1-3 connectivity composite structures.[5]

The usual constitutive relations of piezoelectricity are then considered, together with the fundamental equation of dynamics and Poisson's condition for dielectric media:

$$T_{ij}=c_{ijkl}u_{k,l}+e_{lij}\phi_{,l}, \tag{3}$$

$$D_i=e_{ikl}u_{k,l}-\epsilon_{il}\phi_{,l}, \tag{4}$$

$$\rho\frac{\partial^2 u_j}{\partial t^2}=T_{ij,i}, \tag{5}$$

$$D_{i,i}=0. \tag{6}$$

According to Ref. 8, one can define a generalized displacement field $\boldsymbol{u}$ in which u_4 represents the electrical potential ϕ, and generalized stress vectors $\tilde{\boldsymbol{t}}_i=(T_{i1},T_{i2},T_{i3},D_i)^T$.

Inserting Eqs. (1) and (2) into Eqs. (3) and (4) yields the following expressions for the stress and electrical displacement:

$$\sum_{\boldsymbol{G}'} T_{ij\boldsymbol{K}+\boldsymbol{G}'}e^{-j\boldsymbol{G}'\cdot\boldsymbol{r}}=\sum_{\boldsymbol{G}}\sum_{\boldsymbol{G}'}[-j(K_l+G_l')\times(c_{ijkl\boldsymbol{G}}u_{k\boldsymbol{K}+\boldsymbol{G}'}+e_{lij\boldsymbol{G}}u_{4\boldsymbol{K}+\boldsymbol{G}'})e^{-j(\boldsymbol{G}+\boldsymbol{G}')\cdot\boldsymbol{r}}], \tag{7}$$

$$\sum_{\boldsymbol{G}'} D_{i\boldsymbol{K}+\boldsymbol{G}'}e^{-j\boldsymbol{G}'\cdot\boldsymbol{r}}=\sum_{\boldsymbol{G}}\sum_{\boldsymbol{G}'}[-j(K_l+G_l')(e_{ikl\boldsymbol{G}}u_{k\boldsymbol{K}+\boldsymbol{G}'}-\epsilon_{il\boldsymbol{G}}u_{4\boldsymbol{K}+\boldsymbol{G}'})e^{-j(\boldsymbol{G}+\boldsymbol{G}')\cdot\boldsymbol{r}}], \tag{8}$$

where the $'$ is introduced to differentiate summations over space harmonics relative to the dynamic fields from those relative to static distributions of material properties.

B. Matrix formulation of the problem

The orthogonality property of Fourier series components is then used to eliminate the spatial dependence of the stress and electrical displacement. The change of variable ($\boldsymbol{G}''=\boldsymbol{G}+\boldsymbol{G}'$) is first performed, then both expressions (7) and (8) are multiplied by $e^{j\boldsymbol{G}\cdot\boldsymbol{r}}$, and finally the resulting form is integrated over one period, yielding the following definition of $T_{ij\boldsymbol{K}+\boldsymbol{G}}$ and $D_{i\boldsymbol{K}+\boldsymbol{G}}$:

$$T_{ij\boldsymbol{K}+\boldsymbol{G}}=\sum_{\boldsymbol{G}'}[-j(K_l+G_l')(c_{ijkl\boldsymbol{G}-\boldsymbol{G}'}u_{k\boldsymbol{K}+\boldsymbol{G}'}+e_{lij\boldsymbol{G}-\boldsymbol{G}'}u_{4\boldsymbol{K}+\boldsymbol{G}'})], \tag{9}$$

$$D_{i\boldsymbol{K}+\boldsymbol{G}}=\sum_{\boldsymbol{G}'}[-j(K_l+G_l')(e_{ikl\boldsymbol{G}-\boldsymbol{G}'}u_{k\boldsymbol{K}+\boldsymbol{G}'}-\epsilon_{il\boldsymbol{G}-\boldsymbol{G}'}u_{4\boldsymbol{K}+\boldsymbol{G}'})]. \tag{10}$$

This operation is equivalent to an orthogonal projection of the operators on the trigonometric functional basis using the classical scalar product associated to Fourier theory (see, for instance, Ref. 9). The generalized stress vectors are then written

$$j\tilde{\boldsymbol{t}}_{i\boldsymbol{K}+\boldsymbol{G}}=\sum_{\boldsymbol{G}'}(K_l+G_l')A_{il\boldsymbol{G}-\boldsymbol{G}'}\boldsymbol{u}_{\boldsymbol{K}+\boldsymbol{G}'}\quad(i=1,2,3), \tag{11}$$

where

$$A_{il\boldsymbol{G}}(j,k)=c_{ijkl\boldsymbol{G}},\quad A_{il\boldsymbol{G}}(j,4)=e_{lij\boldsymbol{G}},\quad A_{il\boldsymbol{G}}(4,k)=e_{ikl\boldsymbol{G}},\quad A_{il\boldsymbol{G}}(4,4)=-\epsilon_{il\boldsymbol{G}}, \tag{12}$$

with $(j,k)\in[1,3]^2$.

Inserting Eqs. (1) and (2) into Eqs. (5) and (6) yields the following expressions for the propagation equation and the Poisson's condition:

$$\sum_{\boldsymbol{G}'}-j(K_i+G_i')T_{ij\boldsymbol{K}+\boldsymbol{G}'}e^{-j\boldsymbol{G}'\cdot\boldsymbol{r}}=\sum_{\boldsymbol{G}}\sum_{\boldsymbol{G}'}[\rho_{\boldsymbol{G}}(j\omega)^2u_{j\boldsymbol{K}+\boldsymbol{G}'}e^{-j(\boldsymbol{G}+\boldsymbol{G}')\cdot\boldsymbol{r}}], \tag{13}$$

$$\sum_{\boldsymbol{G}'}-j(K_i+G_i')D_{i\boldsymbol{K}+\boldsymbol{G}'}e^{-j\boldsymbol{G}'\cdot\boldsymbol{r}}=0. \tag{14}$$

Applying the same procedure to the stress and the electrical displacement provides the following expression for the generalized stress vectors:

$$(K_i+G_i)j\tilde{\boldsymbol{t}}_{i\boldsymbol{K}+\boldsymbol{G}}=\sum_{\boldsymbol{G}'}\omega^2R_{\boldsymbol{G}-\boldsymbol{G}'}\boldsymbol{u}_{\boldsymbol{K}+\boldsymbol{G}'}, \tag{15}$$

where $R_{\boldsymbol{G}-\boldsymbol{G}'}=\rho_{\boldsymbol{G}-\boldsymbol{G}'}\tilde{I}$, with

$$\tilde{I}=\begin{bmatrix}1&&&\\&1&&\\&&1&\\&&&0\end{bmatrix}.$$

In numerical computations, one has to truncate the Fourier series using a number of spatial harmonics large enough to ensure convergence. Assuming N terms in the Fourier expansions (1) and (2), the following vector notations are considered for the generalized stress and displacement fields:

$$\tilde{\boldsymbol{T}}_{i\boldsymbol{K}}=\begin{pmatrix}\tilde{\boldsymbol{t}}_{i\boldsymbol{K}+\boldsymbol{G}^1}\\ \vdots \\ \tilde{\boldsymbol{t}}_{i\boldsymbol{K}+\boldsymbol{G}^N}\end{pmatrix} \quad \text{and} \quad \tilde{\boldsymbol{U}}_{\boldsymbol{K}}=\begin{pmatrix}\boldsymbol{u}_{\boldsymbol{K}+\boldsymbol{G}^1}\\ \vdots \\ \boldsymbol{u}_{\boldsymbol{K}+\boldsymbol{G}^N}\end{pmatrix}. \tag{16}$$

The same approach is applied to matrices A_{ij} and R introduced in Eqs. (11) and (15) yielding the definitions

$$\tilde{A}_{ij}=\begin{bmatrix} A_{ij\boldsymbol{0}} & A_{ij\boldsymbol{G}^1-\boldsymbol{G}^2} & \cdots & A_{ij\boldsymbol{G}^1-\boldsymbol{G}^N} \\ A_{ij\boldsymbol{G}^2-\boldsymbol{G}^1} & A_{ij\boldsymbol{0}} & \cdots & A_{ij\boldsymbol{G}^2-\boldsymbol{G}^N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{ij\boldsymbol{G}^N-\boldsymbol{G}^1} & A_{ij\boldsymbol{G}^N-\boldsymbol{G}^2} & \cdots & A_{ij\boldsymbol{0}} \end{bmatrix} \tag{17}$$

and

$$\tilde{R}=\begin{bmatrix} \rho_{\boldsymbol{0}}\tilde{I} & \rho_{\boldsymbol{G}^1-\boldsymbol{G}^2}\tilde{I} & \cdots & \rho_{\boldsymbol{G}^1-\boldsymbol{G}^N}\tilde{I} \\ \rho_{\boldsymbol{G}^2-\boldsymbol{G}^1}\tilde{I} & \rho_{\boldsymbol{0}}\tilde{I} & \cdots & \rho_{\boldsymbol{G}^2-\boldsymbol{G}_N}\tilde{I} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\boldsymbol{G}^N-\boldsymbol{G}^1}\tilde{I} & \rho_{\boldsymbol{G}^N-\boldsymbol{G}^2}\tilde{I} & \cdots & \rho_{\boldsymbol{0}}\tilde{I} \end{bmatrix}. \tag{18}$$

Introducing the diagonal matrix Γ_i $(i=1,2,3)$ whose terms are generated by first-order spatial derivatives, Eqs. (11) and (15) can be written in the very compact form

$$j\tilde{\boldsymbol{T}}_i=\tilde{A}_{ij}\Gamma_j\tilde{\boldsymbol{U}} \quad (i=1,2,3), \tag{19}$$

$$\omega^2\tilde{R}\tilde{\boldsymbol{U}}=\Gamma_i(j\tilde{\boldsymbol{T}}_i), \tag{20}$$

with

$$\Gamma_i=\begin{bmatrix} (K_i+G_i^1)I_d & & & 0 \\ & (K_i+G_i^2)I_d & & \\ & & \ddots & \\ 0 & & & (K_i+G_i^N)I_d \end{bmatrix}. \tag{21}$$

C. Calculation of the vibration distribution along x_3

Equations (19) and (20) do not exactly yield a standard eigenvalue problem as usually defined.[7,9] However, due to the large dimension of the matrices in Eqs. (19) and (20), especially when addressing the bi-periodic problem (8 equations times N, the number of spatial harmonics in the Floquet developments along x_1 times those along x_2), one should avoid any matrix inversion. The computation time and numerical difficulties related to this operation may induce major drawbacks to the proposed method. Similarly to the procedure outlined by Peach for SAW propagation on homogeneous substrates,[10] an extended definition of eigenvalue problems is considered here, for an efficient treatment of Eqs. (19) and (20) without any matrix inversion.

The structure is assumed to be nonperiodic along the x_3 axis. As a consequence, G_3 is set to zero and K_3 depends on the other parameters of the model, namely, K_1, K_2, and ω. Writing the system of Eqs. (19) and (20), in order to identify K_3 as the eigenvalue and $(\tilde{\boldsymbol{U}}_{\boldsymbol{K}}, j\tilde{\boldsymbol{T}}_{3\boldsymbol{K}})^T$ as the eigenvector of the problem, yields after some algebra

$$\begin{bmatrix} \omega^2\tilde{R}-B & 0 \\ -C2 & I_d \end{bmatrix}\begin{pmatrix} \tilde{\boldsymbol{U}}_{\boldsymbol{K}} \\ j\tilde{\boldsymbol{T}}_{3\boldsymbol{K}} \end{pmatrix}=K_3\begin{bmatrix} C1 & I_d \\ D & 0 \end{bmatrix}\begin{pmatrix} \tilde{\boldsymbol{U}}_{\boldsymbol{K}} \\ j\tilde{\boldsymbol{T}}_{3\boldsymbol{K}} \end{pmatrix}, \tag{22}$$

with

$$\begin{aligned} &B=\sum_{i,j=1,2}\Gamma_i\tilde{A}_{ij}\Gamma_j, \quad C1=\sum_{i=1,2}\Gamma_i\tilde{A}_{i3}, \\ &C2=\sum_{j=1,2}\tilde{A}_{3j}\Gamma_j, \quad D=\tilde{A}_{33}, \end{aligned} \tag{23}$$

and where $\tilde{\boldsymbol{T}}_{3\boldsymbol{K}}$ is the generalized normal stress vector.

Solving this system yields $8N$ eigenvalues $K_3^{(r)}$ and eigenvectors, which are used to build the generalized displacement and normal stress fields. Since the eigenvectors are determined for an arbitrary constant value of the independent unknowns of the problem, one introduces the relative amplitudes $A^{(r)}$ (as for surface waves, see, e.g., Refs. 8 and 10). Using a comprehensive set of boundary conditions allows one to determine these relative amplitudes. The final form of the generalized displacement and normal stress fields is written

$$\begin{aligned} \begin{pmatrix} \boldsymbol{u}(\boldsymbol{r},t) \\ \tilde{\boldsymbol{t}}_3(\boldsymbol{r},t) \end{pmatrix} &= e^{j(\omega t-K_1x_1-K_2x_2)} \\ &\times\sum_{l=1}^{N}\left[e^{-j\boldsymbol{G}^l\cdot\boldsymbol{r}}\sum_{r=1}^{8N} A^{(r)}e^{-jK_3^{(r)}x_3}\begin{pmatrix} \boldsymbol{u}_{\boldsymbol{K}+\boldsymbol{G}^l}^{(r)} \\ \tilde{\boldsymbol{t}}_{3\boldsymbol{K}+\boldsymbol{G}^l}^{(r)} \end{pmatrix}\right]. \end{aligned} \tag{24}$$

D. Boundary conditions

Note that the case of bulk waves propagating along the (x_1,x_2) plane in an infinite medium along x_3 can be easily computed using Eq. (22), simply by setting $K_3=0$ and by solving the resulting problem $\omega^2\tilde{R}\tilde{\boldsymbol{U}}=B\tilde{\boldsymbol{U}}$, where ω is computed for each (K_1,K_2) defined in the first Brillouin zone $[0,2\pi/p_{1/2}]$ ($p_{1/2}$ being the periods of the structure).

However, a large variety of boundary conditions can be applied to simulate actual operating conditions of a piezoelectric composite structure. The mechanical boundary conditions require the nullity of stress components normal to surfaces (stress free definite boundaries in air or in a vacuum) or the physical validity of the partial modes that have to be included in the normal mode expansion (24) (semi-infinite medium). In this latter case, one has to select the values of K_3 with respect to this condition. Using a criterion based on the sign of the mean Poynting vector[1,10] of each partial mode (in the case of propagative modes) or on the sign of the imaginary part of each $K_3^{(r)}$ (in the case of evanescent modes), an unambiguous modal selection can be performed. For plate devices, both sides are assumed stress free. Other kinds of mechanical boundary conditions may be applied (for instance, radiation conditions, see Ref. 11), but they have not been implemented yet. The stress free condition is then written independently from the space coordinates

x_1 and x_2, as previously explained for the propagation equation and Poisson's equation, yielding the following expressions:

$$\sum_{r=1}^{8N} A^{(r)} T_{3j\mathbf{K}+\mathbf{G}^i}^{(r)} \mathrm{e}^{-jK_3^{(r)}h} = 0 \quad (i=1,...,N) \quad (j=1,2,3), \tag{25}$$

where $x_3 = h$ at the surface.

From the electrical point of view, in the case of a definite boundary, two kinds of boundary conditions have been considered. The first one is the continuity of the potential and of the normal component of the electrical displacement at the surface. In the air, one can write

$$\Delta \phi^{\mathrm{air}} = 0, \tag{26}$$

$$D_3^{\mathrm{air}} = -\epsilon_0 \frac{\partial \phi^{\mathrm{air}}}{\partial x_3}, \tag{27}$$

where Δ is the Laplacian. The continuity conditions at the surface yield

$$\phi^{\mathrm{substrate}}|_{x_3=h} = \phi^{\mathrm{air}}|_{x_3=h}, \tag{28}$$

$$D_3^{\mathrm{substrate}}|_{x_3=h} - D_3^{\mathrm{air}}|_{x_3=h} = 0. \tag{29}$$

The boundary condition expressions are finally obtained after some algebra as

$$\sum_{r=1}^{8N} A^{(r)} [D_{3\mathbf{K}+\mathbf{G}^i}^{(r)} - \epsilon_0 |\kappa| \phi_{\mathbf{K}+\mathbf{G}^i}^{(r)}] \mathrm{e}^{-jK_3^{(r)}h} = 0 \quad (i=1,...,N), \tag{30}$$

with $|\kappa| = \sqrt{(K_1+G_1)^2 + (K_2+G_2)^2}$.

The second boundary condition considered is setting the surface potential homogeneously to a given value V_0, for instance 0 V for a shorted surface or 1 V for an active electrode. Considering one period and infinitely close electrodes, the potential is written

$$\phi(\mathbf{r},t)|_{x_3=h} = V_0 \mathrm{e}^{j\omega t}, \tag{31}$$

yielding

$$\sum_{i=1}^{N} \left[e^{-j\mathbf{G}^i \cdot \mathbf{r}} \sum_{r=1}^{8N} A^{(r)} \phi_{\mathbf{K}+\mathbf{G}^i}^{(r)} \mathrm{e}^{-jK_3^{(r)}h} \right] = V_0 \mathrm{e}^{j(K_1 x_1 + K_2 x_2)}. \tag{32}$$

Applying the orthogonal projection, one finally obtains the desired boundary conditions

$$\sum_{r=1}^{8N} A^{(r)} \phi_{\mathbf{K}+\mathbf{G}^i}^{(r)} \mathrm{e}^{-jK_3^{(r)}h} = V_0 \left(\mathrm{sinc}\left((K_1+G_1^i)\frac{p_1}{2} \right) \right) \left(\mathrm{sinc}\left((K_2+G_2^i)\frac{p_2}{2} \right) \right) \quad (i=1,...,N). \tag{33}$$

In the case of homogeneous boundary conditions, one has to set up a boundary condition system exhibiting $8N$ equations (the number of independent unknowns). The modes of the structure correspond to the zeros of the determinant of this system, for which nontrivial solutions can then be obtained. For each of these modes, it is possible to compute the relative amplitude $A^{(r)}$ by setting one of them equal to an arbitrary value and then deducing the others by solving the boundary condition system. Thus the actual nature of the corresponding vibration can be identified.

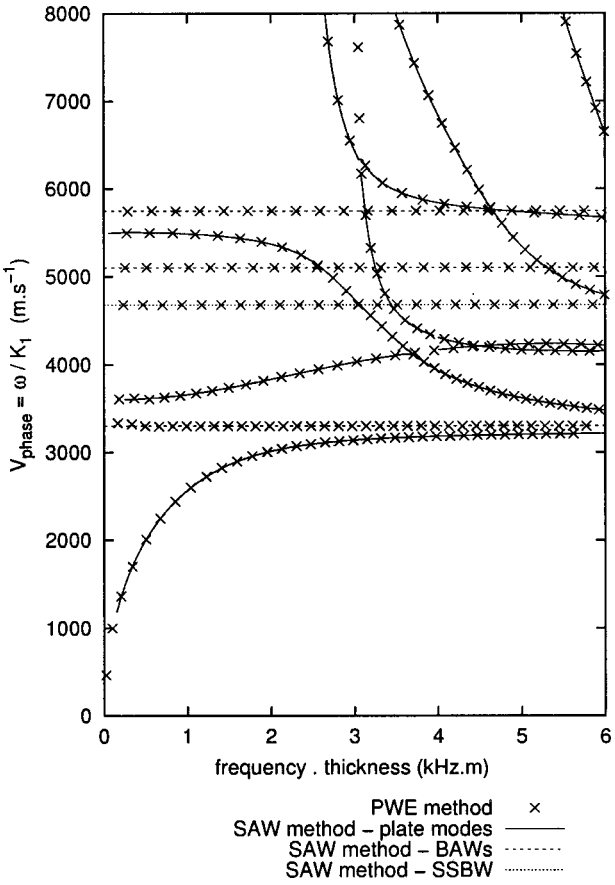


FIG. 2. Comparison between standard Green's function computation (solid and dashed lines) and the proposed approach (cross markers) to identify the acoustic modes of a (ZX) quartz plate. The plate modes (Lamb waves) are found, such as the three BAWs (bulk acoustic waves) and the SSBW (surface skimming bulk wave). The BAWS are the longitudinal, fast shear and slow shear ones at, respectively, 5748, 5103.3, and 3300.5 m s^{-1}. The SSBW celerity is 4678 m s^{-1}. Contrary to the BAWs, its wave vector is not parallel to the surface while its Poynting vector is such. The two isolated points on the top of the figure are in fact in the continuity of a Lamb mode. Indeed, the SAW dedicated method computation is not complete for the corresponding mode because of the difficulty to follow modes which cross each other.

E. Harmonic admittance

Using the above-mentioned electrical boundary conditions allows one to simulate the excitation conditions of piezoelectric composite materials and to derive the harmonic admittance for a single cell by computing the charge distribution on the active electrode. In this approach, one has to consider K_1 and K_2 as excitation parameters governing the excitation potential distribution as follows,

$$V_{mn} = V_0 \mathrm{e}^{-jK_1 m p_1} \mathrm{e}^{-jK_2 n p_2}, \tag{34}$$

where p_1 and p_2 are, respectively, the periodicity along x_1 and x_2 as previously defined (see Fig. 1). For instance, an alternation of the sign of the potential excitation along x_1 is

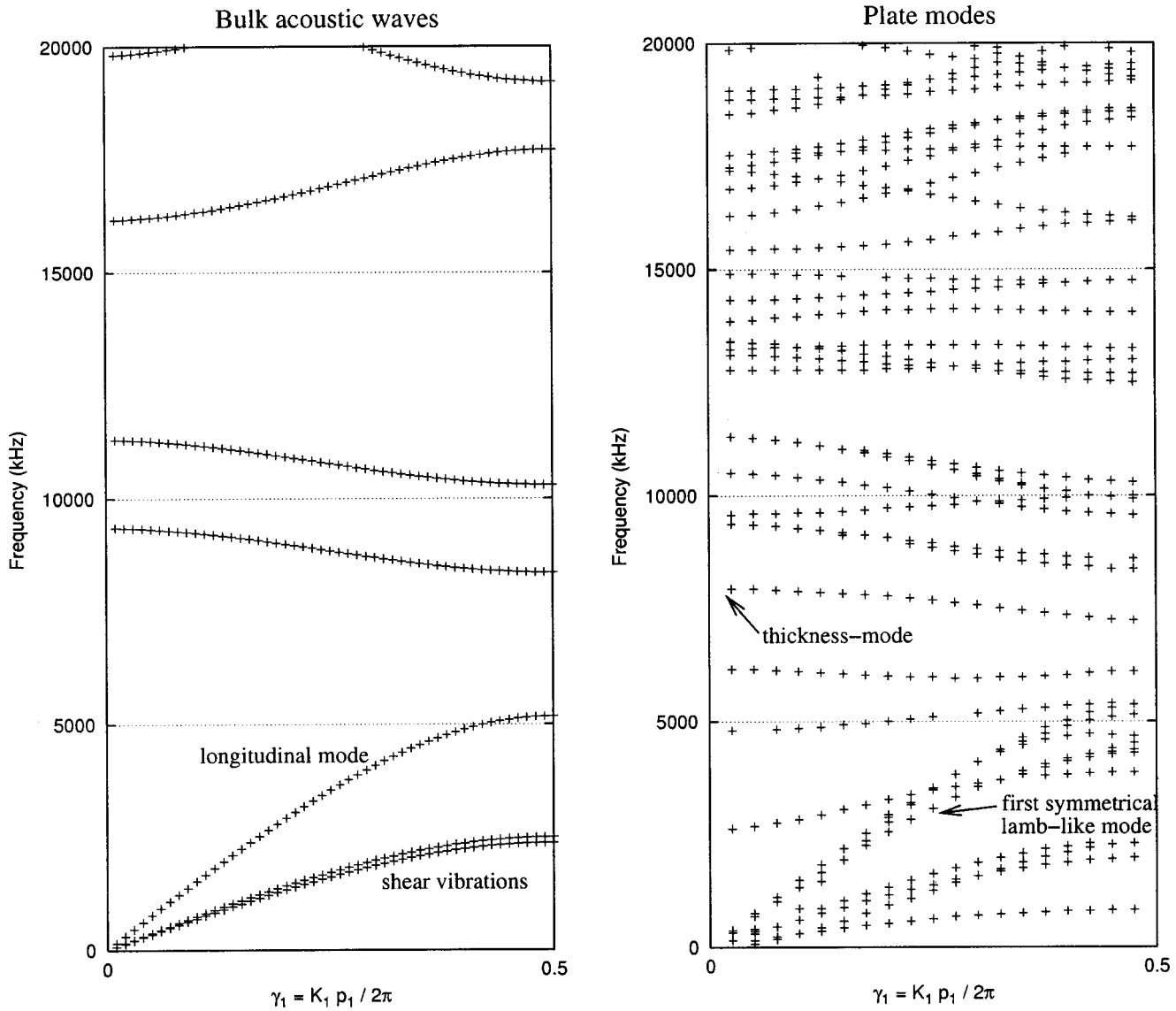


FIG. 3. Dispersion curves of the 2-2 connectivity composite of Fig. 1. Left: bulk acoustic waves propagating in the (x_1,x_2) plane. Right: plate modes of the finite-thickness composite given by the minima of the boundary condition system determinant. Bulk modes are found too with this second computation, so that bulk and plate modes are discriminated comparing the two figures.

obtained by setting $K_1=\pi/p_1$. For more details about the harmonic admittance, see Ref. 12. The charge distribution at the active interface is directly given by the difference between the normal electrical displacement in the material and in a vacuum (assuming ideally thin electrodes). Integrating this distribution along one period and multiplying the result by the angular frequency provides the harmonic admittance written $(i=1,...,N)$

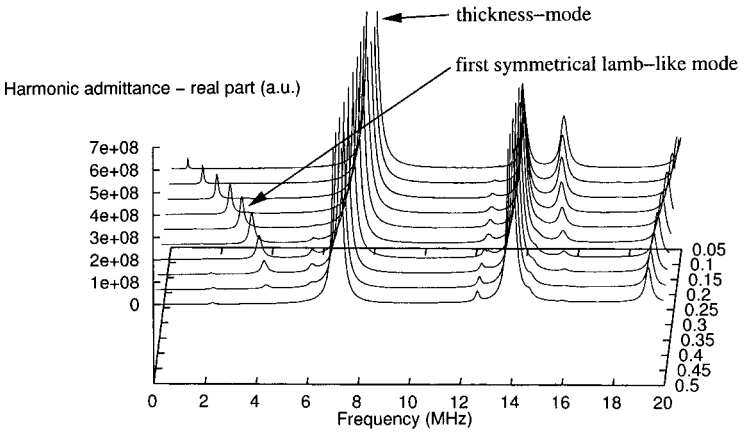


FIG. 4. Real part of the harmonic admittance (conductance) of the 2-2 connectivity piezoelectric composite material of Fig. 1. Resonance frequencies are given by the maxima of the conductance. The admittance is truncated since the thickness mode has a great amplitude compared to the other modes. We can derive from this calculus and the previous curves of Fig. 3 dispersion curves for piezoelectrically coupled plate modes, since conductance maxima indicate vibrating modes in the same manner as dispersion curves.

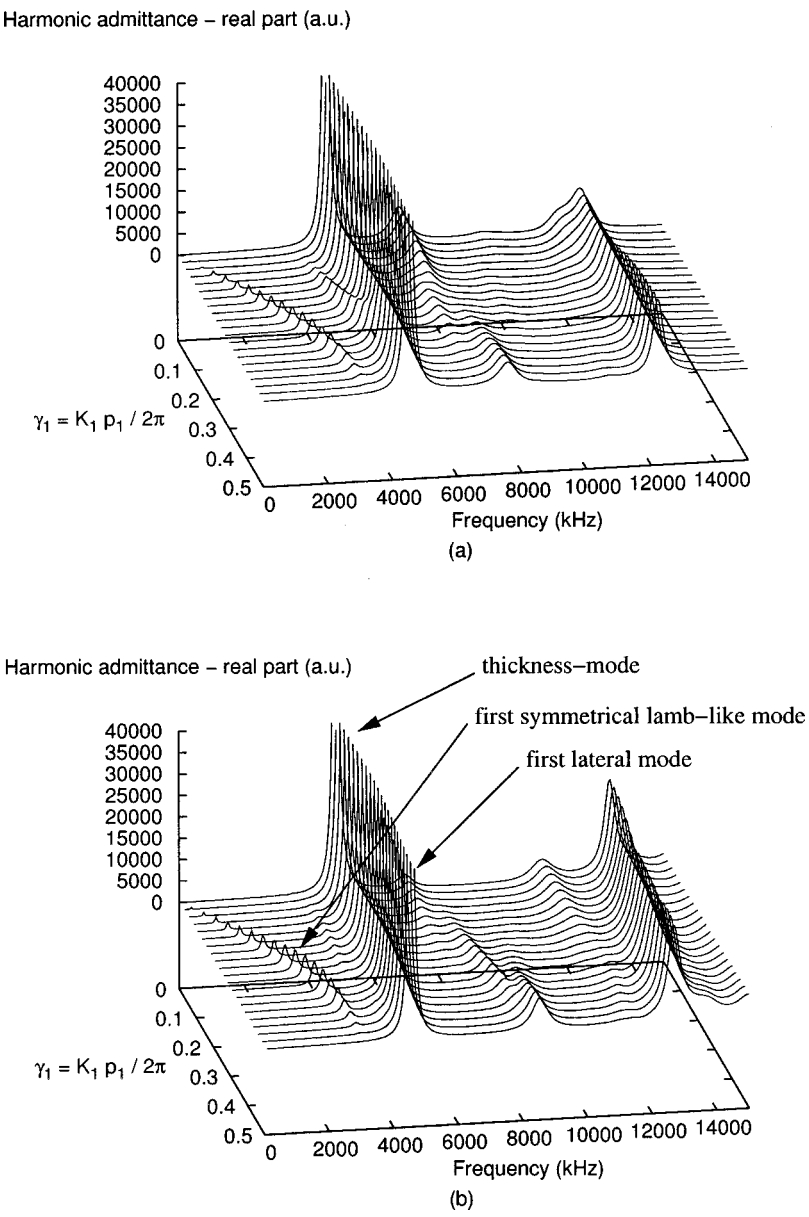


FIG. 5. Comparison between finite element (a) and PWE (b) calculations of the harmonic admittance of the 1-3 piezoelectric composite of Fig. 1. The first symmetrical Lamb-like wave, the thickness mode, and the well-known first lateral mode are indicated in the PWE curves. The displacement fields of these modes are respectively shown in Figs. 6–8.

$$Y(K_1,K_2,\omega) = j\omega \sum_{i=1}^{N} \left[\sum_{r=1}^{8N} A^{(r)} (D_{3\mathbf{K}+\mathbf{G}^i}^{(r)} - \epsilon_0 |\kappa| \phi_{\mathbf{K}+\mathbf{G}^i}^{(r)}) e^{-jK_3^{(r)}h} \right] \times \left(p_1 \mathrm{sinc}\left((K_1 + G_1^i)\frac{p_1}{2} \right) \right) \left(p_2 \mathrm{sinc}(K_2 + G_2^i)\frac{p_2}{2} \right). \quad (35)$$

This harmonic admittance provides precious information about the way the modes are piezoelectrically coupled, but also concerning frequency band-gap phenomena arising in periodic structures. It is also very useful to derive mutual admittances[12] allowing us then to evaluate cross-coupling effects for a given geometry of transducers. This latter point will be addressed in future works.

Note that in all the presented theoretical developments, the material constants are assumed complex, so that elastic, piezoelectric, and dielectric losses can be taken into account.[13]

III. NUMERICAL APPLICATION OF THE MODEL

A. Acoustic plate modes in quartz

This first example was chosen simply to validate the approach. The problem is the calculation of acoustic plate modes of a Z cut plate of quartz, with propagation along the *X* axis. By considering only the fundamental term in the Fourier and Floquet series, one can directly simulate the acoustic propagation in a homogeneous material of finite thickness. The results are easily compared to those of SAW (surface acoustic wave) dedicated simulation tools developed in our group.[14] This comparison is shown in Fig. 2 where the phase velocity is displayed as a function of the frequency-thickness product. A very good agreement between both calculations for piezoelectrically coupled modes can be observed. The proposed approach also allows one to identify the noncoupled modes of the plate, and also modes which energy propagates parallel to the plate surfaces (Poynting vector parallel to the surfaces), but are not affected by the plate thickness and hence are not dispersive. Four modes of

this kind are found in Fig. 2, and are the slow shear, fast shear, and longitudinal bulk acoustic waves, plus a fast shear SSBW (surface skimming bulk wave, see Ref. 1).

B. 2-2 connectivity piezoelectric composite

The geometry of the considered structure is shown in Fig. 1. The thickness of the plate was set to 200 μm; the period along x_1 was 150 μm with a 100-μm width PZT ridge.

In that case, different computations were performed to identify the modes propagating in the structure. Bulk modes were first determined assuming an infinite thickness for the plate and $K_3=0$. The eigenmodes of the finite thickness plate were then extracted assuming that both surfaces are electrically shorted. Finally, the harmonic admittance was computed to discriminate piezoelectrically coupled modes. The top surface is submitted to a harmonic excitation, whereas the back surface is grounded. All computations were performed considering six spatial harmonics. It should be emphasized that using more spatial harmonics yields a more nearly accurate prediction of the frequency location of the modes. Considering six harmonics represents a trade-off between accuracy and computation time. The results obtained for the three different computations are reported in Figs. 3 and 4. Figure 3 shows the bulk eigenmodes propagating in the (x_1,x_2) plane and the plate eigenmodes of the finite-thickness configuration, and Fig. 4 displays the real part of the harmonic admittance of the piezoelectric composite for all couples $(\omega,\gamma_1=K_1p_1/2\pi)$. In the case of infinite thickness of the plate, one can easily identify the first three modes as in-plane and thickness shear vibrations, and in-plane longitudinal propagation. The latter strongly contributes to the first symmetrical Lamb wave. In the case of the finite thickness configuration, many couples (ω,γ_1) are found to minimize the determinant of the boundary condition system. However, some of these values correspond to local minima of this determinant and not to physical solutions of the problem. For instance, bulk modes appear for the finite-thickness configuration, too, and can be discriminated using the results of the case of infinite thickness. In Fig. 4, it can be seen that the piezoelectrically coupled modes of the device can be discriminated efficiently, using the harmonic admittance computation. In that case, eigenmode and harmonic admittance computations are both useful since the dispersion curves of Fig. 3 allow one to distinguish between modes that are very close to one another and not easily independently identifiable by the use of the harmonic admittance.

C. 1-3 connectivity piezocomposite

The last illustration of the proposed theory is a 1-3 connectivity piezoelectric composite. The elementary cell of the studied structure is shown in Fig. 1. The thickness was set to 300 μm, and both periodicities along x_1 and x_2 are equal to 200 μm (PZT width is 100 μm along x_1 and x_2). The same kind of analysis as for the 2-2 piezocomposite was performed. However, due to the large computation time required, the number of spatial harmonics was restricted to five for both directions, i.e., 25 terms in the series. This is a major

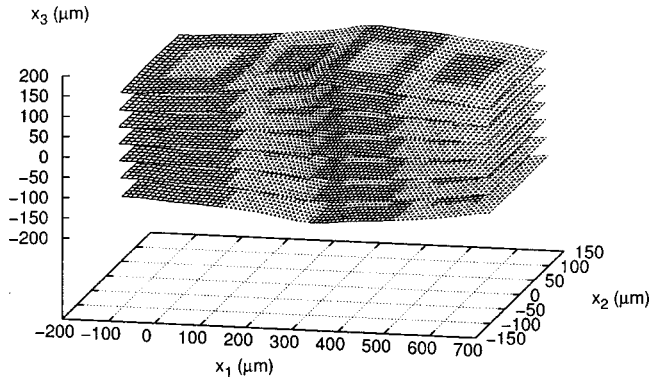


FIG. 6. First symmetrical Lamb wave propagating in four cells ($\gamma_1=0.25$, $\gamma_2=0$, $\omega=2350$ kHz) of the 1-3 piezoelectric composite of Fig. 1.

drawback of the method because computations with too small a number of harmonics yield inaccurate predictions.

Although some minima of the determinant are not physical solutions of the problem, dispersion curves for relatively low-frequency modes are found similar to those computed with a finite element method in Ref. 15.

A qualitative comparison between finite element analysis and PWE analysis of the composite is reported in Fig. 5. Although similar results are obtained in the vicinity of the longitudinal compression mode using the two methods, discrepancies arise for higher-order frequency modes. However, both approaches exhibit prohibitive computation times for the accurate simulation of high-order modes and overtones (increasing the number of elements of the finite element mesh is equivalent to increasing the number of harmonics in the PWE method).

Although the problem of accuracy must be addressed to provide reliable computation results, much pertinent information can be deduced from the proposed calculation. The shape of the mode is accessible by simply calculating the effective values of the mechanical displacement fields of the

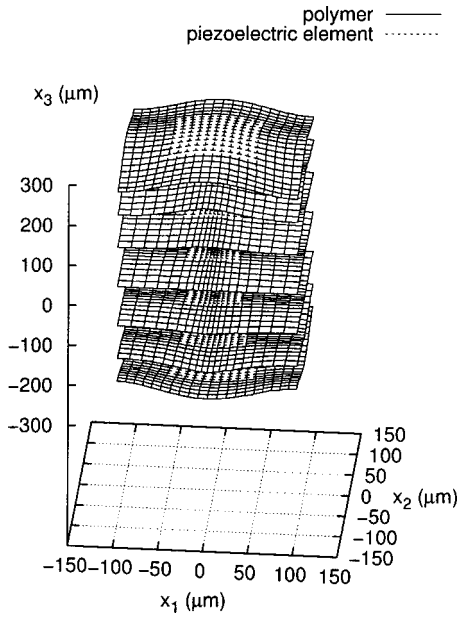


FIG. 7. Fundamental longitudinal mode along x_3 ($\gamma_1=\gamma_2=0$, $\omega=4900$ kHz) of the 1-3 piezoelectric composite of Fig. 1.

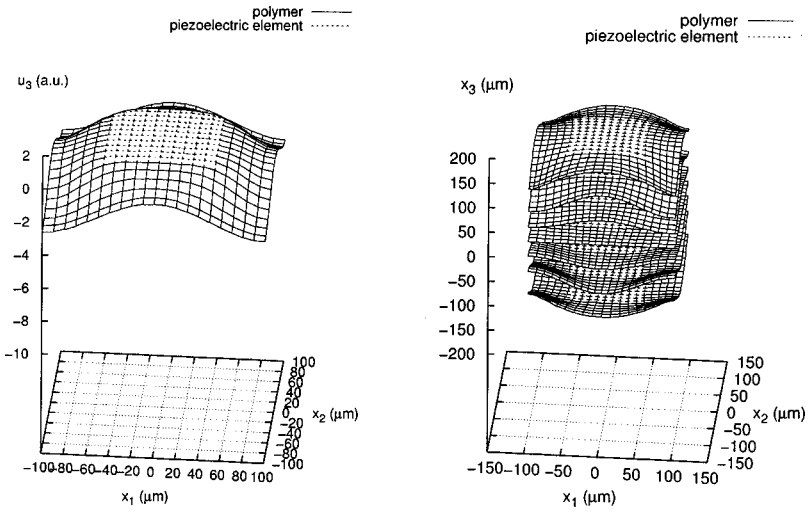


FIG. 8. First lateral mode for infinite thickness and finite thickness ($\gamma_1=\gamma_2=0$, $\omega=7200$ kHz) of the 1-3 piezoelectric composite of Fig. 1.

problem. For a given value of $(\gamma_1,\gamma_2,\omega)$, one can construct the propagation profile for the desired number of cells. This is illustrated in Fig. 6, showing the first Lamb mode at ($\gamma_1=0.25$, $\gamma_2=0$, $\omega=2350$ kHz) propagating in four adjacent cells. The mechanical state of the elementary cell for the fundamental longitudinal compression mode along x_3 and for the first lateral mode ($\gamma_1=\gamma_2=0$) are also reported in Figs. 7 and 8, respectively, demonstrating the capability of the model to fairly predict the actual vibration of a 1-3 connectivity piezoelectric composite.

Figure 9 displays the real part of the harmonic admittance (i.e., the conductance) in the first Brillouin zone[16] of the studied in-plane symmetrical structure. We can recognize for $\gamma_2=0$ the harmonic admittance of Fig. 5. Resonance frequencies are given by the maxima of the conductance. Thanks to these conductance curves, which give the resonant piezoelectrically coupled modes, and the dispersion curves calculated in the case of shorted surfaces, we are able to identify parasitic modes of the thickness mode and acoustic band-gaps of the structure, which are of main interest for ultrasonic imaging applications.

Finally, a quantitative study was performed in the case of the fundamental thickness mode. First, the longitudinal phase velocity v_l and the electromechanical coupling factor k were calculated by varying the volume fraction of ceramic, with our constant sets, according to the well-known Smith and Auld theory.[2] Then the harmonic admittance was calculated at ($\gamma_1=\gamma_2=0$) around the thickness-mode for different numbers of terms in the Bloch–Floquet and Fourier expansions. The maximum of the conductance gives one the resonance frequency f_r, while the antiresonance frequency f_a is given by the maximum of the resistance (real part of the impedance).[17] Consequently v_l and k can be calculated as follows:

$$v_l=2hf_a, \tag{36}$$

where h is the thickness of the plate, and

$$k^2=\frac{\pi}{2}\frac{f_r}{f_a}\tan\frac{\pi}{2}\frac{f_a-f_r}{f_a}. \tag{37}$$

Comparisons between theory[2] and our computations are shown in Figs. 10 and 11. By increasing the number of terms of the expansions, a convergence appears between the Smith and Auld model and the plane-wave-expansion method. For instance, with seven terms along each direction, we have a good agreement for a ceramic volume fraction from 0% to 50%.

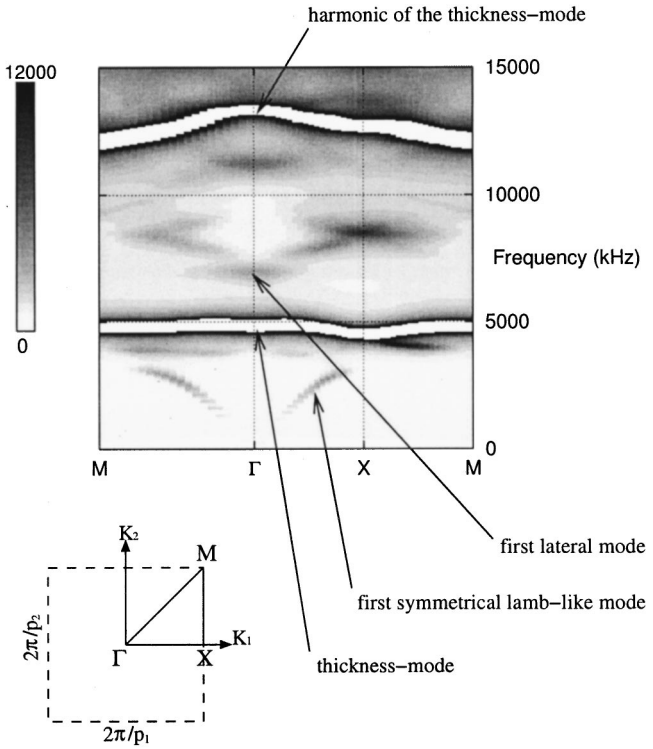


FIG. 9. Real part of the harmonic admittance, for the 1-3 piezoelectric composite, calculated in the first Brillouin zone in the (K_1,K_2) plane. One can deduce the frequencies of the parasitic modes, considering the thickness mode as the useful vibration for imaging applications. For instance, considering the path $X-M$, we can see that a vibrating mode mingles with the strong coupled fundamental thickness mode near the point X. Acoustic band-gaps are retrievable considering piezoelectrically coupled modes. The thickness mode and its harmonic are truncated (white color) because of their magnitudes.

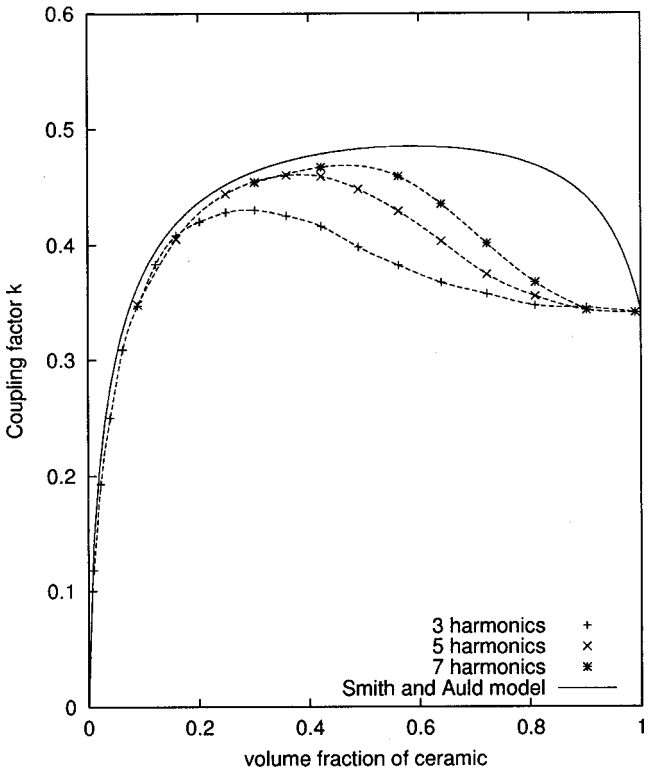


FIG. 10. Electromechanical coupling factor for the thickness mode, calculated for different values of the number of terms of the series, to be compared with the well-established Smith and Auld model. The number of harmonics along each direction is given for each computation.

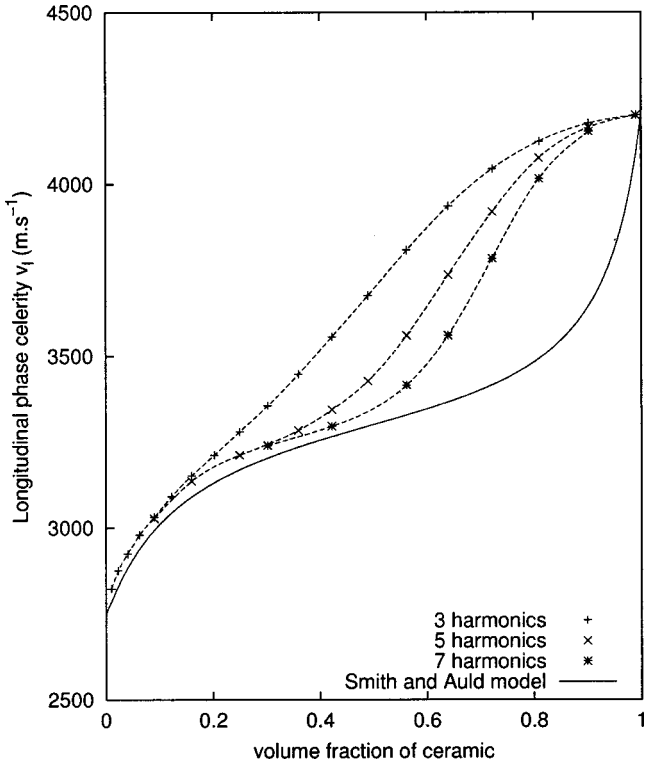


FIG. 11. Longitudinal phase velocity for the thickness mode, calculated for different values of the number of terms of the series, to be compared with the well-established Smith and Auld model. The number of harmonics along each direction is given for each computation.

IV. CONCLUSION

A full 3D model based on a plane wave expansion of the generalized acoustic fields has been developed and implemented for general piezoelectric-based composite materials. It is able to simulate homogeneous materials, but also 1D or 2D periodic devices with a good level of accuracy. It can take into account material losses using the imaginary part of elastic, piezoelectric, and dielectric fundamental constants. This approach is an alternative to finite element computations in the case of the considered regular piezoelectric composite geometry. The model is able to provide much information about the considered device such as the piezoelectric coupling, the resonance efficiency, and the effect of acoustic losses by calculating the harmonic admittance. It can incorporate different boundary conditions corresponding to effective operation of the device. First qualitative computations were performed for 2-2 and 1-3 connectivity piezoelectric composites. Complementary quantitative calculations were performed for the 1-3 connectivity piezoelectric composite thickness mode and compared to a well-established theory. In a future extension of the theory, a particular effort will be paid to the optimization of the computation time and the accuracy of 3D computations. The case of radiation in fluids and solids will also be addressed, to provide a comprehensive description of operating conditions of 2-2 and 1-3 piezoelectric composite-based devices. Even if the computation time, which governs in fact the accuracy of the method, has to be optimized, the PWE method already provides qualitative information, which is of primary interest to understand the operation of such composite structures.

ACKNOWLEDGMENTS

This work was jointly supported by the CNRS, THALES MICROSONICS (TMX), and FRAMATOME-ANP. The authors would like to thank Dr. Abdelkrim Khelif for fruitful discussions concerning the fundamentals of plane-wave-expansion methods, and also W. Steichen (TMX) and G. Pierre (FRAMATOME-ANP) for encouraging and supporting the study.

[1] B. A. Auld, *Acoustic Fields and Waves in Solids*, 2nd ed. (Krieger, Melbourne, FL, 1990).
[2] W. A. Smith and B. A. Auld, "Modeling 1-3 composite piezoelectrics: Thickness-mode oscillations," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **38**(1), 40–47 (1991).
[3] R. E. Newnham, D. P. Skinner, and L. E. Cross, "Connectivity and piezoelectric-pyroelectric composites," Mater. Res. Bull. **13**, 525–536 (1978).
[4] D. Certon, F. Patat, F. Levassort, G. Feuillard, and B. Karlsson, "Lateral resonances in 1-3 piezoelectric periodic composite: Modeling and experimental results," J. Acoust. Soc. Am. **101**, 2043–2051 (1997).
[5] J. O. Vasseur, B. Djafari-Rouhani, L. Dobrzynski, M. S. Kushwaha, and P. Halevi, "Complete acoustic band gaps in periodic fibre reinforced composite materials: The carbon/epoxy composite and some metallic systems," J. Phys.: Condens. Matter **6**, 8759–8770 (1994).
[6] Y. Tanaka and S.-I. Tamura, "Acoustic stop bands of surface and bulk modes in two-dimensional phononic lattices consisting of aluminum and a polymer," Phys. Rev. B **60**, 13294–13297 (1999).
[7] A. H. Fahmy and E. L. Adler, "Propagation of surface acoustic waves in multilayers: A matrix description," Appl. Phys. Lett. **22**, 495–497 (1973).
[8] L. Boyer, "Etude Des Phénomènes de Réflexion-Réfraction Des Ondes Planes Acoustiques Dans Les Milieux Piézoélectriques," Ph.D. thesis, Université Paris 7, France, 1994.

[9] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953).
[10] R. C. Peach, ''A general green function analysis for SAW devices,'' in *IEEE Ultrasonics Symposium*, 1995, pp. 221–225.
[11] X. Geng and Q. M. Zhang, ''Evaluation of piezocomposites for ultrasonic transducer applications—influence of the unit cell dimensions and the properties of constituents on the performance of 2-2 piezocomposites,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**(4), 857–872 (1997).
[12] Y. Zhang, J. Desbois, and L. Boyer, ''Characteristic parameters of surface acoustic waves in a periodic metal grating on a piezoelectric substrate,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**(3), 183–192 (1993).
[13] R. Holland, ''Representation of dielectric, elastic, and piezoelectric losses by complex coefficients,'' IEEE Trans. Sonics Ultrason. **14**(1), 18–20 (1967).
[14] Th. Pastureaud, V. Laude, A. Soufyane, and S. Ballandras, ''Stabilization of the simulation of SAW devices on stratified structures: Application to transverse plate mode resonators,'' IEEE Int. Freq. Control Symp., 661–665 (Seattle, 2001).
[15] P. Langlet, ''Analyse de la Propagation Des Ondes Acoustiques Dans Les Matériaux Périodiques À L'aide de la Méthode Des Éléments Finis,'' Ph.D. thesis, Université de Valenciennes et du Hainaut-Cambresis, France, 1993.
[16] L. Brillouin, *Propagation in Periodic Structures* (Dover, New York, 1953).
[17] ''ANSI/IEEE standard on piezoelectricity, std 176-1987,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**(5), 717–772 (1996).

Wave selection using a magnetomechanical sensor in a solid cylinder

Hocheol Lee[a] and Yoon Young Kim[b]
School of Mechanical and Aerospace Engineering, Seoul National University San 56-1, Shinlim-dong, Kwanak-gu, Seoul 151-742, Korea

(Received 6 July 2001; revised 27 March 2002; accepted 31 May 2002)

Magnetomechanical sensors have received much attention in recent years. However, investigations on the rejection of unwanted wave modes and thus the selection of a desired mode have not been reported. The purpose of this investigation is to present a technique to select only a desired wave mode in a solid ferromagnetic cylinder using a magnetomechanical sensor. To achieve this goal, we propose some bias magnet configurations for each of which either longitudinal or flexural waves alone can be extracted. A simplified theory relating bias magnet arrangement and measured wave modes is developed. Several experimental results confirm the validity of the present investigation. [DOI: 10.1121/1.1497623]

PACS numbers: 43.38.Ct, 43.35.Yb [SLE]

I. INTRODUCTION

Ferromagnetic materials subjected to stresses exhibit some amount of changes in their magnetic states. This effect, known as the Villari effect[1] is the inverse phenomenon of the well-known Joule effect.[2] This Villari effect is also referred to as the inverse magnetostriction or the magnetomechanical effect.

Sensors using this effect, which will be called magnetomechanical sensors, hold their unique application areas as they do not require any direct physical contact. In particular, noncontact magnetomechanical sensors are very useful when a measurement target is in motion or covered with other (nonferromagnetic) materials. They are also useful when the temperature of a target is very high.

Magnetomechanical sensors and magnetostrictive actuators have been used in various applications such as those by Williams,[3] Tzannes,[4] Kim *et al.*,[5,6] Aime *et al.*,[7] Onoe,[8] Kwun *et al.*,[9,10] Kwun and Bartels,[11–13] Kwun and Tellers,[14] and Murayama.[15,16] However, the complex behavior of the Villari effect including nonlinearity and hysteresis makes it difficult to analyze the physics involved in magnetomechanical sensors. In addition, iron that is the most widely used ferromagnetic material exhibits even more complicated behavior. Though a theory of the magnetomechanical effect has been recently developed,[17] this theory appears too complicated to use in practical sensor design applications.

The generation and measurement of elastic waves based on the Joule and Villari effects have also played an important role in dispersion analysis and diagnostics in waveguides. Kwun and Teller[14] have utilized the Joule and Villari effects in the study of the propagation of longitudinal, torsional, and flexural waves in a cylinder. Although no specific bias field configuration or related analysis was provided, they addressed that measured and generated waves by magnetomechanical sensors are strongly influenced by the nature of the applied bias magnetic field near measurement points. Therefore, the use of appropriate bias magnetic field configurations may allow the selective measurement of a desired wave mode, either longitudinal or flexural alone, from a compound wave signal. If the desired wave mode is measured accurately, a sophisticated diagnostic technique may be developed. However, no investigation is reported on how to select a desired mode using magnetomechanical sensors.

The goal of this investigation is to present a technique to select a desired wave mode in a waveguide using magnetomechanical sensors. In the present investigation, we consider the selection of only a flexural or longitudinal mode in a wave consisting of both flexural and longitudinal modes. Since the applied bias field affects the type of the measured wave mode, the focus of the present work lies in identifying appropriate bias magnet configurations.

In the first part of this work, we develop a simplified model to explain the relation between the characteristics of the bias field and the measured wave mode in a solid circular cylinder.

In the second part of this work, we present two bias field configurations for each of which either a longitudinal or a flexural mode alone can be selected. In particular, the bias configuration for the flexural mode measurement rejecting the longitudinal mode is newly proposed. To carry out the present investigation, a wave is generated in a system similar to the Hopkinson bar system. The finite element analysis is performed to justify the present choice of the longitudinal wave-rejection configuration.

II. PROBLEM DEFINITION

Figure 1 shows a transient wave propagating in a long solid cylinder made of a ferromagnetic material such as iron. Our aim is to develop a magnetomechanical sensor-based experimental technique to measure either a longitudinal or a flexural mode from a wave consisting of both longitudinal and flexural modes.

Among several types of magnetomechanical sensors, a solenoid-type magnetomechanical sensor (simply referred as

[a]Electronic mail: hclee21@kornet.net
[b]Author to whom correspondence should be addressed. Electronic mail: yykim@shu.ac.kr

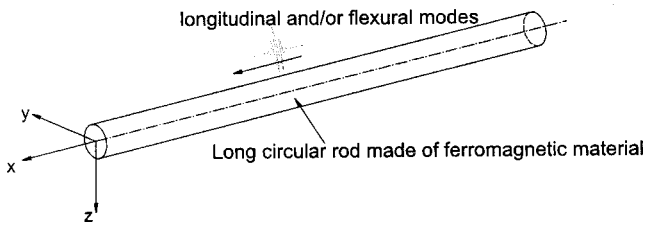


FIG. 1. Propagation of a wave consisting of longitudinal and/or flexural modes in a long solid circular rod made of a ferromagnetic material.

a solenoid sensor in this work) is adopted in this work; see Fig. 2 for its configuration. It may be worth mentioning the advantages of the solenoid sensor: noncontact measurement is possible; relative high transduction efficiency is obtained in comparison with other types of noncontact sensors such as eddy-current type sensors; the sensor operates in a wide frequency band.

A solenoid sensor of length L and diameter D encircles a circular rod of diameter d. Since the sensor diameter D and the rod diameter d are not necessarily equal, the cross sectional areas of the rod and the encircled area by the sensor are denoted by A_r and A_s, respectively. Various kinds of bias magnetic fields generated either by electromagnets or by permanent magnets may be applied in order to increase measurement sensitivity.

Since the principle of the magnetomechanical sensor is based on the Villari effect, we express the Villari effect in one-dimensional form as

$$B = B_H + B_\sigma, \tag{1}$$

$$B_H = \mu^\sigma H, \tag{2a}$$

$$B_\sigma = q\sigma, \tag{2b}$$

where B and H denote magnetic induction and magnetic field strength (along the x axis in Fig. 2). The longitudinal stress is denoted by σ. Two coefficients μ^σ and q represent the permeability under constant stress and the Villari effect coefficient, respectively.

If stresses are developed at the point of a rod which the solenoid sensor encircles, the variation of the magnetic state due to the stresses can be measured through the sensor in the form of voltage output. If the total magnetic flux enclosed by

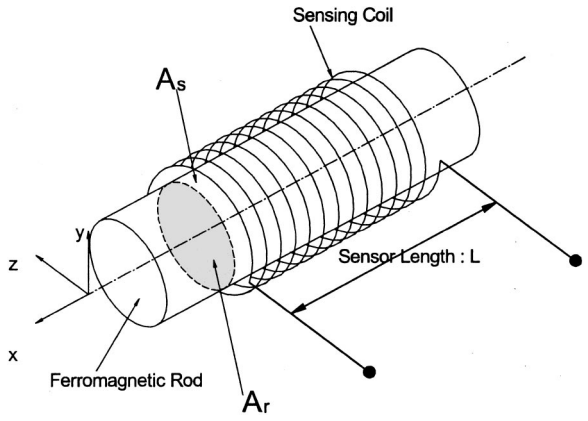


FIG. 2. The configuration of a solenoid sensor.

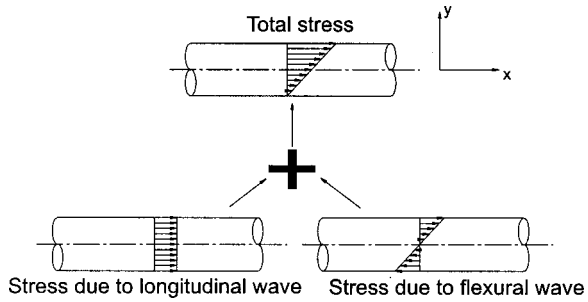


FIG. 3. The resultant stress distribution as a sum of two stress components.

the solenoid sensor is denoted by Φ, the voltage output $V(x,t)$ between both ends of the sensing coil is given by the Faraday–Lentz law as

$$V(x,t) = -\frac{d\Phi}{dt}. \tag{3}$$

If the solenoid sensor is assumed to be ideal, Eq. (3) becomes

$$V(x,t) = -\frac{d}{dt}(N\phi) = -N\frac{d\phi}{dt}, \tag{4}$$

where N is the number of coil turns in the solenoid sensor and ϕ is the magnetic flux encircled by a single turn of coil.

Since the magnetic flux is related to the magnetic induction, the voltage output in Eq. (4) can be written as

$$V(x,t) = -N\frac{d}{dt}\int_{A_s} B\,dA = -N\frac{d}{dt}\int_{A_r} B\,dA, \tag{5}$$

where the area A_s is replaced by the area A_r since the permeability of the rod is much greater than that of air.

Assuming that the applied magnetic field H is time independent, the substitution of Eq. (1) into Eq. (5) yields the relation between the output voltage $V(x,t)$ and the stress σ as

$$V(x,t) = -N\frac{d}{dt}\int_{A_r} q\sigma \cdot d\mathbf{A}. \tag{6}$$

Thus far, we give the problem definition and review the underlying equation relating the output voltage measured through the solenoid sensor and the stress in a rod.

III. MODE SELECTION: ANALYSIS AND EXPERIMENTAL CONFIGURATION

A. Simplified analysis for mode selection

In this section, we develop a simplified theoretical model to link the characteristics of a applied bias field with wave modes measured by a solenoid sensor. The specific analysis will be carried out for the case of coupled longitudinal and flexural wave modes.

When longitudinal and flexural waves propagate in a rod, the simplest approximation of the resulting stress field may be the one shown in Fig. 3. In the present analysis, it is assumed that displacements due to flexural waves occur in the xy plane and that the variations of field quantities along the z direction can be ignored.

The displacement field based on the Timoshenko beam theory[18] can be written as

$$u(x,y,t)=u_0(x,t)-y\,\theta(x,t), \tag{7}$$

where $u_0(x,t)$ denotes the average axial displacement and $\theta(x,t)$, the rotation of a surface normal. In this approximation, the displacement $u_0(x,t)$ and the rotation $\theta(x,t)$ are due to the propagation of longitudinal and flexural waves, respectively. The axial stress due to the displacement field is simply

$$\sigma(x,y,t)=E\frac{\partial u}{\partial x}=E\left(\frac{\partial u_0}{\partial x}-y\frac{\partial \theta}{\partial x}\right), \tag{8}$$

where E represents Young's modulus of the specimen and $\partial u_0/\partial x$, the average axial strain.

In the case when the magnetic induction B_H (actually, the x component of the magnetic induction) is not uniform through the cross section of a rod, B_H may be expanded in terms of the Taylor series in the coordinate y,

$$B_H(x,y)=b_0(x)+b_1(x)y+b_2(x)y^2+b_3(x)y^3+\cdots, \tag{9}$$

where b's are the coefficients used to denote the variation of B_H in x. In Eq. (9), the dependence of B_H on the z coordinate is ignored. By assuming that the coefficient q is proportional to the magnetic induction B_H with the proportional constant k, one can write

$$\begin{aligned} q(x,y)&=kB_H(x,y)\\ &=kb_0(x)+kb_1(x)y+kb_2(x)y^2+kb_3(x)y^3+\cdots\\ &\equiv c_0(x)+c_1(x)y+c_2(x)y^2+c_3(x)y^3+\cdots. \end{aligned} \tag{10}$$

Therefore, the constants c_i's are determined by the applied bias magnetic field distribution.

To express the voltage output V in terms of u_0 and θ, Eq. (10) and Eq. (8) are substituted into Eq. (6) (within the first-order approximation)

$$\begin{aligned} V(x,t)\approx -N\frac{d}{dt}\Bigg\{\int_{A_r}[c_0(x)+c_1(x)y]\\ \times E\left[\frac{\partial u_0(x,t)}{\partial x}-y\frac{\partial \theta(x,t)}{\partial x}\right]dA\Bigg\}. \end{aligned} \tag{11}$$

To simplify Eq. (11), we note that

$$\int_{A_r} y\,dA=0, \quad \int_{A_r} y^2\,dA=I, \tag{12}$$

where I is the mass moment of inertia of the rod cross section about the z axis. Therefore, the output voltage $V(x,t)$ from the sensor can be simplified as

$$\begin{aligned} V(x,t)=NEIc_1(x)\frac{\partial^2\theta(x,t)}{\partial x\,\partial t}\\ -NEA_rc_0(x)\frac{\partial^2 u_0(x,t)}{\partial x\,\partial t}. \end{aligned} \tag{13}$$

Introducing the bending moment $M(x,t)$ and the axial force $P(x,t)$ such that

$$M(x,t)=EI\frac{\partial\theta(x,t)}{\partial x}, \quad P(x,t)=EA_r\frac{\partial u_0(x,t)}{\partial x}. \tag{14}$$

Equation (13) finally reduces to

$$V(x,t)\approx N\left[c_1(x)\frac{\partial M(x,t)}{\partial t}-c_0(x)\frac{\partial P(x,t)}{\partial t}\right]. \tag{15}$$

Equation (15) states that when longitudinal and flexural waves are transmitted simultaneously through a rod, the sensor output is dependent mainly on the coefficients (c_0,c_1) and the wave modes (θ or u_0). Therefore, depending on the bias magnetic induction B_H determined by the applied bias magnetic field H, different wave modes can be measured. For instance, if $C_0(x)\simeq 0$, the sensor output measures the time variation of M, which represents flexural waves.

Case 1. Longitudinal mode selection

If the coefficient set (c_0,c_1) satisfies the following condition:

$$c_0(x)\neq 0, \quad c_1(x)=0, \tag{16}$$

the following equations hold at the measurement point x:

$$B_H(x,y)\approx b_0(x), \quad q(x,y)\approx c_0(x). \tag{17}$$

In this case, Eq. (15) becomes

$$V(x,t)\approx -Nc_0(x)\frac{\partial P(x,t)}{\partial t}. \tag{18}$$

Therefore, the solenoid sensor can measure longitudinal waves as long as the bias magnetic induction B_H is uniform along the y axis [see Eq. (17)]. The bias magnet configuration to yield the desired magnetic induction B_H will be discussed in the next section.

Case 2. Flexural mode selection

If the coefficients set (c_0,c_1) satisfies the following condition:

$$c_0(x)=0, \quad c_1(x)\neq 0, \tag{19}$$

the following equations hold at the measurement point x:

$$B_H(x,y)\approx b_1(x)y, \quad q(x,y)\approx c_1(x)y. \tag{20}$$

In this case, Eq. (15) reduces to

$$V(x,t)\approx Nc_1(x)\frac{\partial M(x,t)}{\partial t}. \tag{21}$$

Equations (20) and (21) imply that the antisymmetric distribution of the bias magnetic induction B_H along the y axis enables the solenoid sensor to measure flexural waves only.

B. Bias magnet configuration for mode selection

Case 1. Longitudinal mode selection

The foregoing analysis indicates that a signal from the solenoid sensor contains only longitudinal waves if the bias magnetic induction is uniform across the rod cross section at the measurement location. Perhaps the simplest arrangement

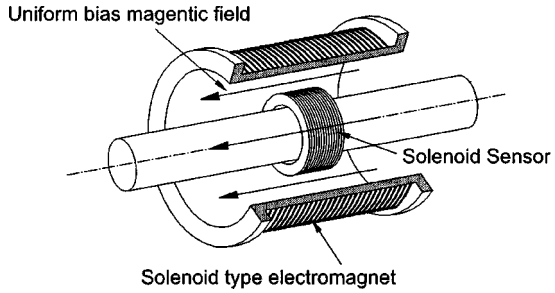


FIG. 4. A solenoid sensor encircled by a solenoid-type electromagnet. This electromagnet is used to produce uniform magnetic induction in a rod.

to achieve uniform magnetic induction is to use a solenoid-type electromagnet as shown in Fig. 4. This well-known configuration has been already employed in delay lines.[3]

If the static current flows in the solenoid-type electromagnet, a uniform magnetic field in the direction of the x axis will be developed. This will cause uniform magnetic induction inside the ferromagnetic rod.

Case 2. Flexural mode selection

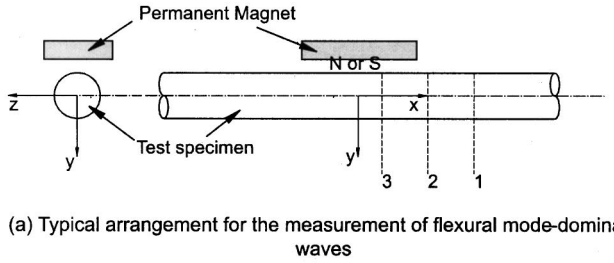
The determination of the bias magnet arrangement to produce a linear antisymmetric magnetic induction ($B_H \approx b_1 y$) is not a simple matter. To reject longitudinal modes, $b_0 = 0$ must be fulfilled [see Eq. (20)]. Otherwise, the measured signals by the solenoid sensor would contain both longitudinal and flexural waves. Figures 5(b) and (c) show the distribution of the magnetic flux lines and the magnetic inductions inside a cylindrical rod under a typical bias magnet arrangement shown in Fig. 5(a). This arrangement may be used for flexural-wave dominant situations. The numerical results are obtained by ANSYS,[19] a finite element code. Though two-dimensional finite element analysis is performed for the present problem, the distribution of the magnetic induction along the y coordinate by the two-dimensional analysis is not far from that by three-dimensional analysis (see Ref. 20).

The distributions of the axial magnetic induction B_H shown in Fig. 5(c) are not uniform along the y coordinate. However, this distribution has both symmetric and nonsymmetric parts, so that both b_0 and b_1 are nonzero. Therefore, the configuration suggested in Fig. 5(a) can capture not only flexural modes but also longitudinal modes. This phenomena will be clearly observed in actual experiments, which will be dealt with in the next section.

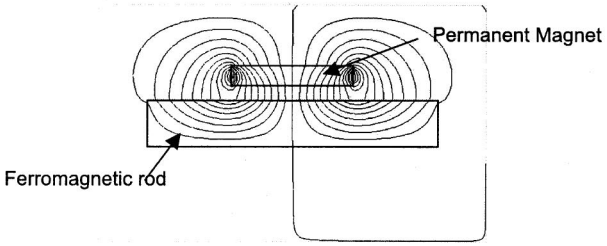
In order to reject longitudinal waves and thus to capture flexural waves only, it is critical to make the magnetic induction B_H inside the rod antisymmetric with respect to y. Motivated by this observation, we propose the bias magnet configuration shown in Fig. 6.

In Fig. 6, we locate two permanent magnets having opposite polarities symmetrically with respect to the rod axis (i.e., the x axis). To guarantee the generation of antisymmetric magnetic induction and to find an optimal sensor location, we carry out a finite element analysis.

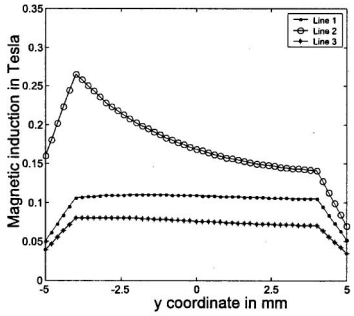
Figure 7 shows the contour of the magnetic flux lines for the bias magnet configuration of Fig. 6. Figure 8 plots the axial magnetic induction B_H as the function of y at different



(a) Typical arrangement for the measurement of flexural mode-dominant waves



(b) Flux lines contour for the arrangement described in (a)



(c) Magnetic induction variation along the lines 1, 2, 3 described in (a)

FIG. 5. A typical bias magnet configuration and the resulting magnetic induction inside a rod.

axial locations (lines 1, 2, and 3 of Fig. 6). It is clear that all distributions of B_H shown in Fig. 8 are antisymmetric, and thus b_0 vanishes. Although the distribution is not linear, the distribution of the magnetic induction along the line 2 is very close to a linear one.

In the actual sensor application, the sensor location with high sensitivity is always preferred. To this end, the axial variation of the magnetic induction B_H needs to be investigated. Figure 9 shows the distribution of B_H along the lines 4 and 5 marked in Fig. 6.

Figures 8 and 9 suggest that the maximum sensitivity will be achieved when the magnetomechanical sensor is lo-

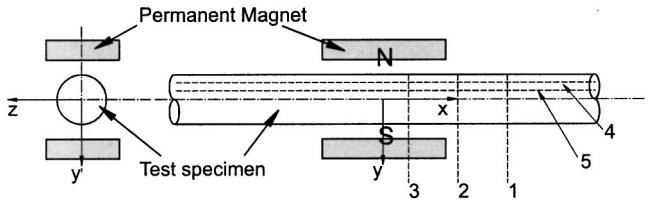


FIG. 6. The proposed bias configuration for the measurement of flexural waves only.

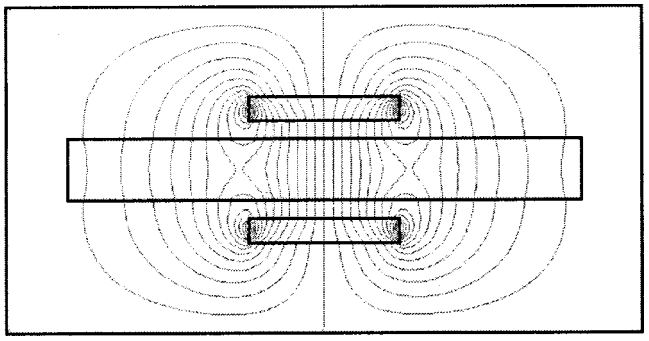


FIG. 7. The contour of the magnetic flux lines for the bias magnet configuration described in Fig. 6.

cated near the ends of the permanent magnets. This is marked by the line 2 in Fig. 6.

IV. EXPERIMENTAL VERIFICATION

A. Experimental setup

Figure 10 shows a schematic diagram of the experimental setup used for the present investigation. The system consists of three main parts: a ferromagnetic rod to be tested, a projectile shooting system for desired wave generation, and a set of measurement equipment. The circular rod is made of mild steel and its length and diameter are 2 m and 10 mm, respectively. The rod is laid down on end supports. The projectile shooting system is similar to the basic Hopkinson bar experiment system (see, e.g., Ref. 21). A projectile of diameter 10 mm is shot by an air gun. The speed of the projectile can be controlled by the pressure of the compressed air and the time during which the projectile is pushed. The length of the projectile governs the duration of a generated pulse.

Some degree of eccentricity makes this device launch longitudinal and flexural waves simultaneously. Only longitudinal waves will be generated if the center of the projectile coincides with the center of the rod. If there is some eccentricity, however, both longitudinal and flexural waves will be generated simultaneously. The validity of the present wave

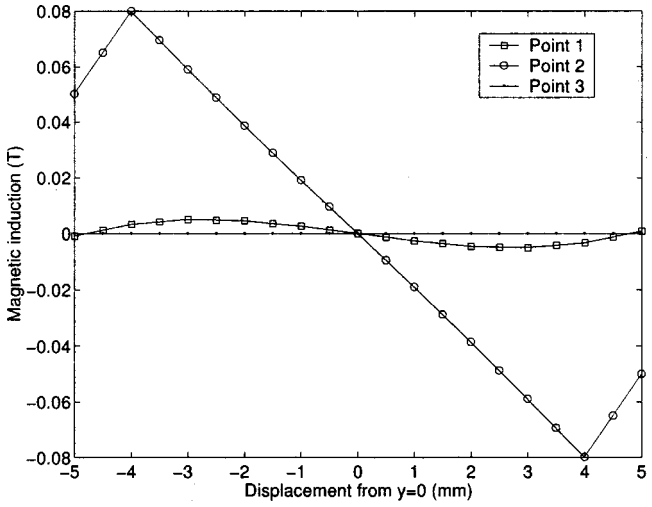


FIG. 8. The distribution of the axial magnetic induction B_H as the function of y at several locations of x.

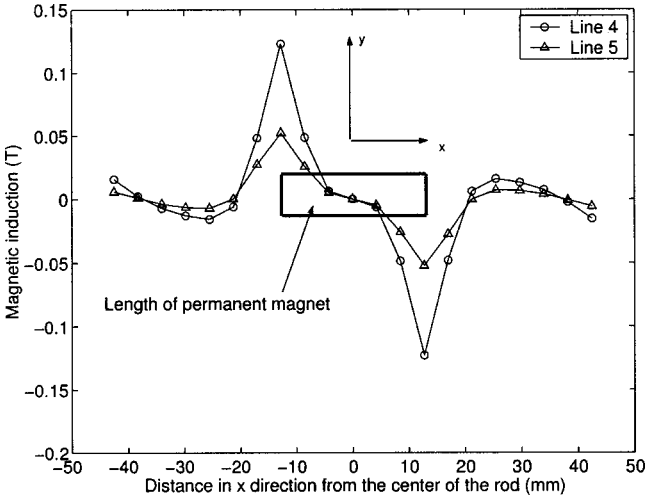


FIG. 9. The distribution of the axial magnetic induction B_H along the lines 4 and 5 marked in Fig. 5.

generation strategy has been confirmed by investigating strain signals of longitudinal and flexural waves measured by strain gages (see Ref. 20 for more details).

The length and diameter of the solenoid sensor used for the present experiment are 5 mm and 13 mm, respectively. The voltage signals captured from the solenoid sensor are sent through a preamplifier (Stanford Research Systems Model SR560) to a digital oscilloscope (Lecroy 9310). The stored signals in the oscilloscope are then transferred to a personal computer for signal processing.

B. Experimental results

This section presents several experimental results directly related to wave mode selection. For all experiments, we use a steel rod of diameter 10 mm and length 2 m. The length and the radius of the projectile used are 30 mm, and 10 mm, respectively. To produce eccentricity, the rod axis and the projectile axis are misaligned by 5 mm. This distance is appropriate to generate both longitudinal and flexural wave modes. To identify the mode characteristics of measured waves by the sensor, we plot the theoretical group velocities of the first branches of the longitudinal and flexural waves (see Refs. 22–24) in Fig. 11 and Fig. 12, respectively.

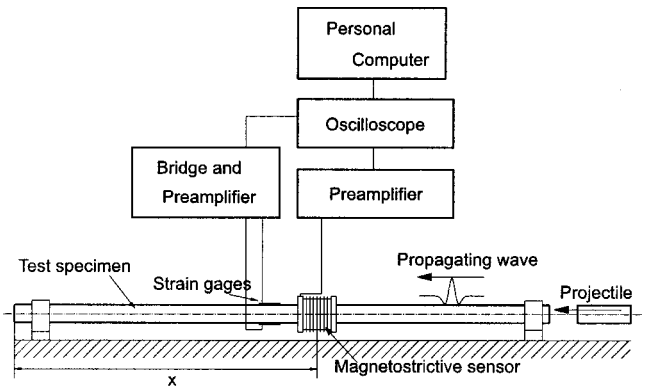


FIG. 10. A schematic diagram for the experimental setup.

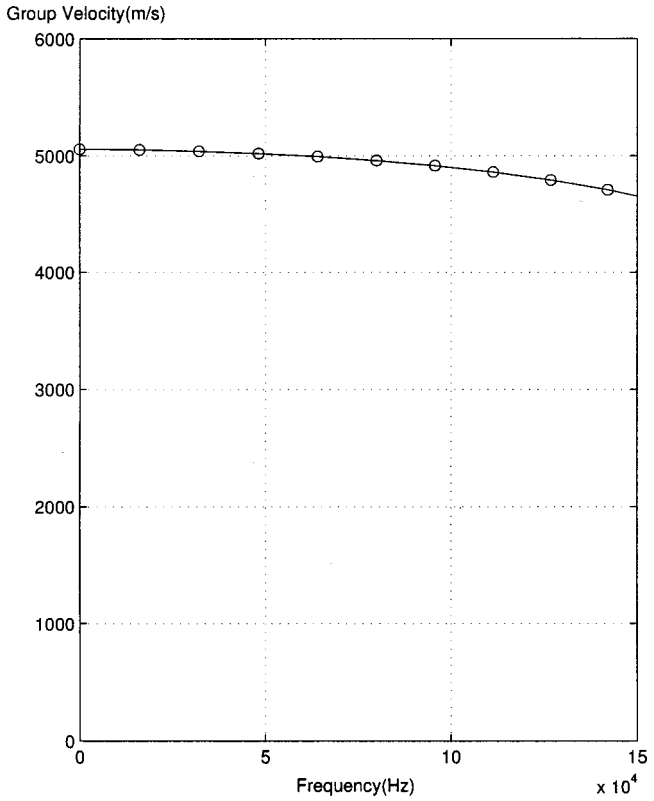


FIG. 11. The theoretical group velocity of the first branch of longitudinal waves in the present steel rod.

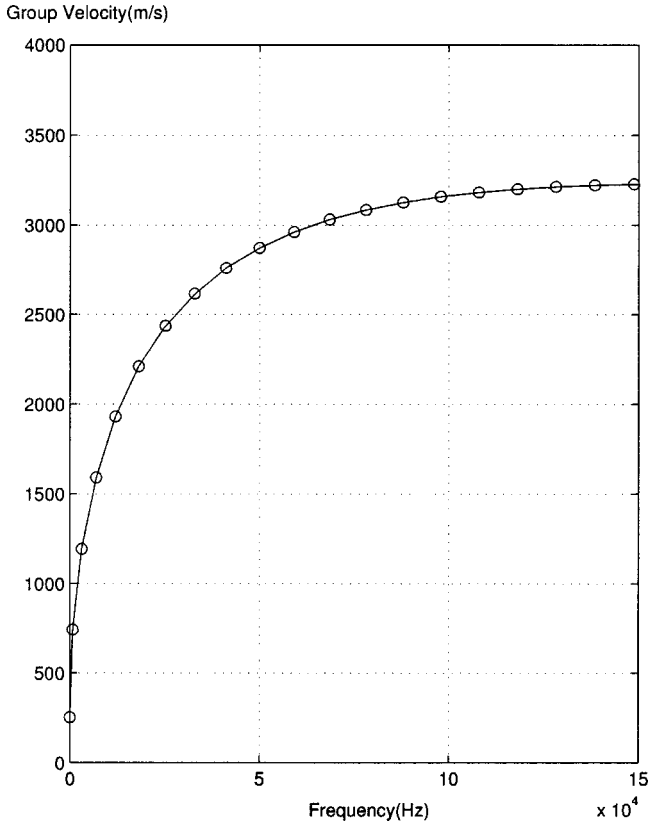


FIG. 12. The theoretical group velocity curve of the first branch of flexural waves in the present steel rod.

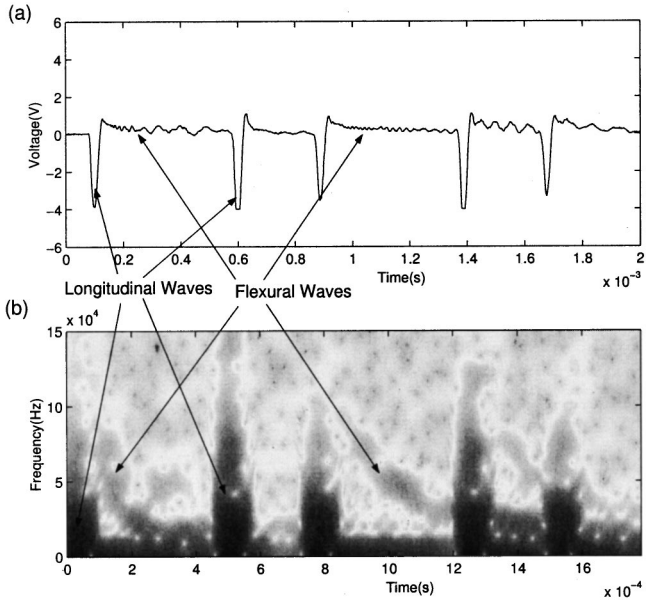


FIG. 13. The captured wave signal with the bias magnet configuration shown in Fig. 5(a). The time history of the signal (a) and its short-time Fourier transform (b).

As the first experiment, the bias magnet configuration shown in Fig. 5(a) is employed. With this configuration, both longitudinal and flexural wave modes may be captured. Figure 13(a) shows the time history of the wave signal measured by the solenoid sensor at location 2 shown in Fig. 5. The spikes in Fig. 13(a) represent longitudinal waves whereas trailing wiggles represent flexural waves. Though the time origin does not affect the subsequent analysis, it may be convenient to use the consistent convention. In plotting the time history of the signal, the time origin $t=0$ is set as the time 10^{-4} s before the arrival of the first peak. To analyze better the characteristics of the mode composition in Fig. 13(a), we carry out the time-frequency analysis of captured signals.

For the analysis of the present signal, we use the short-time Fourier transform,[25] which has been also used in Refs. 9, 11, and 13. Though other transforms are effective for time-frequency analysis (see, e.g., Refs. 26 and 27), the short-time Fourier transform appears most effective for the signals in consideration. The transformed result of Fig. 13(a) by the short-time Fourier transform is plotted in Fig. 13(b).

We will not give the detailed account of the dispersion phenomena of elastic waves in a cylindrical rod, but we mention that the thick close-to-vertical lines represent the incident and reflected waves of the first branch of the longitudinal mode and the thick close-to-hyperbolic curves represent those of the first branch of the flexural mode (see also Refs. 9, 11, and 13). To support this argument, we can extract the group velocities at particular frequencies from the ridge analysis of the time-frequency plot of Fig. 13(b) (see Ref. 25 for the ridge analysis). Since the wave traveling distance by reflection is known, one can convert the flight time to the group velocity. The result at two particular frequencies are summarized as Table I. Table I confirms that both longitudinal and flexural waves are measured by the sensor under the bias magnet configuration shown in Fig. 5(a).

TABLE I. Comparison of the group velocities.

Mode type/Frequency	Experiment	Theory (from Figs. 11 and 12)
Longitudinal/100 kHz	5060 m/s	4912 m/s
Flexural/50 kHz	2836 m/s	2870 m/s

From this preliminary experiment, we observed that unless a special experimental setup is employed, it is not possible to capture longitudinal or flexural waves alone. To capture longitudinal waves alone, we now use the electromagnet configuration shown in Fig. 4. The wave generated in the rod is exactly the same as that for the first experiment. The time history of the wave signal measured by the solenoid senor and its short-time Fourier transform are plotted in Figs. 14(a) and (b), respectively. Apparently, no flexural wave is measured by the solenoid sensor under the bias magnet configuration shown in Fig. 4. Using the same process, the sharp peaks in Fig. 14 can be shown to be correspondant to longitudinal wave.

As the last experiment, which is the most important experiment in the present investigation, we consider the measurement of flexural waves only. As in the previous experiments, both longitudinal and flexural waves are generated simultaneously. The bias magnet configuration we now use for this experiment is shown in Fig. 6. The center of the solenoid sensor is located along the line 2. The distribution of the magnetic induction along the line 2 is already shown in Fig. 8. The measured signal and its short-time Fourier transform are given in Figs. 15(a) and (b), respectively. Unlike the results shown in Fig. 13, the adoption of the proposed bias magnet configuration virtually rejects longitudinal waves. The flexural characteristics of the measured wave may be demonstrated better by the short-time Fourier transform in Fig. 15(b).

It may be worth mentioning that although the same wave is generated for the three cases shown in Figs. 13–15, the signal in Fig. 13(a) is not a composite of the signals in Fig. 14(a) and Fig. 15(a). This is because different bias magnet configurations are used in each case.

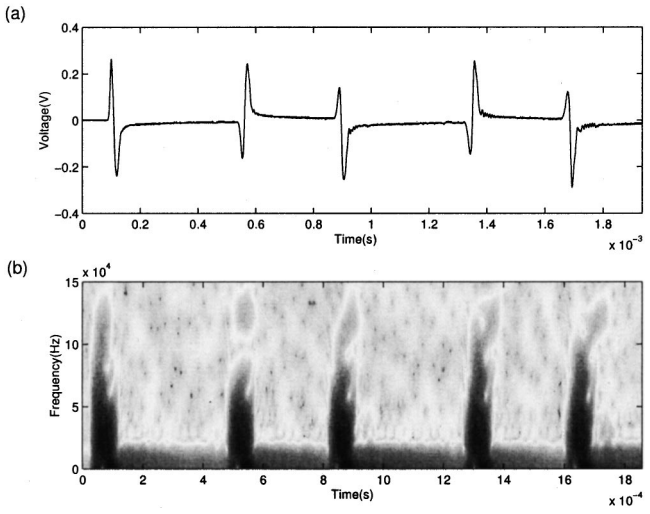


FIG. 14. The captured wave signal with the bias magnet configuration shown in Fig. 4. The time history of the signal (a) and its short-time Fourier transform (b).

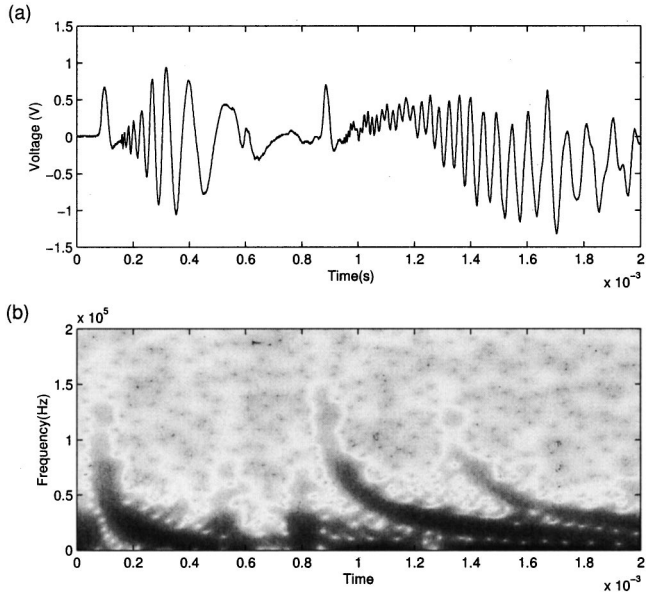


FIG. 15. The captured wave signal with the bias magnet configuration shown in Fig. 6. The time history of the signal (a) and its short-time Fourier transform (b).

We have also conducted the experiments by putting the sensor at some locations such as the lines 1 and 3 of Fig. 6, but we have obtained the best result only at the location of the line 2. This has already been predicted by the numerical analysis in the foregoing section. The use of yokes or other arrangements of bias magnetic field is expected to enhance the performance of flexural mode-sensing solenoid sensors. In this respect, extensive numerical optimization may be performed. As the final remark, we address that this mode-selectable sensor may be very useful to develop an advanced damage detection strategy. Work is in progress towards this direction.

V. SUMMARY

In this paper, we showed that either longitudinal or flexural waves alone can be extracted by a magnetomechanical sensor from a signal consisting of coupled longitudinal and flexural waves. To achieve the mode selection, two bias magnet configurations have been considered. In particular, the bias magnet configuration that allows the selection of flexural waves is newly proposed and its validity is experimentally verified. A simplified theory relating bias magnet arrangements and measured wave modes is also developed during the course of the present investigation.

ACKNOWLEDGMENT

This work is supported by the Brain Korea 21 project.

[1] E. Villari, ''Change of magnetization by tension and by electric current,'' Ann. Phys. (Leipzig) **126**, 87–122 (1865).
[2] J. P. Joule, ''On the effects of magnetism upon the dimensions of iron and steel bars,'' Philos. Mag. **30**, 76 (1847).
[3] R. C. Williams, ''Theory of magnetostrictive delay lines for pulse and continuous wave transmission,'' IEEE Trans. Ultrason. Eng. **UE-7**, 16–38 (1959).

[4] N. S. Tzannes, ''Joule and Wiedemann effects—The simultaneous generation of longitudinal and torsional stress pulses in magnetostrictive materials,'' IEEE Trans. Sonics Ultrason. **SU-13**, 33–41 (1966).
[5] J. O. Kim, Y. Wang, and H. H. Bau, ''The effect of an adjancent viscous fluid on the transmission of torsional stress waves in a submerged waveguide,'' J. Acoust. Soc. Am. **89**, 1414–1422 (1991).
[6] J. O. Kim and H. H. Bau, ''On line, real-time densimeter—Theory and optimization,'' J. Acoust. Soc. Am. **85**, 432–439 (1989).
[7] J. C. Aime, M. Brissaud, and L. Laguerre, ''Spatial analysis of torsional wave propagation in a cylindrical waveguide. Application to magnetostrictive generation,'' J. Acoust. Soc. Am. **109**, 51–58 (2001).
[8] M. Onoe, ''Theory of ultrasonic delay lines for direct-current pluse transmission,'' J. Acoust. Soc. Am. **34**, 1247–1254 (1962).
[9] H. Kwun, K. A. Bartels, and C. Dynes, ''Dispersion of longitudinal waves propagating in liquid-filled cylindrical shells,'' J. Acoust. Soc. Am. **105**, 2601–2611 (1999).
[10] H. Kwun, K. A. Bartels, and J. J. Hanley, ''Effects of tensile loading on the properties of elastic-wave propagation in a strand,'' J. Acoust. Soc. Am. **103**, 3370–3375 (1998).
[11] H. Kwun and K. A. Bartels, ''Experimental observation of elastic-wave dispersion in bounded solids of various configuration,'' J. Acoust. Soc. Am. **99**, 962–968 (1996).
[12] H. Kwun and K. A. Bartels, ''Magnetostrictive sensor technology and its application,'' Ultrasonics **36**, 171–178 (1998).
[13] H. Kwun and K. A. Bartels, ''Experimental observation of wave dispersion in cylinder shells via time-frequency analysis,'' J. Acoust. Soc. Am. **97**, 3905–3907 (1995).
[14] H. Kwun and C. M. Teller, ''Magnetostrictive generation and detection of longitudinal, torsional, and flexural waves in a rod,'' J. Acoust. Soc. Am. **96**, 1202–1204 (1994).
[15] R. Murayama, ''Driving mechanism on magnetostrictive type electromagnetic acoustic transducer for symmetrical vertical-mode Lamb wave and for shear horizontal-mode plate wave,'' Ultrasonics **34**, 729–736 (1996).
[16] R. Murayama, ''Study of driving mechanism on electromagnetic acoustic transducer for Lamb wave using magnetostrictive effect and application in drawability evaluation of thin steel sheet,'' Ultrasonics **37**, 31–38 (1999).
[17] D. C. Jiles and D. L. Atherton, ''Theory of the magnetization process in ferromagnets and its application to magnetomechanical effect,'' J. Phys. D **17**, (1984).
[18] W. J. Weaver, S. P. Timoshenko, and D. H. Young, *Vibration Problems in Engineering, 7th Ed.* (Wiley, New York, 1990).
[19] ANSYS/EMAG User Guide, Swanson Analysis System Inc., 1993.
[20] H. Lee, Ph.D. thesis, Seoul National University, Seoul, 2001.
[21] K. F. Graff, *Wave Motion in Elastic Solids* (Ohio State University Press, Columbus, OH, 1975).
[22] J. Zemanek, ''An experimental and theoretical investigation of elastic wave propagation in a cylinder,'' J. Acoust. Soc. Am. **51**, 265–283 (1972).
[23] Y. H. Pao and R. D. Mindlin, ''Dispersion of flexural waves in an elastic, circular cylinder,'' J. Appl. Mech. **27**, 513 (1960).
[24] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland/American, Elsevier, Amsterdam, New York, 1973).
[25] S. Mallat, *A Wavelet Tour of Signal Processing* (Academic, New York, 1998).
[26] Y. Y. Kim and E. H. Kim, ''Effectiveness of the continuous wavelet transform in the analysis of some dispersive elastic waves,'' J. Acoust. Soc. Am. (in press).
[27] T. Önsay and A. G. Haddow, ''Wavelet transform analysis of transient wave propagation in a dispersive medium,'' J. Acoust. Soc. Am. **95**, 1441–1449 (1994).

The effect of internal point masses on the radiation of a ribbed cylindrical shell

Martin H. Marcus[a] and Brian H. Houston
Naval Research Laboratory, Washington, DC 20375-5320

(Received 13 December 2001; revised 23 May 2002; accepted 23 February 2002)

A finite element analysis is performed on a submerged cylindrical shell with internal frames and point masses attached to the frames. A point force on the shell is shown to excite resonances of the frames. Without the point masses, most of these resonances are evanescent and do not radiate to the far field. With the point masses, the high circumferential order resonances couple with those at low circumferential order with the result of increasing radiation to the far field. In this investigation, the radiation increase is about 10 dB over a broad frequency range (1.5 to 3 times the ring resonance).
 [DOI: 10.1121/1.1497618]

PACS numbers: 43.40.Dx, 43.40.Ey, 43.40.Rj [MO]

I. INTRODUCTION

The effect of complex non-axisymmetric internal structure on the noise level of submerged cylindrical shells has received much interest in recent years. Strasberg[1,2] and Choi[3] analyze simple axisymmetric structures with internal spring-mass oscillators. Both observe a dampening effect in the acoustic radiation. Achenbach[4] takes a similar approach and determines that the radiation could be higher or lower, depending on where the applied load is relative to the spring mass system. Li[5] breaks the symmetry of an infinite cylindrical shell with a floor to show that radiation efficiency decreases with increased coupling. Laulagnet[6] also breaks the symmetry of an infinite cylindrical shell, but this time with a compliant layer on an arc of the shell. Under these circumstances, radiation increases at low frequencies. All of the above works pertain to analytical models that cannot be compared directly to experiments. Experiments on submerged, ribbed, cylindrical shells with hundreds of internal oscillators show how the radiation[7] and scattering[8] are modified by these oscillators. Bucaro[9,10] analytically applies these oscillators to a cylinder of the same shape as in the experiments to reveal angular spreading of the scattering pressures that is quite similar to what is seen in the experiments. Unfortunately, this work lacks elastic effects.

In this paper, a finite element analysis is performed on a submerged, ribbed, cylindrical shell with hundreds of internal masses to break the symmetry. It has already been shown that an axisymmetric, submerged, cylindrical shell with internal frames has resonances[11,12] only some of which radiate to the far field.[13] This is because some of the resonances have such a high circumferential order that they are evanescent. With the introduction of the internal masses, radiation increases. It is shown in this paper that these high circumferential modes couple into the low, and consequently radiating, circumferential modes. Under these circumstances, resonances from frames cause increased far-field radiation.

[a]Electronic mail: marcus@astro.nrl.navy.mil

II. FINITE ELEMENT COMPUTER MODELING

In this paper, the radiation from an axisymmetric shell with and without point masses is analyzed. The finite element model of the base axisymmetric shell shown in Fig. 1 has been previously reported.[14] It is a model of a cylindrical shell with 80 straight frames, a cone-sphere endcap, and made of nickel with an intrinsic loss of 0.001. The load is a radial point force placed halfway between frames 38 and 39 and is applied at 295 frequencies from 122 Hz to 36 kHz. Because the shell is axisymmetric, finite element program sara2d[15] was used. However, because the load is not axisymmetric, it is split into circumferential modes 0 through 30. Consequently, 31 finite element computations are performed across the range of frequencies indicated above. The shell has six quadratic elements per frame spacing, as well as four elements per frame, to make a total of 914 elements. Including the five layers of quadratic fluid elements, the overall model has 3884 elements. The mesh of the structure has 1829 nodes and the entire model has 10 739 nodes. The computations require 6.4 s for one frequency and circumferential mode on a Silicon Graphics power challenge computer.

It is appropriate to discuss the acoustic element technique used in this study. The finite elements on the outer surface of the structure are coupled to finite acoustic elements by including the acoustic pressure as right-hand sides in the structural finite element equations and by including the structural displacements as right-hand sides in the acoustic equations. Several layers of finite acoustic elements, in this study four layers, surround the structure. The number of layers is determined by what is necessary for mesh convergence at the highest frequency. As the frequency changes, the acoustic element sizes are automatically revised to maintain the element layer thickness as one-fourth of a wavelength. The final element layer has infinite acoustic elements[16] with a radial exponential decay relationship. This relationship satisfies the Sommerfeld radiation condition,[17] but does not decay as the inverse of the radius. Despite the approximation, the authors have demonstrated that the finite element software shows excellent agreement with several classical solutions and experimental data,[14] the ultimate test.

When point masses are included in the axisymmetric

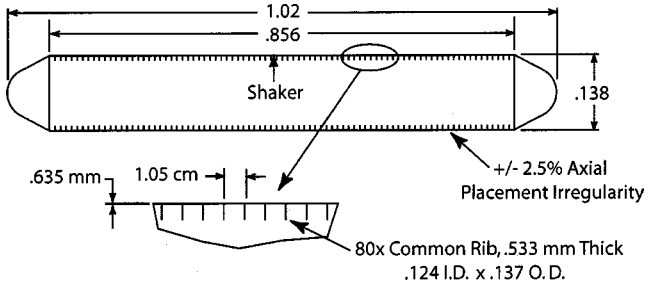


FIG. 1. Axisymmetric shell cross section.

shell, the structure is no longer axisymmetric. Although a three-dimensional finite element computer program can perform the analysis, it is most efficiently done using substructuring.[18,19] Using this method, the results for the axisymmetric shell computations are used to generate the matrices for an analysis of the three-dimensional structure. In this case, there are 11 equally spaced point masses attached to the inner edge of each frame. With 80 frames, this makes 880 nodes or 2640 degrees of freedom. This matrix must be inverted once for each of the 31 modes. This requires 57 min per frequency on a Silicon Graphics power challenge computer. Unfortunately, obtaining the axisymmetric results needed for the three-dimensional calculation requires 86 min per frequency, due to the storage requirements associated with the many degrees of freedom. In contrast, performing a conventional three-dimensional calculation would require at least an order of magnitude more in computer time, as well as storage.

In this paper, results from three different calculations are presented. The first is for the case of no point masses. In the second of these calculations, the masses are attached to the frames (with uniform spacing) and all weigh 3 g each. In the last calculation the uniform spacing is maintained, but the attached masses have a random distribution from 0 to 6 g.

III. AXISYMMETRIC RESULTS

The response to a point force on a submerged, framed cylindrical shell has been previously reported[14] and we briefly review this here. The computation yields surface normal velocities at many frequencies and circumferential modes. Along the cylindrical shell length, the velocity can be expressed as

$$v(x,\theta,t)=\sum_{j=0}^{n} v_j(x)\cos j\theta e^{i\omega t}, \tag{1}$$

where x is the axial position, n is the highest circumferential mode computed, ω is the radian frequency, θ is the angular position from the top, and t is time. The velocities are discrete, so a fast Fourier transform (FFT) along the length of the shell provides the axial mode components of these velocities as represented by v_{jk_x} in Eq. (2):

$$v(x,\theta,t)=\sum_{k_x=-m}^{m}\sum_{j=0}^{n} v_{jk_x}\cos j\theta e^{ik_x x}e^{i\omega t}. \tag{2}$$

The velocity versus axial wave number and frequency for circumferential mode zero $(n=0)$ is plotted in Fig. 2. The

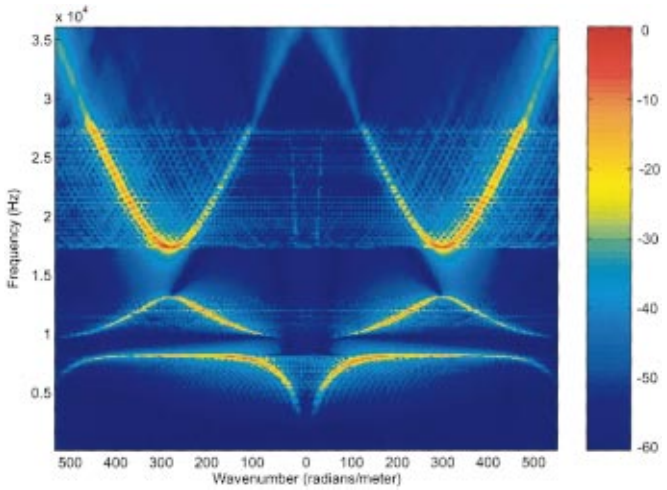


FIG. 2. Circumferential mode 0 component of the surface normal velocity of the axisymmetric shell from a point load. Colors refer to the dB level *re* 1 m/s for a 1 Newton load.

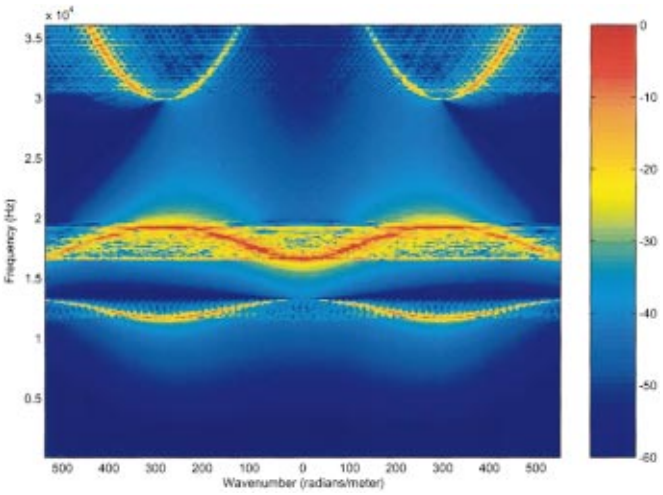


FIG. 3. Circumferential model 11 component of the surface normal velocity of the axisymmetric shell from a point load. Colors refer to the dB level *re* 1 m/s for a 1 Newton load.

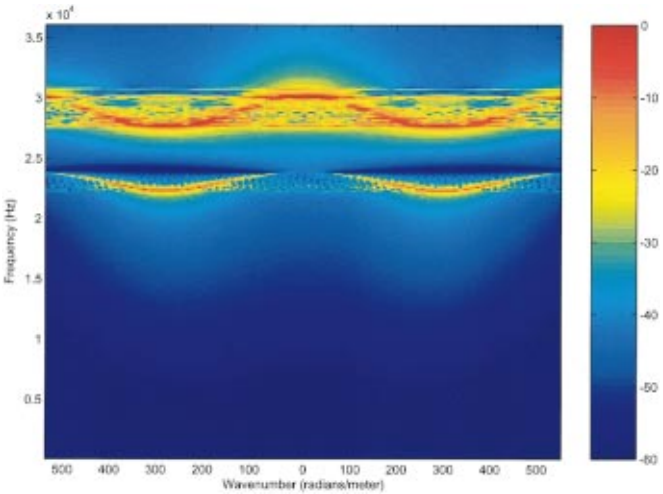


FIG. 4. Circumferential mode 22 component of the surface normal velocity of the axisymmetric shell from a point load. Colors refer to the dB level *re* 1 m/s for a 1 Newton load.

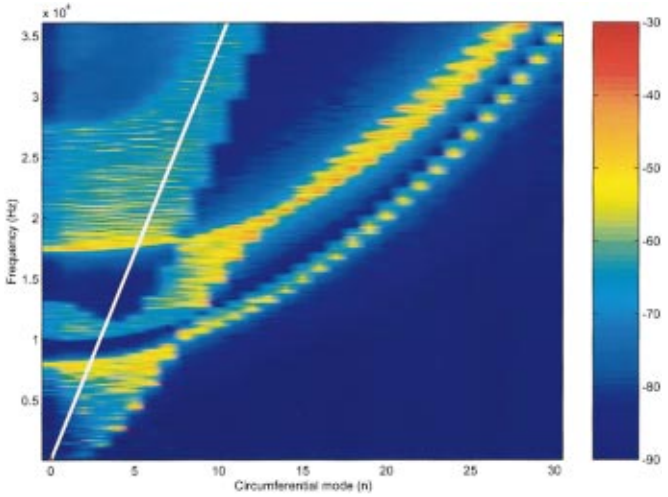


FIG. 5. The rms average of the surface normal velocity of the axisymmetric shell from a point load. Colors refer to the dB level *re* 1 m/s for a 1 Newton load. Velocities to the right of the white line do not radiate.

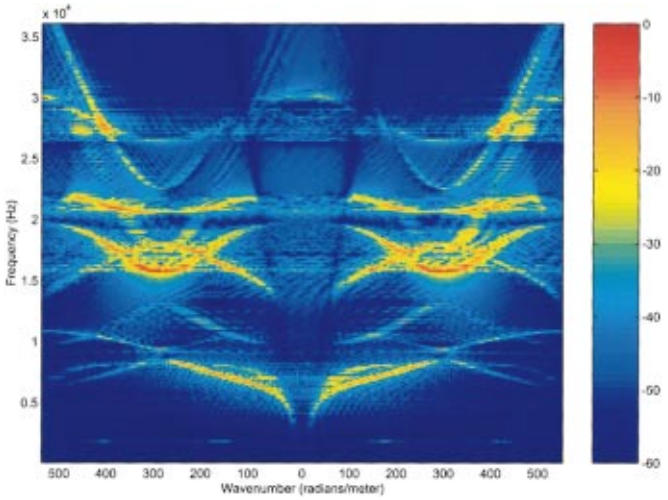


FIG. 7. Circumferential mode 0 component of the surface normal velocity of the uniform mass shell from a point load. Colors refer to the dB level *re* 1 m/s for a 1 Newton load.

first two pass bands, horizontal stripes at about 9 and 13 kHz, are associated with significant frame motion.[11,12,14] The third, much broader, stripe is associated with flexural shell motion scattered by the frames.[20–22] These pass bands appear for higher circumferential modes at higher frequencies.[14,23] For example, Fig. 3 shows another frequency wave number plot at circumferential mode 11 with similar pass bands appearing at about 13, 18, and 33 kHz. Figure 4 shows circumferential mode 22 pass bands at about 23 and 29 kHz.

Figures 2–4 will be of greater significance later in this paper, but for this section, they show the pass band frequencies monotonically increasing with circumferential mode, a trend that is most easily observed in Fig. 5. In Fig. 5, we present the rms average along the length of the shell of the surface normal velocities plotted against frequency and circumferential mode number. The straight line overlay indicates the ''sonic line'' for axial mode zero. Components with phase velocities to the right of this line are subsonic and do not radiate efficiently to the far field in the external fluid. Figure 6 shows the far-field pressure (in the plane and on the side containing the shaker) as a function of frequency and angle (measured from the end of the shell). As expected, it shows very little radiation outside of the pass bands in Fig. 2 and is especially high above the beginning of the third pass band, at about 17 kHz. The radiation levels are high across most of the angles from 17 to 27 kHz, after which it shows a strong dependence between frequency and angle due to the low-order axial Bloch modes becoming supersonic.[7,23]

IV. NON-AXISYMMETRIC RESULTS AND DISCUSSION

The response to a radial point force located halfway between frames 38 and 39 is computed for the same shell, now with 3-g point masses attached to the inner edges of the frames as well as for the shell with a random mass distribution similarly attached. Henceforth, we will call them the uniform mass shell and random mass shell, respectively. The circumferential mode 0 axial FFT of the surface normal velocities is displayed in Figs. 7 and 8 for the uniform mass shell and random mass shell, respectively. Figure 7 shows

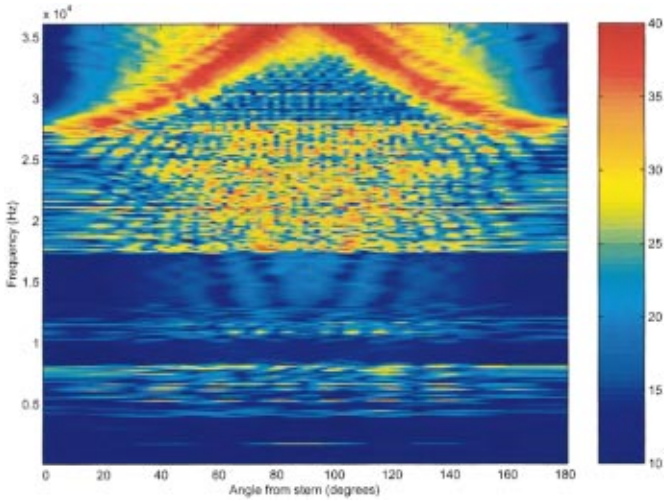


FIG. 6. Far-field radiated pressure from a point load on the axisymmetric shell. Colors refer to the dB level *re* 1 Pa for a 1 Newton load.

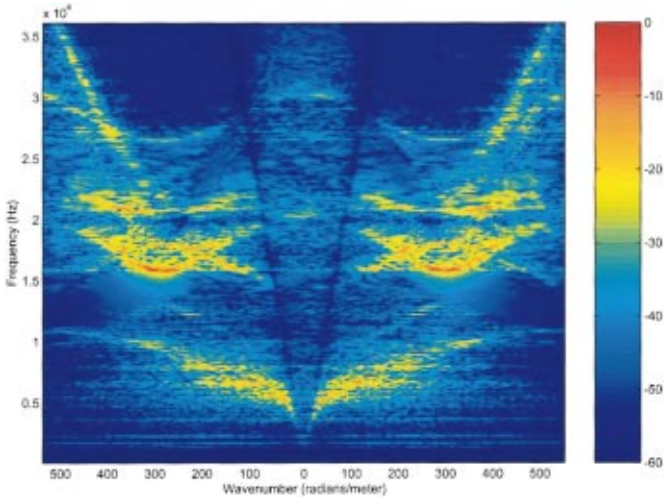


FIG. 8. Circumferential mode 0 component of the surface normal velocity of the random mass shell from a point load. Colors refer to the dB level *re* 1m/s for a 1 Newton load.

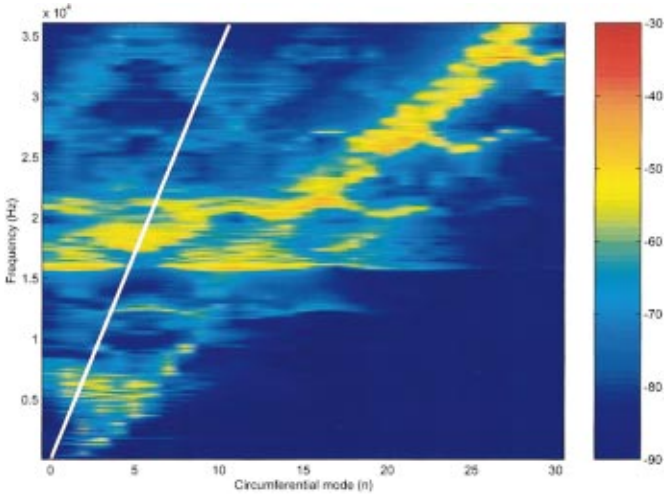


FIG. 9. The rms average of the surface normal velocity of the axisymmetric shell from a point load. Colors refer to the dB level *re* m/s for a 1 Newton load. Velocities to the right of the white line do not radiate.

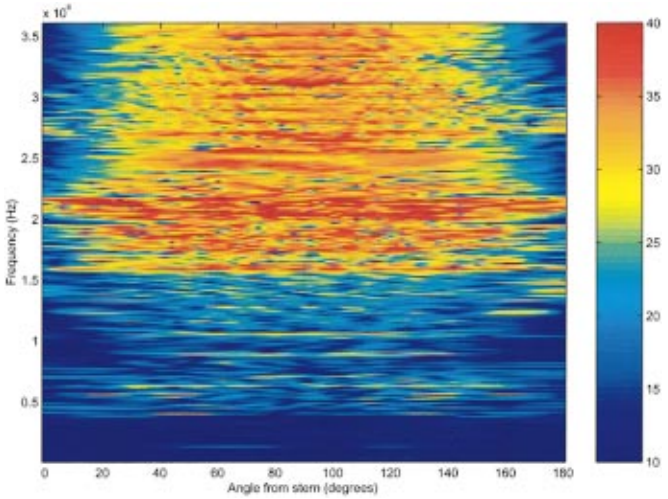


FIG. 10. Far-field radiated pressure from a point load on the uniform mass shell. Colors refer to the dB level *re* 1 Pa for a 1 Newton load.

evidence of a superposition of dispersions that correspond to the axisymmetric shell modes 0, 11, and 22 in Figs. 2–4. Since there are 11 equally spaced masses on each rib, we expect and observe the strongest coupling for mode 11 and its multiples in Figs. 7 and 8. Clearly, the presence of point masses in both the uniform and the random mass shell cause strong coupling among the circumferential modes. Figure 8 looks like a slightly disordered version of Fig. 7. In this case, we should expect these modes to couple, but not as strongly; we would also expect participation from additional circumferential orders.

The most dramatic observation to make about Figs. 7 and 8 is the distinctly different character that is seen when compared to Fig. 2. The addition of the point masses causes circumferential modes to come about that are supersonic and that are not found on the axisymmetric structure. This is a result of coupling between the higher and the lower n's through the attachment of discontinuities. Phrased another way, the masses introduce additional modes that allow leakage of energy from the high circumferential modes to the low modes. The above observations would suggest that the degree of radiation from the non-axisymmetric structures should be higher than that for the axisymmetric case.

Figure 9 gives the rms average along the length of the surface normal velocities versus frequency and circumferential mode number for the uniform mass shell. The straight line overlay indicates the ''sonic line'' for axial mode 0. Above 6 kHz, velocities to the left of this line are somewhat higher than they are in Fig. 5, the axisymmetric case. Since these velocities are expected to radiate efficiently to the far field, we should see more radiation above 16 kHz for the uniform mass shell than for the axisymmetric shell.

Figure 10 shows the uniform mass shell radiated pressure versus frequency and angle. Figures 10 and 6 are significantly different. The degree of radiation has generally increased across the band; however, the most dramatic difference occurs within the third (Bloch) pass band. In addition to the cutoff frequency for this pass band being lowered from 17.5 to 16 Hz, assumedly due to mass loading, the angular distribution and levels within the pass band have been modified. The onset of Bloch radiation at 27.5 kHz

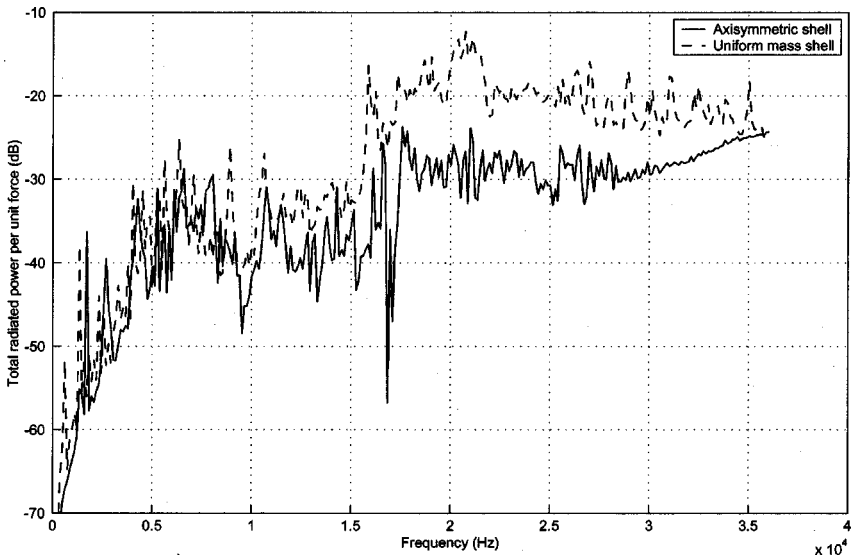


FIG. 11. Total radiated power from a point load on the axisymmetric shell and the uniform mass shell. Magnitude refers to the dB level *re* 1 W for a 1 Newton load.

(eminating from the ends of the cylinder) has disappeared in the uniform mass shell and has been replaced with a less spatially coherent, broad-angled, radiation structure that is more concentrated away from the ends. The random mass radiation result is similar to Fig. 10 and need not be shown.

Perhaps a simpler and more quantitative way of looking at the far-field radiated pressure is by plotting the total radiated power. This is computed by integrating the square of the far-field pressure on the surface of a far-field sphere and is performed for each circumferential mode separately, then the contributions from each mode are added together. The power is displayed in Fig. 11 versus frequency for both the axisymmetric shell and uniform mass shell. We observe that from about 16 to 36 kHz, the uniform mass shell radiates about 10 dB more than does the axisymmetric shell. It is clear that this is due to high circumferential order resonances that are coupled to the lower n's.

V. CONCLUSIONS

In this paper, we illustrate in a preliminary way the importance of non-axisymmetric structure in the radiation physics of fluid-loaded framed cylindrical shells. With the inclusion of internal point masses, our finite element computations reveal a significant restructuring of the normal modes with an increase in radiation over that of an axisymmetric shell. The radiation shows its most striking differences at frequencies that correspond to rib resonances that are accentuated by the placement pattern of the point masses. This implies that rib resonance energy that results in evanescent motion for axisymmetric shells gains access to the fluid by the breaking of symmetry. Though the extent that these results generalize is presently not known, particularly with regard to the amount of structural damping, the fundamental nature of the change in the basic physics of axisymmetric versus non-axisymmetric shells is clear. In this investigation, the radiation increase is about 10 dB over a broad frequency range (1.5 to 3 times the ring resonance).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Office of Naval Research for support of this work.

[1] M. Strasberg and D. Feit, ''Vibration damping of large structures induced by attached small resonant structures,'' J. Acoust. Soc. Am. **99**, 335–344 (1996).
[2] M. Strasberg, ''Continuous structures as fuzzy substructures,'' J. Acoust. Soc. Am. **100**, 3456–3459 (1996).
[3] S. H. Cho, T. Igusa, and J. D. Achenbach, ''Acoustic radiation from a finite-length shell with non-axisymmetric substructures using a surface variational principle,'' J. Sound Vib. **197**(3), 329–350 (1996).
[4] J. D. Achenbach, J. Bjarnason, and T. Igusa, ''Effect of a vibrating substructure on acoustic radiation from a cylindrical shell,'' Trans. ASME, J. Vib. Acoust. **114**, 312–318 (1992).
[5] D. S. Li, L. Cheng, and C. M. Gosselin, ''Analysis of structural acoustic coupling of a cylindrical shell with an internal floor partition,'' J. Sound Vib. **250**(5), 903–921 (2002).
[6] B. Laulagnet and J. L. Guyader, ''Sound radiation from finite cylindrical shells, partially covered with longitudinal strips of compliant layer,'' J. Sound Vib. **186**(5), 723–742 (1995).
[7] D. M. Photiadis, B. H. Houston, A. J. Bucaro, and E. G. Williams, ''Resonant responses of and acoustic radiation from complex shell structures,'' J. Acoust. Soc. Am. **108**, 1027–1035 (2000).
[8] D. M. Photiadis, J. A. Bucaro, and B. H. Houston, ''The effect of internal oscillators on the acoustic response of a submerged shell,'' J. Acoust. Soc. Am. **101**, 87–92 (1997).
[9] J. A. Bucaro, A. Romano, A. Sarkissian, D. M. Photiadis, and B. H. Houston, ''Local admittance model for acoustic scattering from a cylindrical shell with many internal oscillators,'' J. Acoust. Soc. Am. **103**, 1867 (1998).
[10] J. A. Bucaro, D. M. Photiadis, and B. H. Houston, ''The effect of internal oscillators on the acoustic response of a submerged shell,'' in *Proceedings of the ASME, Noise Control and the Acoustics Division* (ASME, New York, 1997), NCA 22, pp. 895–899.
[11] M. H. Marcus and A. Sarkissian, ''Rib resonances present in the scattering response of a ribbed cylindrical shell,'' J. Acoust. Soc. Am. **103**, 1864–1866 (1998).
[12] M. H. Marcus, ''The effect of rib resonances on the vibration and wave scattering of a ribbed cylindrical shell,'' J. Acoust. Soc. Am. **103**, 3055(A) (1998).
[13] E. G. Williams, ''Supersonic acoustic intensity on planar sources,'' J. Acoust. Soc. Am. **104**, 2845–2850 (1998).
[14] M. H. Marcus, B. H. Houston, and D. M. Photiadis, ''Wave localization on a submerged cylindrical shell with rib aperiodicity,'' J. Acoust. Soc. Am. **109**, 865–869 (2001).
[15] H. Allik, *Sara-2d user's manual, version 95-3*, 1995, BBN Systems and Technologies, New London, CT.
[16] O. C. Zienkiewicz, *The Finite Element Method* (McGraw–Hill, New York, 1994), pp. 660–664.
[17] A. D. Pierce, *Acoustics* (Acoustical Society of America, New York, 1991), pp. 177–178.
[18] H. Allik, R. N. Dees, and S. W. Moore, ''Efficient structural acoustic analysis using finite and infinite elements,'' 1995, ASME 15th Biennial Conference on Vibration and Noise, 1995, Vol. 3, part B, pp. 87–92.
[19] H. Allik, *Sara-3d user's manual*, 1992, BBN Systems and Technologies, New London, CT.
[20] C. H. Hodges, J. Power, and J. Woodhouse, ''The low frequency vibration of ribbed cylinder, part one: Theory,'' J. Sound Vib. **101**(2), 219–235 (1985).
[21] D. M. Photiadis, ''The effect of irregularity on the scattering of acoustic waves from a ribbed plate,'' J. Acoust. Soc. Am. **91**, 1897–1903 (1992).
[22] L. Brillouin, *Wave Propagation in Periodic Structures* (Dover, New York, 1946).
[23] B. Houston, D. M. Photiadis, J. A. Bucaro, and H. J. Simpson, ''Vibratory states and radiation from a framed cylindrical shell with internal complexity,'' ASME International Mechanical Engineering Congress and Expo., November 1997, NCA Vol. 24, pp. 25–35.

Acoustic characteristics of twin jets

F. He and X. W. Zhang
Department of Engineering Mechanics, Tsinghua University, Beijing 100084, People's Republic of China

(Received 4 February 2002; revised 15 May 2002; accepted 16 May 2002)

Experiments were conducted to investigate the acoustic characteristics of underexpanded supersonic twin jets in different azimuthal measurement planes. Compared with two independent jets, the twin jets produced additional noise due to the enhanced mixing and entrainment. The larger pressure ratio for switching from the axisymmetric mode to the helical mode led to lower noise levels at 90° than for two independent jets. For pressure ratios greater than 5.00, the noise reduction was due to cessation of screeching of the twin jets while screeching of a single jet was still detected. The apparent shielding phenomenon was measured for the screech helical mode. The screech tone intensities were attenuated largely due to the shielding effects. The noise reductions due to shielding were obtained over a wide range of pressure ratios relative to the sum of two independent jets. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1492818]

PACS numbers: 51.40.+p, 43.28.+h, 51.35.+a [MSH]

I. INTRODUCTION

Twin jets are widely used in flight propulsion systems as well in other engineering applications such as painting, dust removal and cooling. In practice the reduction of jet exhaust noise is a very important issue.[1] Most work on low-speed twin jets has been performed with incompressible turbulent jets but there have been very few studies of high-speed twin jets. Two jets can couple to produce higher dynamic pressures, which can cause structural fatigue.[2–6] Norum[2] found that twin jet acoustic levels that propagate upstream of the exhaust nozzles can exceed the single jet levels by more than 20 dB. For closely spaced nozzles, coupling occurred at low Mach numbers but was suppressed at high Mach numbers.[3] For twin jets, Umeda observed two counter rotating helical oscillation modes and reversed lateral oscillation modes along the center line of the two jets for two different pressure ratios.[7] The acoustic features of rectangular supersonic jets were investigated by Raman *et al.*[4] Supersonic multi-jet interaction is likely to be important to enhance mixing and noise reduction in supersonic mixer-ejector nozzles. The peak jet noise source moves upstream and radiates noise at larger angles to the flow direction.[8] Noise radiation from a source can be shielded by an adjacent jet, as was recognized in the 1970s[9–11] where early work focused on engine arrangement studies. The governing mechanisms for shielding are refraction and diffraction of sound by the jet flow.[12] Predictions for acoustic shielding have been based on an idealized cylindrical uniform jet model in the absence of aerodynamic interaction.[13–16]

This paper presents experimental results for the acoustic characteristics of twin jets in different azimuthal measurement planes. The objective is to investigate the noise reduction due to interaction or shielding.

II. EXPERIMENTAL APPARATUS

Screw compressors were used to provide high-pressure airflow to a settling chamber with two axisymmetric circular exhaust nozzles with diameters of $d=6$ mm and length to diameter ratios of 3. The spacing between the twin nozzles was $s=1.5d$ or $s=2d$. The stagnation/ambient pressure ratios p_0/p_a were set by controlling a regulating valve. Sound measurements were conducted inside a semi-anechoic chamber 3.2 m×3.2 m×2.2 m. An ACO7610 microphone, ACO 6139 microphone amplifier and an FFT CF5220 analyzer were used to measure the noise. The microphone had an open circuit frequency response of up to 100 kHz and was mounted on a support that pivoted around an axis passing through the jet centerline. The microphone was set at 15° intervals starting from an angle 15° from the jet axis and 1 m from the center of the two nozzles. Figure 1 shows the azimuthal measurement planes in the $\psi=90°$ x-y plane containing both jet axes and in the $\psi=0°$ x-z plane perpendicular to both jet axes and containing the jet centerline along the x axis. The single jet tests were run on the same apparatus.

III. TWIN JET SCREECH MODES

For imperfect supersonic jets, the discrete fundamental frequencies for various screech modes decrease with increasing pressure ratios and eventually switch from one mode to another. The flow structure and the radiation levels are associated with the screech modes.[17] Figure 2 shows the screech modes for one and two jets, where $S_t=f\,d/u$, f is the fundamental frequency, d is the nozzle diameter and u is the jet mean velocity. The fundamental frequency and modes for two jets were similar to those for a single jet; however, the following differences can be seen. (1) The pressure ratios to switch from mode B to mode C were greater for two jets.

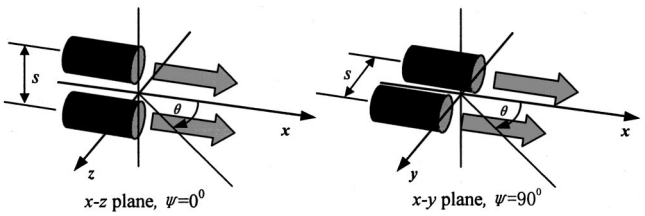


FIG. 1. Sketch of azimuthal measurement planes.

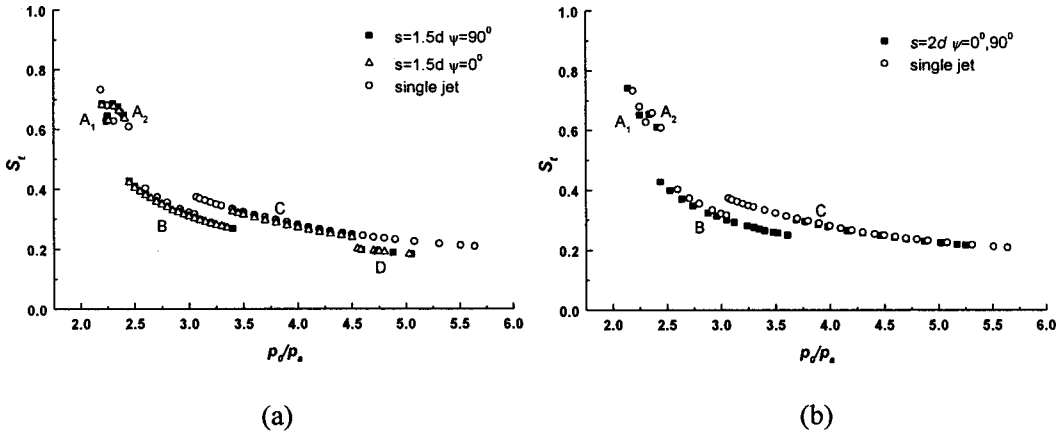


FIG. 2. Fundamental frequencies and mode switching. A_1, A_2, B, C and D are different screech modes.

The same phenomenon, which did not arouse attention, was also seen in Thomas's experimental data.[1] (2) The five stages for the two closely spaced jets labeled A_1, A_2, B, C and D in Fig. 2(a), were more than the four stages for a single jet. (3) The pressure ratios corresponding to cessation of the screeching differed for the two different measurement planes, $\psi=0°$ and $\psi=90°$, but both were lower than that of the single jet. Table I compares the pressure ratios corresponding to cessation of the screeching. Figure 3 shows that for twin jets at a pressure ratio $p_0/p_a=5.00$, screeching was not detected at $\psi=90°$ but was detected at $\psi=0°$.

IV. TWIN JET INTERACTION

A. Acoustic directivity for $p_0/p_a<3.00$

The turbulent eddies in the shear layer are formed by entraiment. The quadrupole radiation is then generated by the fluctuating turbulent stresses. For $p_0/p_a<3.00$, the acoustic features of the twin jets show that turbulent mixing noise from the quadrupole sound sources dominate, even when the discrete frequencies of screech modes A_1, A_2, and B are detected at both $\psi=0°$ and $\psi=90°$, as shown in Fig. 4.

B. Coupling interaction

With large pressure ratios, the directivity is related to the orientation of the azimuthal measurement plane. The sound measurements at $\psi=0°$ reflect acoustic coupling of the two jets.

Figure 5 shows the directivities in the azimuthal measurement plane $\psi=0°$. Because of the enhanced noises associated with the shock structure, the directivity function is not the same as for subsonic or low-underexpanded twin jets as shown in Fig. 4. The data are also compared with two independent jet data which equal to a single nozzle data added by 3 dB, as shown in Fig. 6. No distinctive features were found for both $s=1.5d$ and $s=2d$. The difference of the nozzle spacing affects the coupling interaction. Figure 7 shows the sound pressure level over the full range of pressure ratios at the directivity angles $\theta=30°$, 90° and 135° for $s=1.5d$,$2d$, respectively. In Fig. 7(a) for $s=1.5d$, the noise decreases when the screech mode switches to the mode D for $p_0/p_a=4.40-4.52$. However, for this range of pressure ratios, the mode C still dominates the single jet noise. In Fig. 7(b), for $s=2d$, the noise is greater than for a single jet $+3$ dB for pressure ratios from 3.00 to 4.50.

TABLE I. Pressure ratios corresponding to screeching cessation.

	$s=1.5d$		$s=2d$		Single jet
azimuthal planes	$\psi=0°$	$\psi=90°$	$\psi=0°$	$\psi=90°$	—
p_0/p_a	>5.41	>5.28	>5.33	>5.00	>5.70

The enhanced mixing and entrainment generates additional noise. The larger pressure ratios for switching from axisymmetric mode B to helical mode C leads to higher noise levels for $\theta=90°$ than for the single jet for both $\psi=0°$ and $\psi=90°$. For pressure ratios $p_0/p_a>5.00$, the noise levels decrease as shown in Fig. 7 which is when the screeching of the twin jets ended but the screeching of the single jet still continued.

C. Acoustic shielding

The sound measurements at $\psi=90°$ reflect the two jets shielding effect. One jet acts as a partial barrier between the other jet and the receiver.

For subsonic and low-underexpanded supersonic twin jets, no shielding effect was detected at either $\psi=0°$ or $\psi=90°$.

The directivities were quite different for $\psi=0°$ and $\psi=90°$ for the supersonic twin jets. The twin jet shielding effect is shown in Fig. 8 as a plot of the difference ΔL_p, between the directivities at $\psi=0°$ and 90°. $\Delta L_p<0$ indicates a shielding effect. There was little or no shielding effect for underexpanded supersonic jets with screech modes A_1, A_2, B, or D, but shielding was apparent for screech mode C. Noise reductions due to shielding of 2 to 10 dB were measured for angles $\theta<70°$ and $\theta>100°$.

Figure 9 compares the frequency spectra for $\psi=0°$ and $\psi=90°$. The intensities of the screeching at $\psi=90°$ are significantly reduced by the shielding effect.

Figure 10 shows that the sound pressure level generally increased with increasing pressure ratio for $\theta=30°$, 90° and

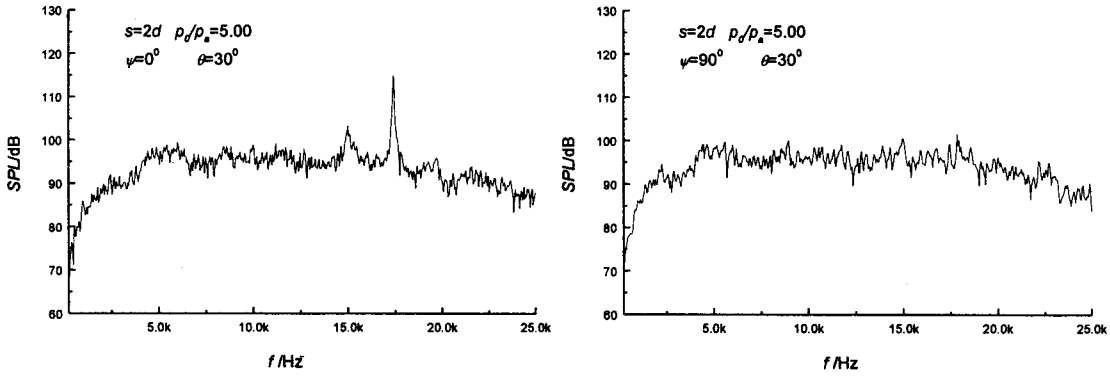


FIG. 3. For twin jets, screeching was not detected at $\psi=90°$ but was detected at $\psi=0°$.

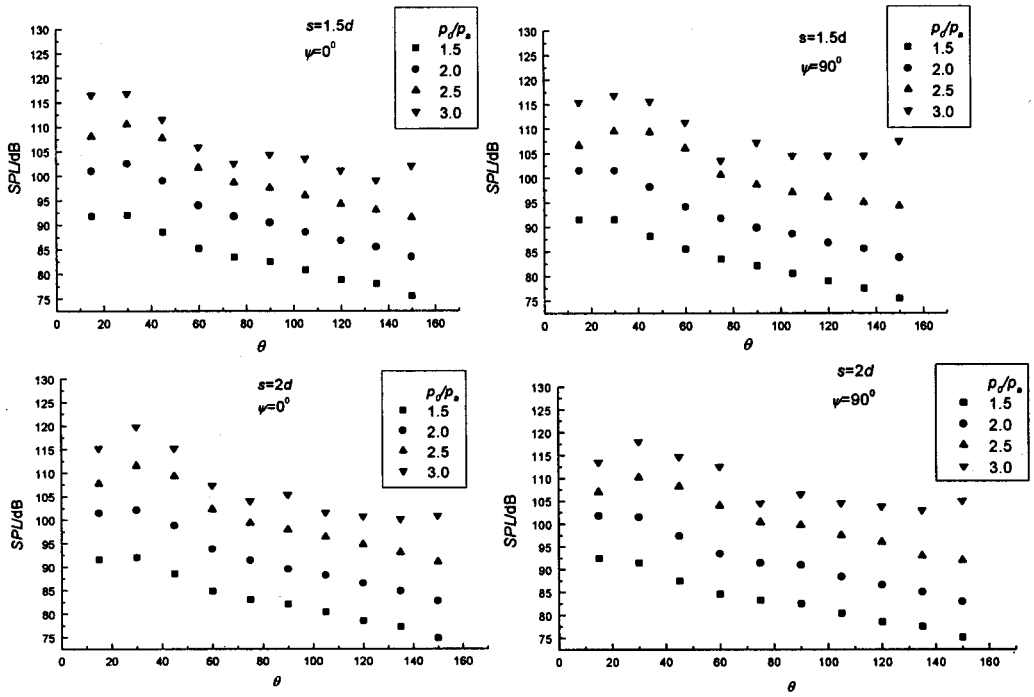


FIG. 4. Acoustic directivities for $p_0/p_a<3.00$.

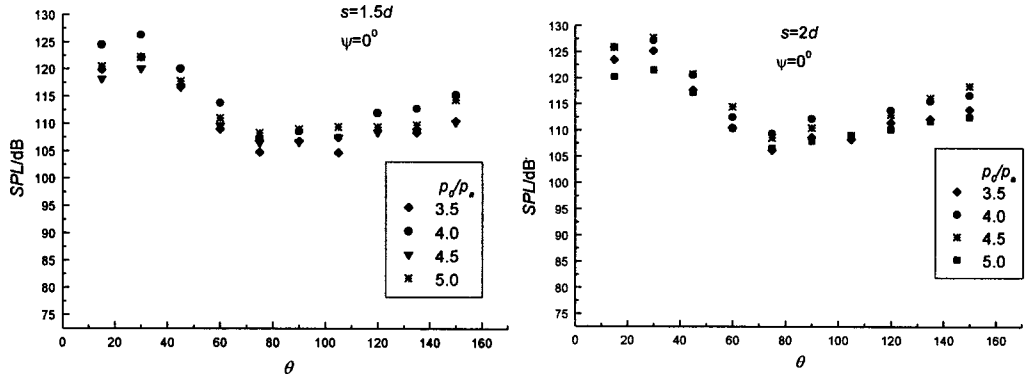


FIG. 5. Directivities in the azimuthal measurement plane $\psi=0°$ for $p_0/p_a>3.00$.

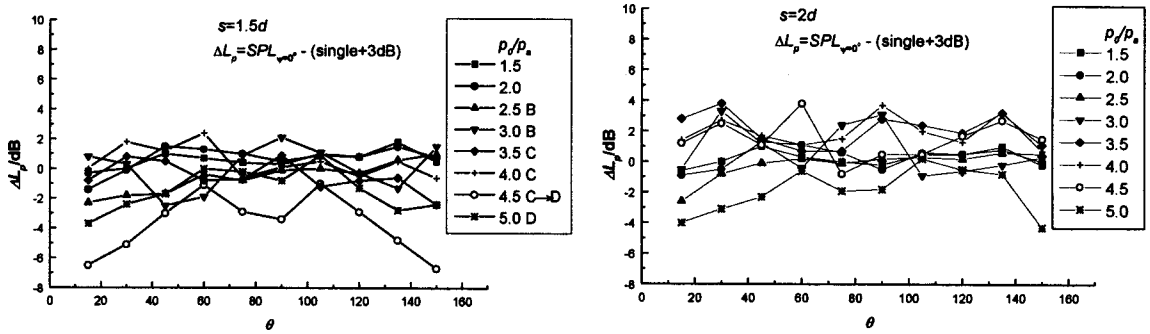


FIG. 6. Comparison of twin jets noise levels and two independent jets (single jet data+3 dB).

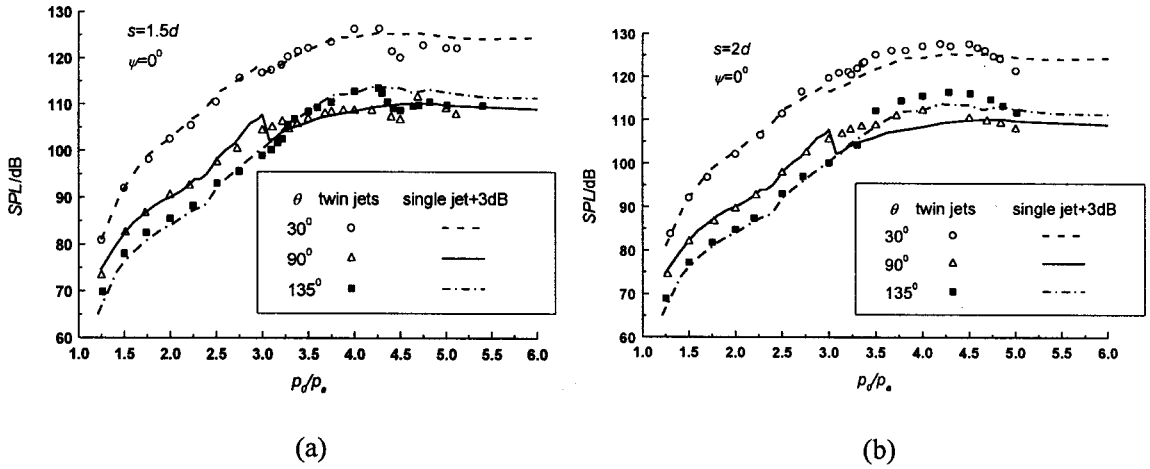


FIG. 7. Sound pressure levels for twin jets and two independent jets over the full range of pressure ratios at $\psi=0°$.

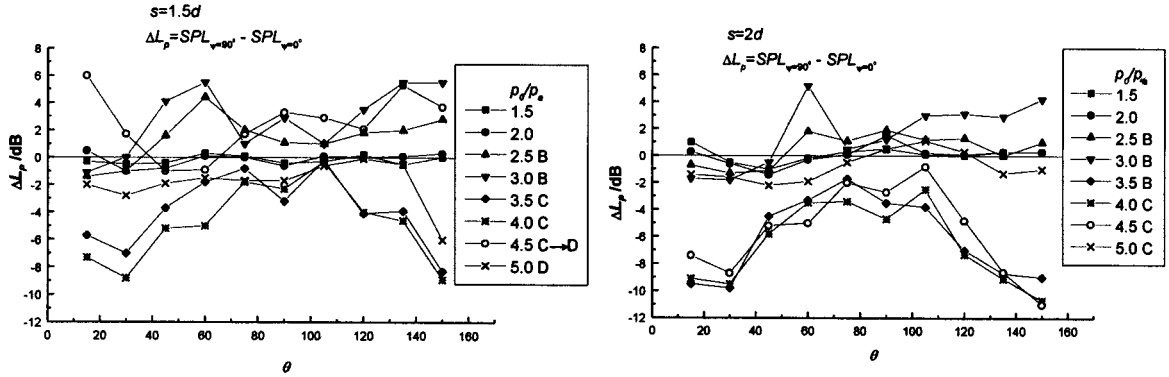


FIG. 8. Differences in the directivity function between $\psi=0°$ and 90°.

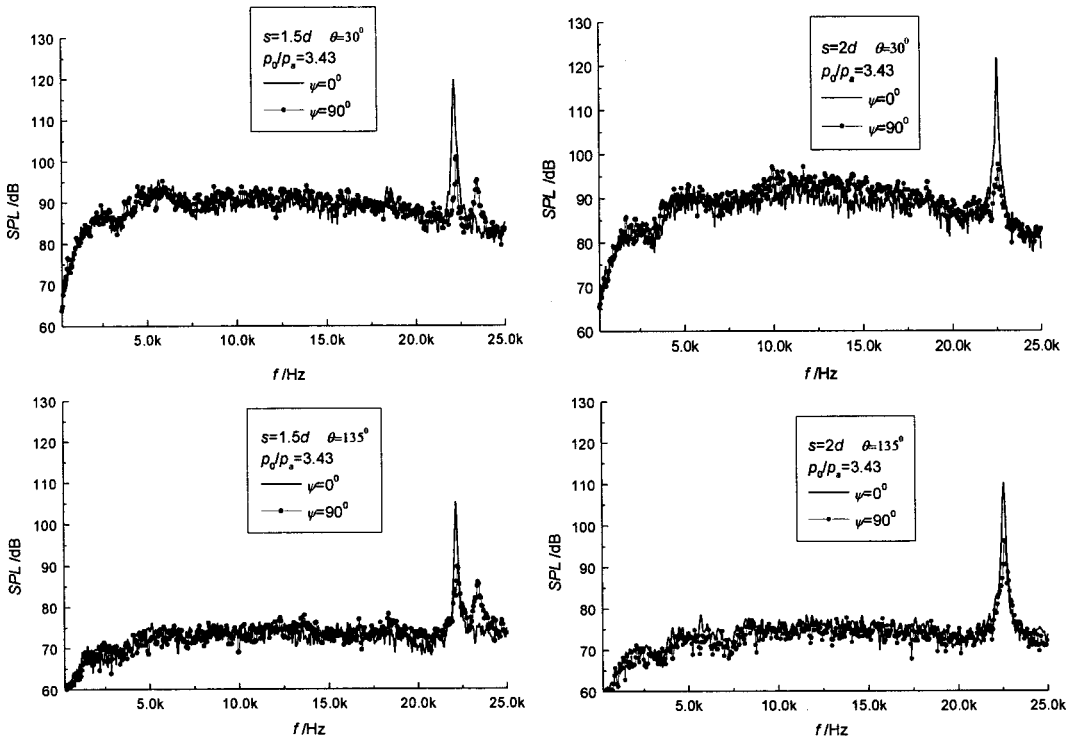


FIG. 9. Screeching intensity reduction at $\psi=90°$ due to shielding.

135°. Except for mode C, where there is little shielding effect, the noise level decreased only at the mode switching pressure ratios, such as switching from mode B to mode C, or from mode C to mode D. Hence, for a wide range of pressure ratios, the two jets together will make less noise than the sum of two independent jets. The differences of the sound pressure levels, ΔL, at the mode switching points are listed in Table II.

V. CONCLUSIONS

For twin jets, the pressure ratios for switching from the axisymmetric to the helical mode are higher than for a single jet. The larger switching pressure ratios result in noise levels at $\theta=90°$ which are greater than for two independent jets. The pressure ratios at which the twin jets no longer screech differ in the different measurement planes. Twin jets can be used for noise attenuation since the screeching of the twin jets stopped at a lower pressure ratio than for a single jet. However, the enhanced mixing and entrainment caused additional noise compared with two independent jets.

Shielding occurred for underexpanded supersonic twin jets, especially for the helical screech mode C. The shielding most effectively reduced the screech intensities in the upstream and downstream directions approaching the jet axis. The noise reductions due to shielding are achieved over a wide range of pressure ratios compared with the noise of two independent jets.

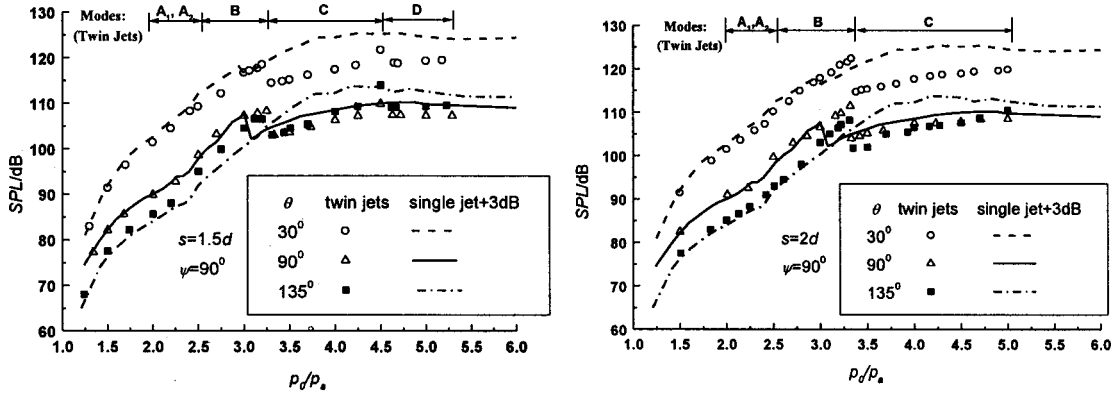


FIG. 10. Sound pressure levels for twin jets and two independent jets over the full range of pressure ratios at $\psi=90°$.

TABLE II. Sound pressure level differences between modes.

	$s=1.5d$			$s=2d$		
θ	30°	90°	135°	30°	90°	135°
$\Delta L_{B\text{-}C}$ (dB)	4.5	5.3	3.0	7.7	7.5	6.4
$\Delta L_{C\text{-}D}$(dB)	2.8	2.6	4.4	–	–	–

ACKNOWLEDGMENTS

The authors would like to thank their colleagues Professor X. F. Wang and D. Q. Yang for their interest and constructive criticism of this work.

[1] M. R. Green and D. G. Crighton, ''Instability properties of interacting jets,'' J. Fluid Mech. **350**, 331–349 (1997).
[2] T. D. Norum and J. G. Shearin, ''Dynamics loads on twin jet exhaust nozzles due to shock noise,'' J. Aircr. **23**, 728–729 (1986).
[3] R. W. Wlezien, ''Nozzle geometry effects on supersonic jet interaction,'' AIAA J. **27**, 1361–1367 (1989).
[4] G. Raman and R. Taghavi, ''Coupling of twin rectangular supersonic jets,'' J. Fluid Mech. **354**, 123–146 (1998).
[5] G. H. Mostafa, ''Studies on the flowfield of multijets with square configuration,'' AIAA J. **31**, 1189–1190 (1993).
[6] J. M. Seiner, J. C. Manning, and M. K. Ponton, ''Dynamic pressure loads associated with twin supersonic plume resonance,'' AIAA Pap. **86–1539** (1986).
[7] Y. Umeda and R. Ishii, ''Oscillation modes of supersonic multijets,'' J. Acoust. Soc. Am. **101**, 3353–3360 (1997).
[8] D. E. Zilz and R. W. Wlezien, ''The sensitivity of near-field acoustic to the orientation of twin two-dimensional supersonic nozzle,'' AIAA Pap. **90–2149** (1990).
[9] D. S. Jones, ''The mathematical theory of noise shielding,'' Prog. Aerosp. Sci. **17**, 149–229 (1977).
[10] S. J. Cowan and R. W. Crouch, ''Transmission of sound through a two-dimensional shielding jet,'' AIAA Pap. **73–1002** (1973).
[11] M. S. Howe, ''The attenuation of sound by a randomly irregular impedance layer,'' Proc. R. Soc. London, Ser. A **347**, 513–535 (1976).
[12] J. C. Yu and D. J. Fratello, ''Measurement of acoustic shielding by a turbulent jet,'' J. Sound Vib. **98**, 183–212 (1985).
[13] C. H. Gerhold, ''Analytical model of jet shielding,'' AIAA J. **21**, 694–698 (1982).
[14] T. F. Balsa, ''The shielding of a convected source by an annular jet with an application to the performance of multitube suppressors,'' J. Sound Vib. **44**, 179–189 (1976).
[15] R. A. Kantila, ''Acoustic properties of heated twin jets,'' J. Sound Vib. **79**, 79–106 (1981).
[16] J. S. Clauss, Jr., B. R. Wright, and G. E. Bowie, ''Twin-jet noise shielding for a supersonic cruise vehicle,'' J. Aircr. **17**, 627–632 (1980).
[17] Y. Umeda and R. Ishii, ''On the sound sources of screech tones radiated from choked circular jets,'' J. Acoust. Soc. Am. **110**, 1845–1858 (2001).

A dimensionless mobility formulation for evaluation of force and moment excitation of structures

A. T. Moorhouse
Acoustics Research Unit, University of Liverpool, Liverpool L69 3BX, United Kingdom

(Received 2 November 2001; revised 24 May 2002; accepted 30 May 2002)

The problem addressed is the simultaneous excitation of a structure by a force and moment. These two forms of excitation are dimensionally incompatible, and so their relative importance is normally difficult to assess. Here, a dimensionless form of the mobility matrix is developed thereby allowing proper comparison. The off-diagonal elements of the matrix quantify the coupling between two excitations of whatever dimensions, in other words the extent to which an applied force or moment affects the power input by another force or moment. It is shown that off-diagonals of unit magnitude indicate the maximum possible extent of coupling between the corresponding excitations. A precise definition of weak and strong coupling is therefore possible. Illustrative examples of dimensionless mobilities for infinite, semi-infinite, and finite plates are given. It is also shown that the eigenvalues of the dimensionless mobility matrix relate the actual delivered power to the ''point-by-point'' power input. The results are shown to have wider relevance than originally intended; specifically, the dimensionless mobilities provide a measure of the suitability for active vibration control of different points on a structure, or mixed excitations. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1497622]

PACS numbers: 43.40.−r, 43.40.Vn, 43.40.At, 43.40.Dx [EGW]

I. INTRODUCTION

The problem addressed in this paper is the simultaneous excitation of a structure by one or more forces and moments. Often in structural analysis only force terms are included and moments neglected. However, this is usually motivated by a desire to avoid measurement difficulties, rather than by a knowledge that the effect of moments is insignificant. Some of the difficulties of measurement are evident from Refs. 1–3. Understanding of the role of moments has significantly improved as a result of studies such as Refs. 4–7. However, it is likely that rotational excitations and responses will always be more difficult to handle experimentally than forces. An improved understanding of the relative importance of moments would help to identify situations where they can be safely ignored and those where there is no alternative but to measure.

One barrier to understanding is that when both forces and moments are included in an analysis the mobility matrix is made up of dimensionally incompatible elements. The significance of the various elements cannot then be directly compared. The objective of this paper is to present a dimensionless form of the mobility matrix so as to facilitate comparison between elements, even when they have different units. Examples of calculated mobilities for various idealized structures are given, first to illustrate the dimensionless mobility and second to provide insight into the behavior of such structures. Additionally, although not the original intention, this formulation also turns out to provide some more general insight into coupling between any degrees of freedom. Therefore, an analysis is also included of the transfer coupling between dimensionally compatible elements (two translations or two rotations for example).

The basis of the formulation is that forces and moments can be compared in terms of power: it is meaningless to compare a force and a moment directly, but their relative importance in exciting a structure can be properly compared on a power basis.[6–8] The active power (the real part of the complex power) only is included here since it is argued that most excitations of interest will involve dissipation of power. The reactive power (imaginary part) is by definition stored during one half of an oscillation, and fully returned to the exciting source in the other half of the cycle. In many situations such temporary exchanges are not of interest since they play no part in far field radiation from the structure, or transmission to the far field, and are not considered in detail here. Thus, for simplicity, unless stated otherwise ''power'' in this paper refers to the active power only.

II. DIMENSIONLESS MOBILITY MATRIX

Consider an arbitrary structure excited by a number of applied forces, Fig. 1. The active power is given by:[9]

$$P = \overline{f}^H \,\mathrm{Re}(\overline{Y})\,\overline{f}, \tag{1}$$

where f is the vector of applied generalized forces (which may include moments) and Y is the mobility matrix. H indicates Hermitian transpose, i.e., $(\,)^H = (\,)^{*T}$, and barred quantities are complex. Note that the active power depends only on the real part of the mobility matrix.[9] Equation (1) can be rewritten in a dimensionless form,[10] i.e.:

$$P = \overline{q}^H C_r \overline{q}. \tag{2}$$

This is achieved by introducing a diagonal matrix, D, made up of the real parts of the point mobilities, $\overline{y_{ii}}$, i.e.,

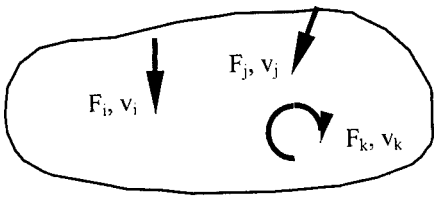


FIG. 1. Structure excited by arbitrary forces and moments.

$$D=\begin{bmatrix} \sqrt{\mathrm{Re}(\overline{y}_{11})} & & 0 \\ & \sqrt{\mathrm{Re}(\overline{y}_{22})} & \\ & \cdot & \\ 0 & & \sqrt{\mathrm{Re}(\overline{y}_{nn})} \end{bmatrix}. \tag{3}$$

The square symmetric matrix C_r in Eq. (2) is a dimensionless mobility matrix defined as,

$$C_r=(D^{-1}\,\mathrm{Re}(\overline{Y})D^{-1})=\begin{bmatrix} 1 & c_{12} & \cdot\cdot & c_{1n} \\ c_{21} & 1 & & \vdots \\ \vdots & & \cdot & \vdots \\ c_{n1} & \cdot\cdot & \cdot\cdot & 1 \end{bmatrix}, \tag{4}$$

in which

$$c_{ij}=\frac{\mathrm{Re}(y_{ij})}{\sqrt{\mathrm{Re}(y_{ii})\mathrm{Re}(y_{jj})}}. \tag{5}$$

Thus, the off-diagonal terms are rendered dimensionless by dividing by the geometric mean of the two corresponding point mobilities; a cross mobility is therefore normalized by a point force mobility and a moment mobility. The diagonal terms are all ones. This dimensionless matrix, and its use in Eq. (2), are the central concepts in this paper and will now be interpreted.

In Eq. (2), $\overline{q}=D\overline{f}$ is a forcing vector with units of square root power. The magnitude squared of each term is $|\overline{q_i}|^2=|f_i|^2\,\mathrm{Re}(\overline{y_{ii}})$, which is the power that would be input by a single force f_i. The squared length of this vector,

$$\overline{q}^H\overline{q}=P_u \tag{6}$$

is thus the power that would be input by all the applied forces if they acted independently, and can be termed the ''point-by-point power.''

Note that the matrix C_r describes the coupling between the various excitation components, and is independent of the actual level of excitation, and the distribution of the forces. On the other hand the forcing vector q describes the excitation and is independent of any coupling within the structure.

A response vector can also be defined through the dimensionless mobility matrix:

$$\overline{u}=C_r\overline{q}, \tag{7}$$

which has units square root of power. This is analogous to the velocity response vector in the classical formulation, or more strictly, to that part of the velocity vector $\overline{v'}=\mathrm{Re}(Y)\overline{f}$ whose dot product with the forcing vector yields the power. The actual power delivered is the inner product of the forcing and response vectors, i.e.,

$$P=\overline{q}^H\overline{u}. \tag{8}$$

Thus, the dimensionless mobility matrix can be seen as a transformation from the forcing vector to the response vector which are linked, respectively, to the point-by-point and actual delivered power.

A. Fully coupled case

One of the properties of C_r is positive definiteness,[9,11] since the power is always positive into the passive receiving structure. It follows that while off-diagonals can be positive or negative, their magnitude must be less than unity:[12]

$$|c_{ij}|\leqslant 1. \tag{9}$$

This fact has a several implications: for example, the inequality (9) can serve as a check on the validity of measured mobility data, since off-diagonals of greater than unit magnitude are unphysical. This check has been used successfully by the author to identify errors in measured mobilities, although no examples are given here since this is not the main focus of the paper. A further implication is that the limiting case $|c_{ij}|=1$ indicates the maximum possible coupling between points (and/or directions) i and j. We are thus able to quantify what is meant by fully coupled. (In Ref. 13 a rather different analysis led to a similar result.) In such a case, the matrix is rank deficient and has at least one zero eigenvalue. Rank deficiency is of interest since it implies that the structure can be fully described with less than a full matrix. This indicates a possibility for reducing the amount of measurement. The importance of a zero eigenvalue is that it implies that there exists at least one combination of generalized forces applied at the given points which will result in zero power transmission. This has implications for active vibration control since it implies that correct choice of control forces can completely eliminate transmission into the given structure. An illustrative example will be given later.

From the above it is of interest to investigate under what conditions the fully coupled case is achieved, and some examples will be given later.

B. Uncoupled case

We now turn our attention to the other limiting case, i.e., when the off-diagonal terms are small. $|c_{ij}|\ll 1$ implies that the coupling between forces i and j is negligible. This can be illustrated by the expansion of Eq. (2) for a system with two applied forces:

$$P=|q_1|^2+|q_2|^2+2|q_1||q_2|\cos(\theta)c_{12}, \tag{10}$$

where θ is the phase angle between the forces. The power input by point-by-point forces would be $|q_1|^2+|q_2|^2$. The actual delivered power input can be more or less, depending on the magnitude of the third term on the right hand side, i.e., depending on c_{12} and the phase angle. It is clear from Eq. (10) that for $|c_{ij}|\ll 1$, the actual power absorbed will never differ significantly from the point-by-point power and the applied forces can be considered independent.

The largest difference between the point-by-point and actual delivered power in Eq. (10) occurs when $|q_1|^2=|q_2|^2$ and is $P/P_u=(1+c_{12})$, where $P_u=|q_1|^2+|q_2|^2$ is

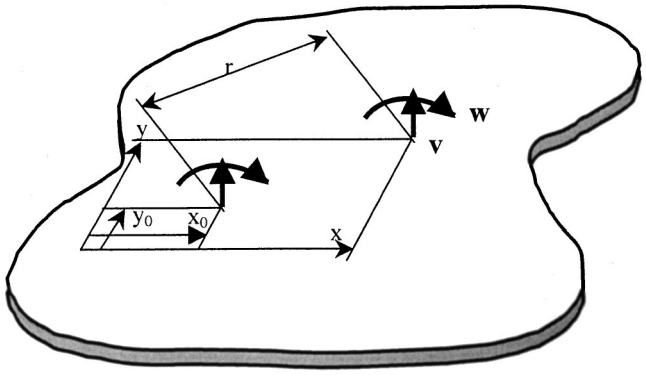


FIG. 2. Sign convention for the infinite plate with force and moment excitation.

the point-by-point power. Thus, the condition $|c_{ij}| \ll 0.3$ will ensure a maximum error of less than ± 1.5 dB, due to completely neglecting coupling between these two particular points (or directions). For multiple excitation points, neglecting all coupling terms could result in an accumulation of errors, but the limit $|c_{ij}| \ll 0.3$ remains a valid criterion for neglecting coupling between any two points (or directions) and will be adopted in the following sections.

C. Reactive power

A similar analysis to above can be carried out on the reactive power. In this case the power in Eq. (1) is the imaginary part of the power, and the imaginary part of the mobility matrix is used on the right hand side. All the steps as above can be followed through to yield a dimensionless mobility matrix. However, an important difference is that the resulting matrix is not positive definite, therefore the eigenvalues may be positive or negative and off-diagonal terms can be greater than unity. The reactive power will not be pursued further in this paper.

III. EXAMPLES OF DIMENSIONLESS MOBILITIES

In this section the off-diagonal elements of the dimensionless mobility matrix for idealized structures are examined. The objectives are to illustrate the concept of the dimensionless mobility, and through it to gain physical insight into the behavior of the structures.

A. Infinite plate

Initially we examine an infinite plate with force and moment excitation at a point as shown in Fig. 2. The required mobility relations (which are derived in Appendix A) are given below. These relations are based on thin plate theory, and are therefore subject to the restrictions pointed out in Ref. 14.

Force transfer mobility

$$Y_{vF} = \left(\frac{v}{F}\right)_{M=0} = Y_c \left\{ H_0^{(2)}(kr) - \frac{2i}{\pi} K_0(kr) \right\}; \tag{11}$$

Cross mobility

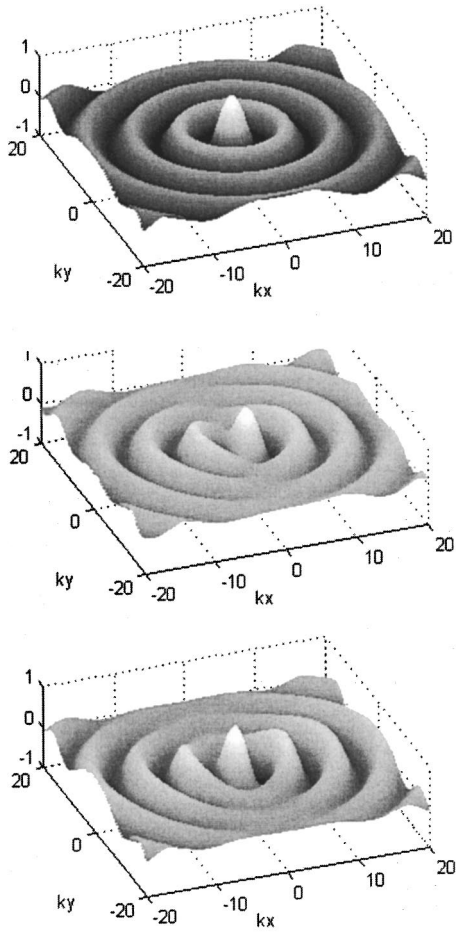


FIG. 3. Dimensionless mobility as a function of the distance between source and receiver point: (a) transfer mobility C_{vF}, (b) cross mobility $C_{wF} = C_{vM}$, (c) moment transfer mobility C_{wM}.

$$Y_{wF} = \left(\frac{w}{F}\right)_{M=0} = Y_c k \frac{(x-x_0)}{r} \left[H_1^{(2)}(kr) - \frac{2i}{\pi} K_1(kr) \right]; \tag{12}$$

Moment transfer mobility

$$Y_{wM} = \left(\frac{w}{M}\right)_{F=0} = Y_c k^2 \left\{ \left(\frac{x-x_0}{r}\right)^2 \left[\frac{1}{2} H_0^{(2)}(kr) - \frac{1}{2} H_2^{(2)}(kr) + \frac{i}{\pi} K_0(kr) + \frac{i}{\pi} K_2(kr) \right] + \left(\frac{y-y_0}{r}\right)^2 \left[H_1^{(2)}(kr) - \frac{2i}{\pi} K_1(kr) \right] \right\}. \tag{13}$$

Note that by reciprocity, the cross mobility can also be expressed as

$$\left(\frac{w}{F}\right)_{M=0} = \left(\frac{v}{M}\right)_{F=0}. \tag{14}$$

The dimensionless mobilities are derived by substituting Eqs. (11)–(14) into Eq. (3). In Fig. 3(a) is plotted the dimensionless transfer mobility, that is the real part of the term in brackets in Eq. (11). This function is well known, but is-

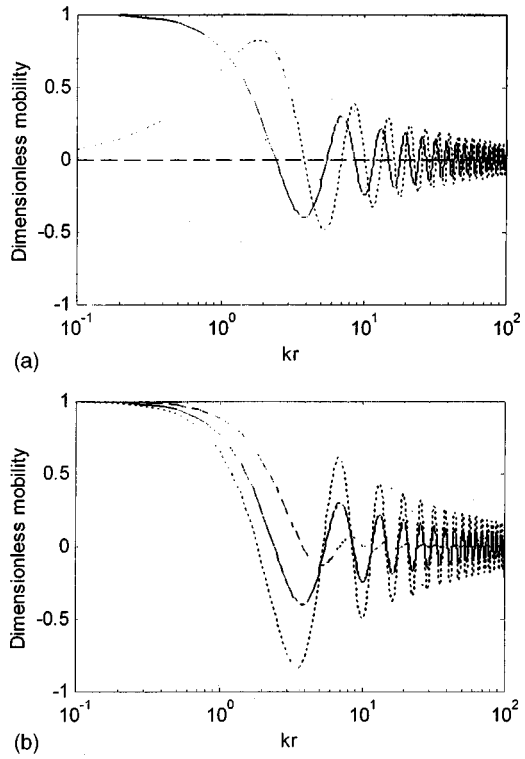


FIG. 4. Dimensionless mobility as a function of the distance between source and receiver point: (a) —— transfer mobility C_{vF}, · · · · cross mobility perpendicular to moment axis $C_{wF}=C_{vM}$, — — — cross mobility on moment axis. (b) —— transfer mobility C_{vF}, · · · · moment transfer mobility perpendicular to moment axis, — — — moment transfer mobility on axis.

given for completeness. Given in Fig. 3(b) is the dimensionless cross mobility, that is the real part of Eq. (12) normalized to the force and moment mobilities according to definition (5). The moment is about the *y* axis. It is seen that this is an antisymmetric function, and therefore zero on the *x* axis. It bears some resemblance to an acoustic dipole field, and has a strong directivity.

Shown in Fig. 3(c) is the dimensionless moment transfer mobility. This is a symmetric function about the *x* axis. Apart from the phase it has a similar directivity to the cross mobility, and the influence of the moment extends further along the *x* axis than the *y* axis.

Cross sections through Figs. 3(b) and (c) are also given in Figs. 4(a) and (b), giving cross mobility and moment transfer mobility, respectively. The dimensionless force transfer mobility is shown in both plots for comparison. It is seen that the transfer mobility falls below 0.3 for distances greater than about $kr=6$, in other words, the points can be considered uncoupled in force when separated by more than about a wavelength. We can therefore talk of a ''coupling distance'' for an infinite plate of about one wavelength. Figure 4(b) shows that in the case of moment excitation the influence extends to a somewhat greater distance along the *x* axis, but dies away more quickly along the *y* axis. For all cases, full coupling occurs only when the two points are closer than a sixth of a wavelength, $kr<1$. All the findings of this subsection are consistent with those in Ref. 15 where the same problem was studied from the point of view of active control of a primary force by a secondary moment.

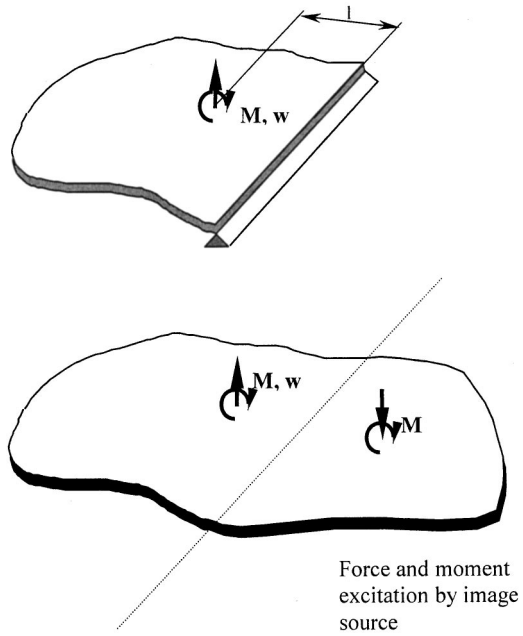


FIG. 5. Sign convention for an excitation point close to a simple supported edge, and image source model.

B. Semi-infinite plate

The above results illustrate among other things the well-known fact that cross coupling at the point of excitation is zero for an infinite plate due to symmetry. However, Petersson[6,7] has shown that cross coupling can become significant close to either a clamped or simply supported edge as shown in Fig. 5, which also illustrates the image source model used to derive the mobilities. The smooth, bold line in Fig. 6 illustrates the dimensionless cross coupling close to the simply supported edge of a semi-infinite plate (the derivation of the mobility equations is given in Appendix B). There is full coupling close to the edge, which confirms Petersson's findings. This indicates that there is the theoretical possibility of complete cancellation of a force and moment resulting in zero power transmission. Further than 1/2 wavelength away the coupling could be neglected to engineering accuracy, and full cancellation is impossible with only a force and moment applied.

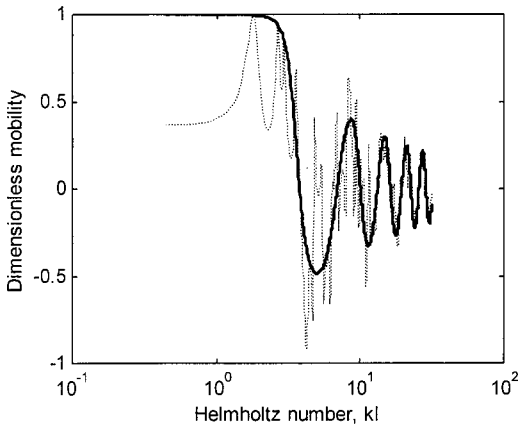


FIG. 6. Dimensionless cross mobility close to a simply supported edge: —— semi-infinite plate, · · · · finite (SSSS) plate.

The potential for cancellation of a force excitation by a suitable moment was first noted in Ref. 16 and studied by Petersson[6,7] and Koh.[17–19] The dimensionless cross mobility provides an alternative means of establishing when such cancellation is possible, and also generalizes the idea to include any type of forced excitation.

C. Finite plate

The simply supported plate shown in Fig. 7 has also been analyzed. The mobility formulas, which are derived in Appendix C are based on the well-known series solution. The plate is the same thickness and material as the semi-infinite plate given above, and the excitation point is the same distance from the edge. In Fig. 6 is shown the dimensionless cross mobility. It is seen that coupling is strong at every natural frequency. For comparison the semi-infinite plate result is also given. The finite plate curve is seen to follow the trends of the semi infinite case over most of the frequency range, with resonances superimposed. These results are reminiscent of those given by Skudrzyk's mean value theorem.[20] The low frequency asymptote is however different in the finite and semi-infinite case, and full coupling is not achieved for the finite plate. The strong coupling at resonance can be explained by considering the mobility matrix for a resonant structure with mode shapes φ_n:

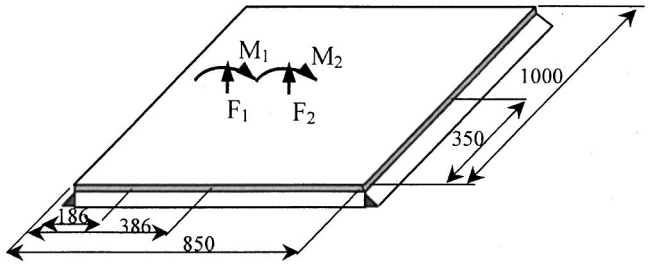


FIG. 7. SSSS plate with force and moment applied at two points used to obtain results in Figs. 8–10.

$$Y=\frac{i\omega}{M}\sum_{n=1}^{\infty}\frac{1}{\omega_n^2-\omega^2}\begin{bmatrix}\varphi_n(x_1,y_1)\varphi_n(x_1,y_1) & \varphi_n(x_1,y_1)\varphi_n(x_2,y_2) & \cdots & \varphi_n(x_1,y_1)\varphi_n(x_N,y_N)\\ \varphi_n(x_2,y_2)\varphi_n(x_1,y_1) & \varphi_n(x_2,y_2)\varphi_n(x_2,y_2) & & \cdot\\ \cdot & & \cdot & \cdot\\ \varphi_n(x_N,y_N)\varphi_n(x_1,y_1) & \cdots & & \varphi_n(x_N,y_N)\varphi_n(x_N,y_N)\end{bmatrix}. \tag{15}$$

The above was derived from the modal expansion of the response of a structure at point (x_p,y_p) subject to point excitation at (x_q,y_q),[16] that is:

$$\frac{v_p}{F_q}=\frac{i\omega}{M}\sum_{n=1}^{\infty}\frac{\varphi_n(x_p,y_p)\varphi_n(x_q,y_q)}{\omega_n^2-\omega^2}. \tag{16}$$

Note that if rotations are included then derivatives of the mode shapes can appear in this matrix, for example from

$$\frac{w_p}{F_q}=\frac{i\omega}{M}\sum_{n=1}^{\infty}\frac{\varphi_n'(x_p,y_p)\varphi_n(x_q,y_q)}{\omega_n^2-\omega^2},\qquad \frac{w_p}{M_q}=\frac{i\omega}{M}\sum_{n=1}^{\infty}\frac{\varphi_n'(x_p,y_p)\varphi_n'(x_q,y_q)}{\omega_n^2-\omega^2}, \tag{17}$$

where the prime indicates differential. Thus, the form of Eq. (15) covers all combinations of normal forces and rotations.

At frequencies where a single resonance dominates the response, one can take a single term in the series of Eq. (15). This can then be rendered dimensionless using the matrix

$$D=\left(\mathrm{Re}\left(\frac{i\omega}{M}\frac{1}{\omega_n^2-\omega^2}\right)\right)^{1/2}\times\begin{bmatrix}\varphi_n(x_1,y_1) & & & 0\\ & \varphi_n(x_2,y_2) & & \\ & & . & \\ 0 & & & \varphi_n(x_N,y_N)\end{bmatrix} \tag{18}$$

as described in Eq. (4), giving

$$C=\begin{bmatrix}1 & 1 & & \\ 1 & 1 & & \\ & & . & \\ & & & 1\end{bmatrix}. \tag{19}$$

The dimensionless mobility matrix in this special case thus has all terms equal. This confirms the observation from Fig. 6 that all excitations are strongly coupled whenever a single mode dominates its response. The rank of the matrix is unity and it has one nonzero eigenvalue, which indicates that complete cancellation of the power input is theoretically possible for a single isolated mode. A physical explanation is that the structure is constrained to move according to its mode shape and therefore has only one degree of freedom; all points are therefore fully coupled, and all motion can be suppressed by a control force which can be applied at any point on the structure. In practice, this is a limiting case which can never be fully achieved since there will always be contributions from other modes.[11] Nevertheless, the analysis gives some insight into the results of Fig. 6.

We conclude that cross coupling should be taken into account, not just near a discontinuity, but also at resonance frequencies. A finite plate was also studied in Ref. 15 from the point of view of active control. The findings were consistent with those presented here.

Comparing the results from the last three subsections it is seen that for a transfer mobility, the distance to a neighboring excitation point is the factor which determines the

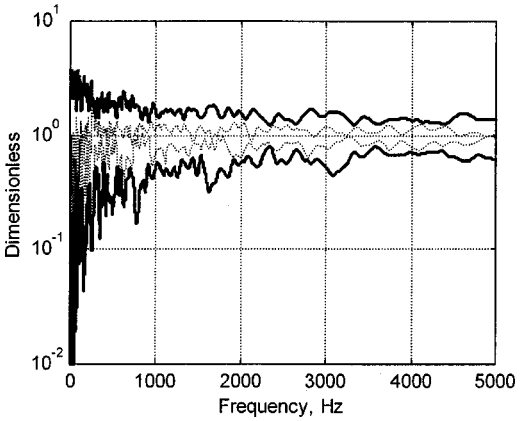


FIG. 8. Eigenvalues for the dimensionless mobility matrix of the SSSS plate shown in Fig. 7.

degree of coupling, whereas for a cross mobility it is the distance to a constrained edge.

As a final comment on finite plates, it is interesting to consider whether one can define a ''coupling distance'' in the same way as was done for infinite plates. It was found by simulation that when excited away from a resonance, the coupling distance of about one wavelength obtained for the infinite plate was valid also for the finite plate. However, close to a resonance, all points on the structure are coupled as shown above, and a corresponding coupling distance cannot be found.

IV. RELATIONSHIP BETWEEN POINT-BY-POINT AND DELIVERED POWER

Figure 8 illustrates the eigenvalues of the dimensionless mobility matrix for a simply supported rectangular plate excited at two points by a force and moment, as shown in Fig. 7. The eigenvalues (which are a function of frequency) can be interpreted as the ratio of delivered to point-by-point power. This can be seen from the well-known relationship[12]

$$\lambda_{\min} \leqslant \frac{P}{\Sigma_{i=1}^{n} |\overline{q}|^2} \leqslant \lambda_{\max}, \tag{20}$$

where $\lambda_{\min}$, $\lambda_{\max}$ are the maximum and minimum eigenvalues of C_r, respectively, and $\Sigma_{i=1}^{n}|\overline{q}|^2$ is the point-by-point power. Note that $\lambda_{\min}<1$ and $\lambda_{\max}>1$, thus, the delivered power can be more or less than the point-by-point power by a factor that can fall anywhere between the bold lines in Fig. 8. Where the delivered power falls between these extremes depends on the relative magnitude and phase of the applied forces and moments. For this example, at frequencies above 680 Hz (which corresponds to a single wavelength separation of the points) the bounds are relatively narrow, and the delivered power will not differ greatly from the point-by-point power. This is significant because the point-by-point power can be estimated without knowledge of the phase of the applied forces or any of the off-diagonal terms of the matrix. At lower frequencies there is the possibility of significant errors if the coupled power is estimated from the point-by-point power, and the coupling must be taken into account. The widest variations are at resonance where the full coupling condition is approached as was anticipated from the previous section. On the other hand, the results in Fig. 8 can be viewed from the perspective of active vibration control where one or more of the forces or moments is a control force. It has been shown by Bardou *et al.*,[15] and Pavic[21] that optimum control is achieved by minimizing the power input to a structure. The minimum eigenvalue indicates the minimum possible power, and hence the maximum control that can be obtained with forces and moments applied at the given points. It is seen that in this example the minimum eigenvalue is around 0.5–0.7 above 1000 Hz, indicating a maximum reduction of 1.5–3 dB from the point-by-point power (which corresponds to a value of unity in Fig. 8). At lower frequencies, particularly at resonances there is more scope for control, for example, below 200 Hz a reduction by a factor of 10 is theoretically possible. Such an analysis provides a measure of the inherent suitability for active control of a structure excited at given points.

The combination of forces that minimizes power input is given by the eigenvector corresponding to the minimum eigenvalue. This eigenvector is a function of frequency and the magnitude of its elements are plotted in Fig. 9. The four terms are the elements of the forcing vector corresponding to force and moment at position 1 and position 2. They are defined by $\overline{q}=D\overline{f}$ where each of the forces or moments is multiplied by the square root of its corresponding point mobility:

$$\begin{aligned} &\overline{q}_{f1}=\overline{f}_1\sqrt{\mathrm{Re}(y_{v1,f1})}, \quad \overline{q}_{f2}=\overline{f}_2\sqrt{\mathrm{Re}(y_{v2,f2})}, \\ &\overline{q}_{m1}=\overline{m}_1\sqrt{\mathrm{Re}(y_{w1,m1})}, \quad \overline{q}_{m2}=\overline{m}_2\sqrt{\mathrm{Re}(y_{w2,m2})}. \end{aligned} \tag{21}$$

These terms are squared in Fig. 9 so as to give the more familiar units of power. The plot is not a full representation of the eigenvector, since only the magnitude is plotted, the sign having been dropped in order to clarify the trends.

Figure 9(a) shows that below 16 Hz there exists a quasi static regime in which the minimum power input is achieved with a combination of two forces of comparable magnitude (these are actually of opposite sign, although this is not shown in the figure). In this frequency range the moments [indicated by the dotted lines in Fig. 9(a)] play virtually no part in the cancellation.

However, at higher frequencies, the force terms do not in general dominate. In fact, the excitation of forces and moments (judged in terms of their point-by-point power) are equally important over much of the range. Indeed, over certain frequency ranges the role of moments in cancelling the power is dominant, for example over the ranges 1.5–1.8 kHz, 2.8–3.3 kHz, and 4.5–5 kHz. This indicates that over these frequency ranges the most effective cancellation would be achieved by a correct combination of moments, with virtually no applied normal forces. An example is given in Fig. 10(a) where the ''minimum'' eigenvector at 3421 Hz is plotted. The lengths of the lines are proportional to the strength of the force and moment. Although, the main purpose of this paper is not to examine active control, this analysis gives a clear indication that this is an inherently unsuitable structure over these ranges in view of the difficulty of applying control moments.

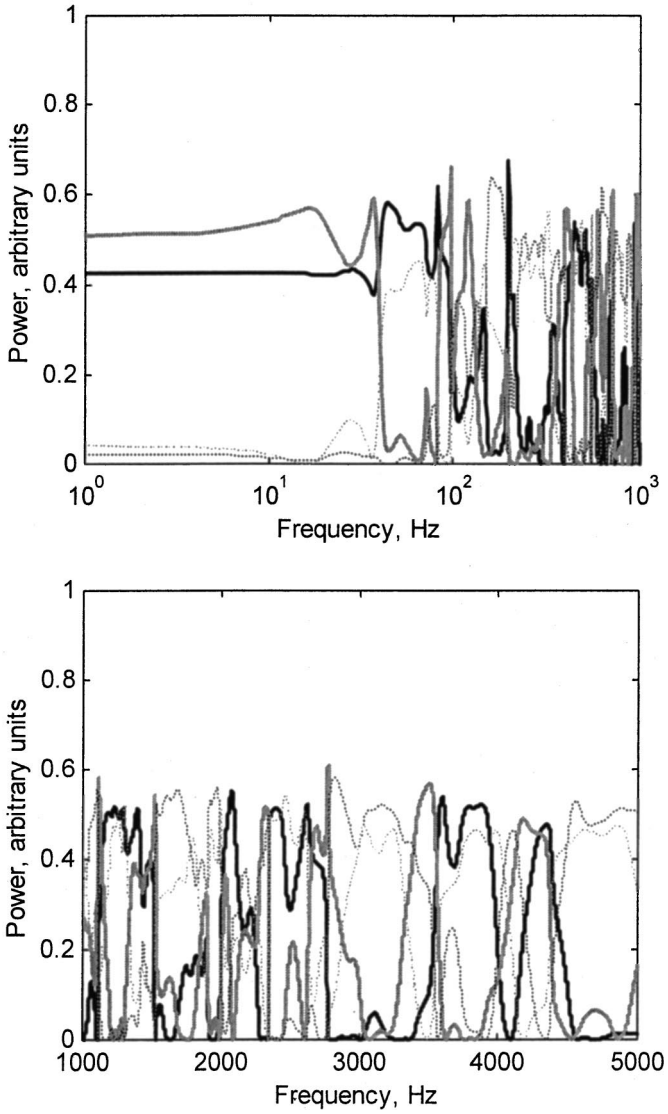


FIG. 9. Eigenvector corresponding to the minimum eigenvalue from Fig. 8: —— force excitation, q_{f1}, q_{f2}; · · · · moment excitation, q_{m1}, q_{m2}.

Shown in Fig. 10(b) is the eigenvector corresponding to minimum power at the lowest resonance frequency, 16 Hz. This corresponds to the deepest trough in Fig. 8. The minimum is achieved in this case with two nearly equal and opposite forces and very small moments. This is not a general result, however, as at other resonance frequencies moments were found to be equally important.

V. CONCLUDING REMARKS

A dimensionless mobility matrix has been presented based on the real part of the mobility matrix, which controls the power input to the structure. The diagonals of the dimensionless mobility matrix are all ones, and the off-diagonals, which quantify the coupling between two points or directions, are normalized by the geometric mean of the corresponding point mobilities. Thus, a cross mobility is normalized by the geometric mean of a force and a moment mobility, for example.

It is shown that the magnitude of a real part, off-diagonal term cannot exceed unity, and that when it equals

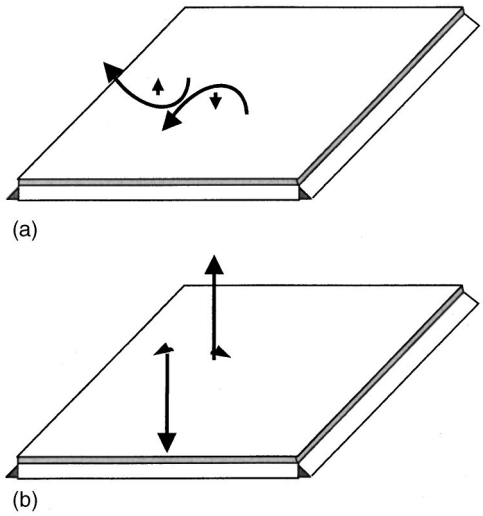


FIG. 10. Eigenvectors at (a) 3421 Hz and (b) 16 Hz (lowest resonance). The lengths are proportional to the force or moment.

unity the corresponding two degrees of freedom can be considered ''fully coupled.'' In such cases the coupling must be taken into account in calculating power input. An implication of ''full coupling'' is that the matrix is rank deficient and therefore has at least one zero eigenvalue. The eigenvalues of the dimensionless mobility matrix relate the delivered and point-by-point power input, and a zero eigenvalue means that there exists a combination of generalized forces (nonzero) which completely cancel and result in no net power input. This has potential application in active vibration control: points on a structure or directions of excitation which are fully coupled could be considered inherently suitable for active control. Simulation on idealized structures has demonstrated that such cases are limited to close points, a combined force and moment close to a fixed edge, and isolated resonances. In other cases, the minimum eigenvalue provides a measure of the minimum power inherent to the given points, and the eigenvector gives the combination of forces to achieve this minimum.

At the other extreme, small off-diagonals in the dimensionless mobility matrix indicate that coupling between the relevant points or directions can be ignored. Simulations have shown that cross coupling can be ignored away from discontinuities and away from resonances. Similarly, transfer coupling can be ignored away from other excitation points and away from resonance frequencies.

The above results provide useful insights relating to the power absorbing properties of structures. It should be appreciated that ''connectivity'' properties are also needed to determine the contact forces when two or more structures are connected, where both the real and imaginary parts of the mobility are equally important. It would be interesting to extend the above approach, although the brief analysis in Sec. II C suggests that this may not be completely straightforward because the positive definite property does not apply to the imaginary part of the mobility matrix.

ACKNOWLEDGMENT

This work was financially supported by the Engineering and Physical Sciences Research Council, whose support is gratefully acknowledged.

APPENDIX A: FORCE AND MOMENT MOBILITIES FOR AN INFINITE PLATE

The sign convention is given in Fig. 2. The transfer mobility of an infinite plate is given by Cremer and Heckl[16] as

$$Y_{vF}=\frac{v}{F}=Y_c\cdot\Pi(kr), \tag{A1}$$

in which v is the velocity at the response point a distance r from the excitation point, and F is the exciting force. Y_c is the characteristic mobility of the plate, and k the flexural wave number. The function:

$$\Pi(kr)=H_0^{(2)}(kr)-H_0^{(2)}(-jkr)=H_0^{(2)}(kr)-\frac{2i}{\pi}K_0(kr) \tag{A2}$$

is the difference between two Hankel functions of zero order and the second kind. The first form on the right hand side of Eq. (A2) is given in Ref. 16, but the second form, in which $K_0(kr)$ is the modified Bessel function of the second kind, is more convenient for analysis and will be used in the following. Similar analyses are given in Refs. 15 and 22.

The angular velocity is evaluated by differentiation of Eq. (A2) with respect to the x coordinate of the receiver point:

$$\begin{aligned}Y_{wF}=\frac{w}{F}&=-\frac{1}{F}\frac{\partial v}{\partial x}\\&=-Y_c\cdot\frac{\partial}{\partial x}\{\Pi(kr)\}\\&=-Y_ck\cdot\Pi'(kr)\frac{\partial r}{\partial x}\\&=Y_ck\frac{(x-x_0)}{r}\left[H_1^{(2)}(kr)-\frac{2i}{\pi}K_1(kr)\right],\end{aligned} \tag{A3}$$

where the prime indicates differentiation with respect to kr. The moment mobility is evaluated by further differentiating the Green's function (A1) with respect to the coordinate of the excitation point, x_0:

$$\begin{aligned}Y_{wM}=\frac{w}{M}&=Y_c\cdot\frac{\partial}{\partial x}\frac{\partial}{\partial x_0}\{\Pi(kr)\}\\&=Y_ck^2\left\{\Pi''(kr)\frac{\partial r}{\partial x_0}\frac{\partial r}{\partial x}+\Pi'(kr)\frac{\partial^2 r}{\partial x\,\partial x_0}\right\}.\end{aligned} \tag{A4}$$

The following relationships are now required:

$$\begin{aligned}&r=[(x-x_0)^2+(y-y_0)^2]^{1/2},\\&\frac{\partial r}{\partial x}=\frac{x-x_0}{r},\quad \frac{\partial r}{\partial x_0}=-\frac{x-x_0}{r},\\&\frac{\partial^2 r}{\partial x\,\partial x_0}=-\frac{(y-y_0)^2}{r^3},\end{aligned} \tag{A5}$$

$$\Pi'(kr)=-H_1^{(2)}(kr)+\frac{2i}{\pi}K_1(kr), \tag{A6}$$

$$\begin{aligned}\Pi''(kr)=&\tfrac{1}{2}[-H_0^{(2)}(kr)+H_2^{(2)}(kr)]\\&-\frac{i}{\pi}[K_0(kr)+K_2(kr)].\end{aligned} \tag{A7}$$

The latter recurrence relations can be found in Ref. 23. Expressions for the moment and cross mobility are obtained by substituting relations (A5)–(A7) into Eqs. (A3) and (A4) to give Eqs. (12) and (13).

The imaginary part of the last relation for moment mobility increases without bound as $kr\to 0$. However, the real part, which is used for evaluation of the dimensionless mobility, remains finite at the excitation point.

APPENDIX B: FORCE AND MOMENT MOBILITIES FOR A SEMI-INFINITE PLATE

The relationships between forces, moments and responses in terms of velocity and angular velocity in the vicinity of a simply supported edge are derived using the method of image sources as shown in Fig. 5. A similar analysis has been described in Ref. 7. The origin is taken as the point of application of the real forces. The point, cross and moment mobility are the superposition of the contributions from the real and image sources.

$$\begin{aligned}&Y_{vF}=(v/F)_{M=0}=Y_{vF}^{\infty}(0,0)-Y_{vF}^{\infty}(0,2l),\\&Y_{wF}=(w/F)_{M=0}=Y_{wF}^{\infty}(0,0)-Y_{wF}^{\infty}(0,2l),\\&Y_{wM}=(w/M)_{F=0}=Y_{wM}^{\infty}(0,0)+Y_{wM}^{\infty}(0,2l),\end{aligned}$$

where $Y_{vF}^{\infty}(0,2l)$, etc. are the mobilities of an infinite plate from the previous section with excitation at $2l$ and response at 0 (the superscript has been added to indicate an infinite plate).

APPENDIX C: FORCE AND MOMENT MOBILITIES FOR A SIMPLY SUPPORTED PLATE

These are derived using the normal modes expansion for the Green's function of a simply supported plate (see, for example, Ref. 14):

$$Y_{vF}=\frac{4i\omega}{M}\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\frac{\sin(k_mx_0)\sin(k_ny_0)\sin(k_mx)\sin(k_ny)}{\omega_{mn}^2-\omega^2}, \tag{C1}$$

where x_0, y_0 are the coordinates of the excitation point, and x, y those of the response point, M is the mass of the plate, and ω is the radian frequency, $k_m=(m\pi/l_x)$, and $k_n=(n\pi/l_y)$ where l_x, l_y are the dimensions of the plate. The natural frequencies ω_{mn} are given by:

$$\omega_{mn} = c_L \kappa (k_n^2 + k_m^2)(1 + i\eta/2),$$

where η is the loss factor, c_L the longitudinal wave speed, and κ the radius of inertia of the plate section ($\kappa = h/\sqrt{12}$ for a solid plate of thickness h).

As was done for the infinite plate, the angular velocity is derived by differentiation of Eq. (C1) with respect to the response point coordinate, x (the minus sign arrives because of the sign convention):

$$Y_{wF} = -\frac{4i\omega}{M} k_m \times \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{\sin(k_m x_0)\sin(k_n y_0)\cos(k_m x)\sin(k_n y)}{\omega_{mn}^2 - \omega^2}.$$

Again, as for the infinite plate, the response due to a moment is derived by differentiation with respect to the forcing point coordinate x_0:

$$Y_{wM} = \frac{4i\omega}{M} k_m^2 \times \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{\cos(k_m x_0)\sin(k_n y_0)\cos(k_m x)\sin(k_n y)}{\omega_{mn}^2 - \omega^2}.$$

A similar set of equations is given in Refs. 15 and 22.

[1] J. M. Mondot and A. T. Moorhouse, ''The characterization of structure-borne sound sources,'' Proc Inter-noise 96, Book 3, Liverpool (1996), pp. 1439–1446.

[2] M. A. Sanderson and C. R. Fredo, ''Direct measurement of moment mobility. 1. A theoretical study,'' J. Sound Vib. **179**, 669–684 (1995).

[3] M. A. Sanderson and C. R. Fredo, ''Direct measurement of moment mobility. 1. An experimental study,'' J. Sound Vib. **179**, 685–696 (1995).

[4] S. H. Yap and B. M. Gibbs, ''Structure-borne sound transmission from machines in buildings. Part 2: Indirect measurement of force and moment at the machine-receiver interface of a single point connected system by a reciprocal method,'' J. Sound Vib. **222**, 99–113 (1999).

[5] B. A. T. Petersson and B. M. Gibbs, ''Use of the source descriptor concept in studies of multi-point and multi-directional vibrational sources,'' J. Sound Vib. **168**, 157–176 (1993).

[6] B. A. T. Petersson, ''Structural acoustic power transmission by point moment and force excitation, Part I: Beam-like and frame-like structures,'' J. Sound Vib. **160**, 43–66 (1993).

[7] B. A. T. Petersson, ''Structural acoustic power transmission by point moment and force excitation, Part II: Plate-like structures,'' J. Sound Vib. **160**, 67–91 (1993).

[8] C. Kauffmann, ''Input mobilities and power flows for edge-excited, semi-infinite plates,'' J. Acoust. Soc. Am. **103**, 1874–1884 (1998).

[9] Su. Jianxin, A. T. Moorhouse, and B. M. Gibbs, ''Towards a practical characterization for structure-borne sound sources based on mobility techniques,'' J. Sound Vib. **185**, 737–741 (1995).

[10] A. T. Moorhouse, ''A mobility matrix formulation for force and moment excitation of structures,'' Proc. Inter-noise 99, Fort Lauderdale (1999).

[11] A. T. Moorhouse and B. M. Gibbs, ''Simplified characterization of multiple point excited structures using mobility matrix eigenvalues and eigenvectors,'' Acustica **84**, 843–853 (1998).

[12] J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Oxford University Press, London, 1965).

[13] F. Bessac, L. Gagliardini, and J.-L. Guyader, ''Coupling eigenvalues and eigenvectors: A tool for investigating the vibroacoustic behavior of coupled vibrating systems,'' J. Sound Vib. **191**, 881–899 (1996).

[14] I. Dyer, ''Moment impedance of plates,'' J. Acoust. Soc. Am. **32**, 1290–1297 (1960).

[15] O. Bardou, P. Gardonio, S. J. Elliott, and R. J. Pinnington, ''Active power minimization and power absorption in a plate with force and moment excitation,'' J. Sound Vib. **208**, 111–151 (1997).

[16] L. Cremer, M. Heckl, and E. E. Ungar, *Structureborne Sound* (Springer Verlag, Berlin, 1973).

[17] Y. K. Koh and R. G. White, ''Analysis and control of vibrational power transmission to machinery supporting structures subjected to a multi-excitation system, Part I: Driving point mobility matrix of beams and plates,'' J. Sound Vib. **196**, 509–522 (1996).

[18] Y. K. Koh and R. G. White, ''Analysis and control of vibrational power transmission to machinery supporting structures subjected to a multi-excitation system, Part II: Vibrational power analysis and control schemes,'' J. Sound Vib. **196**, 509–522 (1996).

[19] Y. K. Koh and R. G. White, ''Analysis and control of vibrational power transmission to machinery supporting structures subjected to a multi-excitation system, Part III: Vibrational power cancellation and control experiments,'' J. Sound Vib. **196**, 509–522 (1996).

[20] E. Skudrzyk, ''The mean-value method of predicting the dynamic response of complex vibrators,'' J. Acoust. Soc. Am. **67**, 1105–1135 (1980).

[21] G. Pavic, Proc 7th Int. Congress on Sound and Vibration, Garmisch-Partenkirchen. Power and energy in vibroacoustic systems (2000).

[22] P. Gardonio, S. J. Elliott, and R. J. Pinnington, ''Active isolation of structural vibration on a multiple degree of freedom system, Part 1: The dynamics of the system,'' J. Sound Vib. **207**, 61–93 (1997).

[23] E. Skudrzyk, *The Foundations of Acoustics* (Springer Verlag, Wien, 1971), Table VIII.

Combined feedback–feedforward active noise-reducing headset—The effect of the acoustics on broadband performance

Boaz Rafaely[a] and Matthew Jones
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

(Received 10 September 2001; revised 21 June 2002; accepted 24 June 2002)

Active noise-reducing headsets that employ analog feedback control and provide good broadband attenuation are commercially available for a wide range of applications. Recent studies have explored the integration of an adaptive digital feedforward controller with the analog feedback controller to provide additional attenuation of periodic noise components. This paper presents an experimental study of such a combined control system, but with both feedback- and feedforward controllers attenuating broadband noise. Good performance is demonstrated in a reverberant sound field, while under direct sound-field conditions the attenuation performance of the feedforward controller is shown to be dependent on head position. The paper concludes with an analysis of the forward path delay showing how the passive attenuation mechanism improves broadband performance. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1501090]

PACS numbers: 43.50.Hg, 43.50.Ki, 43.66.Vt [MRS]

I. INTRODUCTION

Active noise-reducing headsets use active sound control to reduce the noise level at the user's ears. Passive noise-reducing headsets provide good attenuation at high frequencies, and so the objective of active control in headsets is to improve low-frequency attenuation. Active headsets which employ analog feedback control were studied in the 1950's by Simshauser, Hawley, and Meeker[1–3] and are currently available commercially for aviation, military, industrial, and consumer applications. Recently, active headsets based on digital controllers have been investigated and showed a potential for improved performance. Digital adaptive feedback controllers were used to cancel periodic siren noise,[4,5] or in combination with the analog feedback controller.[6] Digital adaptive feedforward controllers were also investigated in canceling periodic noise[7] and broadband noise.[8,9] Since the delay of the digital system limits broadband performance, multirate sampling methods were proposed to reduce delay.[7,8] In more recent studies, a combined system was investigated, where an analog feedback controller was used to reduce broadband noise, together with an adaptive feedforward controller which was used to eliminate periodic noise components.[10,11] Nevertheless, the ability of a combined system to reduce broadband noise, and the effect of the surrounding sound field and the headset acoustics on performance, were not investigated.

The aim and contribution of this paper is the study of the performance of a combined feedback–feedforward active headset in attenuating broadband noise, and the investigation of the dependence of this performance on the sound field surrounding the headset and the sound transmission through the headset. The results show that good performance of the feedforward part of the combined headset can be achieved in a reverberant sound field, and in a direct sound field when the headset is facing the source, due to the requirement for advanced information in the reference microphone. Nevertheless, this requirement is relaxed by the increase in delay imposed by the passive transmission through the headset. The results presented in this paper could be used to improve performance of practical active headsets under realistic conditions.

The paper is presented as follows. First, the combined active headset system used in this work is described, after which attenuation performance is presented under reverberant and direct sound-field conditions, with the subject head both stationary and moving. The paper is concluded by the analysis of the acoustic delay in the feedforward system and its effect on performance.

II. THE ACTIVE HEADSET SYSTEM

The active headset system used in this work employed a combined analog feedback controller and an adaptive digital feedforward controller, as illustrated in Fig. 1. The headset used was a modified Lectret active noise-reducing headset, which included a circumaural ear defender, fitted with a loudspeaker and a microphone inside the shell, connected via the analog feedback control circuit. The digital feedforward control system was implemented on a two-channel LSI TMS320C30-based DSP card, placed in a PC. An additional electret microphone was positioned centrally on the outer side of the earshell, and was connected to the DSP card via a preamplifier and an 8th-order antialiasing KEMO VBF8 low-pass filter. The output of the internal microphone was connected to the DSP via a similar low-pass filter, while the DSP digital-to-analog converter (DAC) output was connected to the loudspeaker input via a third low-pass filter and the analog control circuit.

Figure 2 presents a block diagram of the combined control system. The plant P controlled by the feedback controller C represents the response from the loudspeaker input to the internal microphone output, and includes the response of

[a]Electronic mail: br@isvr.soton.ac.uk

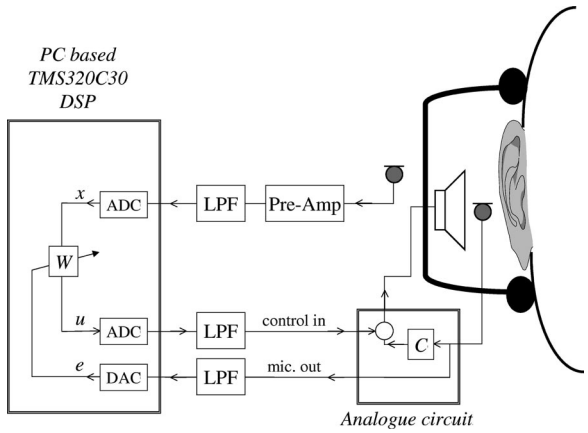


FIG. 1. Experimental setup including the headset and transducers, the analog circuit with the feedback controller (C), the DSP system with the feedforward controller (W), and signal amplifier and filters.

the amplifiers. The plant G controlled by the digital controller W also represents a response between the loudspeaker input and the internal microphone output, but includes the effect of the feedback controller, the low-pass filters, and the ADC and DAC, denoted by H_{ADC} and H_{DAC}, and can be written in the frequency domain as

$$G = H_{\mathrm{DAC}} \frac{P}{1+CP} H_{\mathrm{ADC}}. \tag{1}$$

The summation of the feedback- and feedforward control signals was implemented in the analog control circuit. The filtered-x LMS algorithm[12] was implemented to adapt the coefficients of filter W, while an LMS-based system identification was performed to estimate the model $\hat{G}$ prior to control. The DSP was operating at a sampling frequency of 12 kHz, with the cutoff frequency of the low-pass filters set to 5.5 kHz. This sampling frequency was chosen as a compromise between the need to reduce the electronic delay to enable broadband performance (with consequently high sampling rate) and the need to allow time for the real-time implementation (with consequently low sampling rate). Both the control filter W and model filter $\hat{G}$ were implemented as finite impulse response (FIR) filters with 40 coefficients. This filter length was sufficient to provide reasonable control performance and plant modeling.

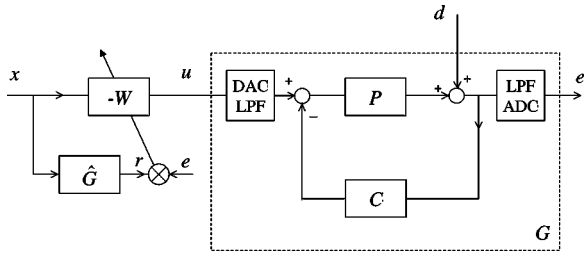


FIG. 2. Block diagram of the combined analog feedback–digital feedforward active noise-reducing headset system, with C and P the analog feedback controller and plant; W and G the digital feedforward controller and plant; and $\hat{G}$ the digital plane model.

III. PERFORMANCE IN REVERBERANT AND DIRECT SOUND FIELDS

A. Introduction

The broadband performance of the combined active noise-reducing headset in reverberant and direct primary sound fields was investigated in this section. Although the performance of the feedback control part is not expected to be dependent on the surrounding sound field, because the feedback controller closes a local control loop inside the earshell, the broadband performance of the feedforward control part is dependent on the advance information in the reference microphone relative to the error microphone, which in turn could be dependent on the sound field around the head. For example, the acoustic delay between the reference and error microphones is largest if the primary field propagates directly from the reference to the error microphone, and smaller if it propagates in other directions or if the sound field is reverberant. Now, if the acoustic delay between the microphones is smaller than the delay of the plant G introduced by the electronics and acoustics, broadband performance can be degraded significantly.[13] Performance was therefore studied by introducing two extreme acoustic conditions, namely, reverberant and direct sound fields, with various head positions relative to the source, as described below.

B. Experimental procedure

The experiment was conducted in two different rooms. The first was a reverberation chamber of dimensions 2.4 $\times$2.5$\times$2.6 m, with hard walls and no furniture except the experimental equipment, and was used to study performance under reverberant conditions. It should be noted that although this room produces a ''reverberant'' sound field, it does not strictly comply with the ISO requirements for reverberation rooms,[14] and therefore the sound field might still have some directivity. The second was a laboratory room of dimensions 5.7$\times$6.0$\times$3.2 m, which included benches, cupboards, and other laboratory equipment, and was used to study performance under direct field conditions by positioning the headset in the near field of a loudspeaker.

The active headset described above was fitted to one subject, with only the right side operating. Only one subject was used since the nature of the study was exploratory. Nevertheless, results were consistent throughout the experiment. An extension of this study to investigate performance across subjects is suggested for future work.

The noise source in this experiment was produced by passing a white-noise signal generated by a two-channel spectrum analyzer (Advantest R9211C) through 8th-order high-pass and low-pass KEMO VBF 8 filters in the range 200–900 Hz, which were then connected to a power amplifier driving a 10-cm-radius loudspeaker. Once the active headset was fitted to the subject and the noise source activated, the power spectral density at the error microphone (internal microphone) was measured by the spectrum analyzer and recorded on a disk for further analysis. The spectrum at the microphone was measured with various conditions of the active headset: no control, with both analog and

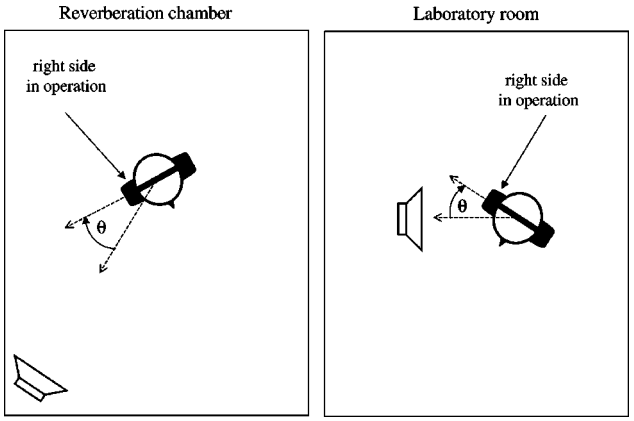


FIG. 3. Illustration of the 2.4×2.5×2.6-m reverberation chamber and the 5.7×6.0×3.2-m laboratory room with the subject wearing the headset placed in a position θ relative to the loudspeaker noise source.

digital controller turned off; analog feedback controller turned on (and digital feedforward controller off), and both controllers turned on.

In the experiment with the reverberant conditions, performed in the reverberation chamber, the source loudspeaker was placed in the room corner, about 2 m away from the subject wearing the headset, ensuring that the indirect sound field is dominant, as illustrated in Fig. 3. The figure also shows the position of the subject head relative to the source, with $\theta=0°$ denoting a position of the subject's right ear pointing directly to the source. The experiment included head positions of $\theta=0°$, 90°, and 180°, to test the performance of the active headset in different orientations of the reference microphone relative to the source, with $\theta=0°$ denoting ''upstream'' orientation (reference microphone nearer the source), and $\theta=180°$ denoting ''downstream'' orientation (reference microphone furthest from the source).

In the experiment with the direct field conditions, performed in the laboratory room, the loudspeaker was placed in the center of the room, about 0.5 m away from the subject, ensuring that the direct sound field is dominant, as illustrated in Fig. 3. Head positions of $\theta=0°$, 90°, and 180° were also investigated here, as discussed above and illustrated in Fig. 3.

C. Results

The power spectral density of the signal at the error microphone was measured in both the reverberation chamber and the laboratory room for the three head positions and with three operation modes of the active headset, as described above. The measured power spectral density recorded by the spectrum analyzer is presented in a 1/3-octave format in Fig. 4. The right column presents results under the reverberant conditions, while the left column presents results under the direct field conditions. In both columns the results for the three head positions are shown ($\theta=0°$, 90°, and 180°), while in each of the plots Figs. 4(a)–(f), the thin curve denotes control off; the normal curve denotes analog feedback control on; and the thick curve denote both controllers on. All plots show the sound-pressure level in decibels as a function of the center frequency of each 1/3-octave band.

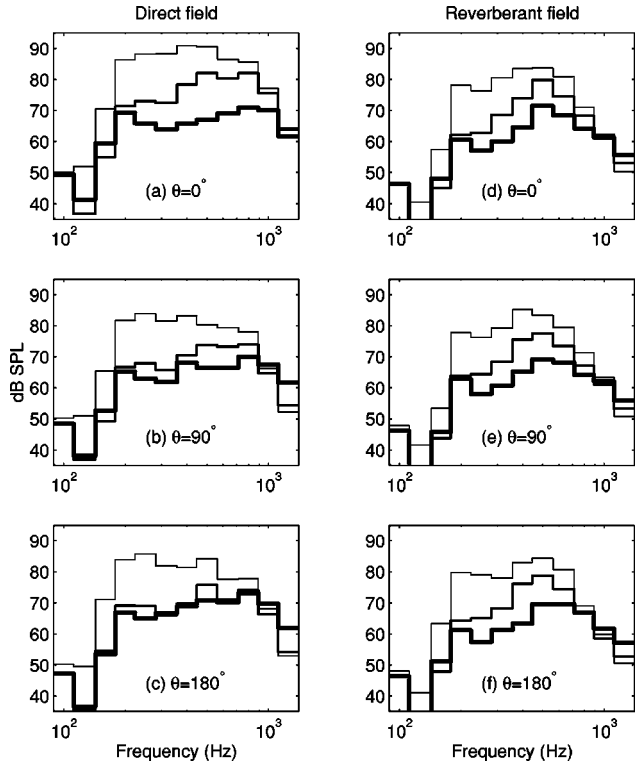


FIG. 4. Measured 1/3-octave power spectral density of the error microphone signal before control (thin line), after control with the analog feedback controller (normal line), and after control with the combined systems (thick line), in the direct field condition, with the head at positions (see Fig. 3) (a) $\theta=0°$; (b) $\theta=90°$; (c) $\theta=180°$; and in the reverberant field condition, with the head at positions (d) $\theta=0°$; (e) $\theta=90°$; (f) $\theta=180°$. A noise source of bandwidth 200–900 Hz was used.

The power spectral density at the error microphone with no control in all plots is concentrated at the frequency range 200–900 Hz, as expected. The spectrum after control with the analog feedback controller is reduced, mainly at the low frequencies, in all plots. This is expected, since the analog controller is designed to reduce the noise at the lower frequencies, with poor performance usually around 1 kHz and above. A more detailed analysis of the attenuation of the analog controller is presented below. The plots also show that the spectrum at the error microphone is further reduced once the digital feedforward controller is turned on. However, the amount of reduction is significantly different under the various conditions. While under the reverberant sound field conditions a similar reduction is achieved for all three head positions, under the direct field conditions a much better reduction is achieved for $\theta=0°$ than for $\theta=180°$.

Figure 5 further illustrates the results presented above by presenting the attenuation levels achieved by the analog feedback controller (left column) and digital feedforward controller (right column, values excluding the contribution of the analog controller), in both reverberant and direct sound field conditions, for all three head positions ($\theta=0°$—thin curve, $\theta=90°$—normal curve, $\theta=180°$—thick curve). The figure confirms that an attenuation of up to 15 dB is achieved by the analog feedback controller at the low frequencies, and that the attenuation is similar for all head positions, in both direct and reverberant sound fields, as expected (see Sec. III A). It should be noted that a slightly narrower attenuation

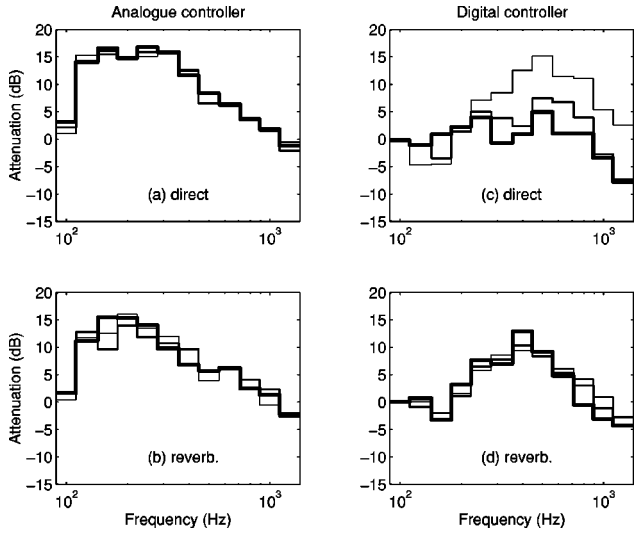


FIG. 5. Measured 1/3-octave attenuation at the error microphone for the head at positions (see Fig. 3) $\theta=0°$ (thin line); $\theta=90°$ (normal line); and $\theta=180°$ (thick line); for the analog feedback controller under (a) direct sound field; (b) reverberant sound field; and for the digital feedforward controller (excluding the attenuation of the analog controller) under (c) direct sound field; (d) reverberant sound field. A noise source of bandwidth 200–900 Hz was used.

bandwidth is achieved by the analog feedback controller under the reverberant field conditions compared to the direct field conditions. Although this is not really expected, it could be explained by the fact that the experiments at the reverberation chamber and lab rooms were separate, and so differences in the headset fitting could account for differences in the response of the analog plant P resulting in differences in the attenuation values.

Figure 5 also clearly shows that the digital feedforward controller achieved an attenuation of up to about 13 dB in the reverberant conditions, and that this result is similar for all head positions. This is expected, since the reverberant sound field is composed of many plane waves arriving from various directions so the properties of the surrounding sound field should not change significantly as a function of direction or position. Nevertheless, the slight variation with position can be explained by the slight directivity in this sound field, which is not strictly reverberant, as discussed above. However, in the direct sound field conditions, it is shown clearly that a much higher attenuation is achieved with the reference microphone facing the source ($\theta=0°$) and the poorest attenuation with the reference microphone facing the other way ($\theta=180°$).

The results of Fig. 5 are further summarized in Table I, which presents the overall attenuation (reduction in the power of the error signal at the internal microphone) of both controllers in the two sound-field conditions and three head positions (note that the attenuation values of the digital controller exclude the contribution of the analog controller). The analog feedback controller achieved an attenuation of around 10 dB in the direct field condition, and over 7 dB in the reverberant field condition. These differences could be explained by the different attenuation curves of the analog controller, as discussed above, together with differences in the power spectral density of the noise in the two sound fields.

TABLE I. Overall attenuation levels in decibels, corresponding to the attenuation curves presented in Fig. 5, of the analog feedback controller (analog) and the digital feedforward controller (digital, excluding the contribution of the analog controller) in the combined system, for various values of θ (see Fig. 3) under direct field and reverberant field conditions, with the noise source of bandwidth 200–900 Hz.

Sound field	Analog	Digital
Direct $\theta=0°$	9.3	10.1
Direct $\theta=90°$	10.0	4.0
Direct $\theta=180°$	10.7	1.7
Reverberant $\theta=0°$	6.8	7.4
Reverberant $\theta=90°$	7.9	7.0
Reverberant $\theta=180°$	7.1	7.2

The digital feedforward controller achieved attenuation of over 7 dB in the reverberant field condition, but attenuation varied from 10.1 dB ($\theta=0°$) to 1.7 dB ($\theta=180°$) in the direct field conditions.

The results presented above confirm that the feedforward digital controller provided some additional attenuation over a broad frequency range, verifying that the combined system is able to improve broadband performance compared to the analog feedback controller alone. The incorporation of the feedforward digital controller can be used therefore not only to attenuate periodic noise components, as previously demonstrated,[10] but also broadband noise components, as demonstrated previously for a stand-alone digital feedforward active headset.[8]

The effect of the surrounding sound field on the feedforward controller is now studied by comparing the impulse response of the real-time FIR control filter W after convergence under the various sound-field conditions, with the subject's head in positions of $\theta=0°$ and 180°. Figure 6(a) shows that the control filters for these two head positions are indeed different under direct field conditions, while Fig. 6(b) shows that they are similar under reverberant field conditions. The behavior of the control filters presented here is in agreement with the noise attenuation results presented above, i.e., the performance of the feedforward controller is significantly more sensitive to head position in the direct field compared to the reverberant field.

The significant variation in the attenuation as a function of head position in the direct field can be explained by the fact that at a position of $\theta=0°$ the sound field propagates from the reference microphone directly to the error microphone, with a maximum acoustic delay between the two microphones and best performance of the feedforward controller.[13] However, at positions of $\theta=90°$ and $\theta=180°$ the reference microphone has less advanced information, if any at all, with expected reduction in performance of the feedforward controller. In the reverberant sound-field conditions the performance is not significantly varied with head position, since the many reflections ensure that no position provides significantly different advance information in the reference microphone. The analog feedback controller, as discussed above, uses a single microphone detecting the pressure inside the headset earshell, and its performance is therefore not significantly affected by the surrounding sound field. A similar behavior of a combined feedback–feedforward system was

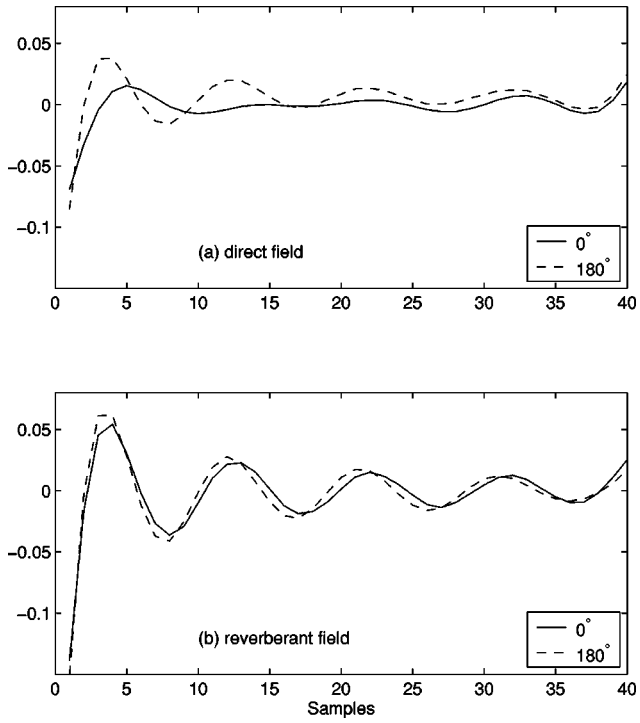


FIG. 6. Impulse response of the FIR feedforward control filters W after convergence, with the head at positions θ=0° (solid curve) and 180° (dashed curve) in (a) direct sound-field conditions and (b) reverberant sound-field conditions.

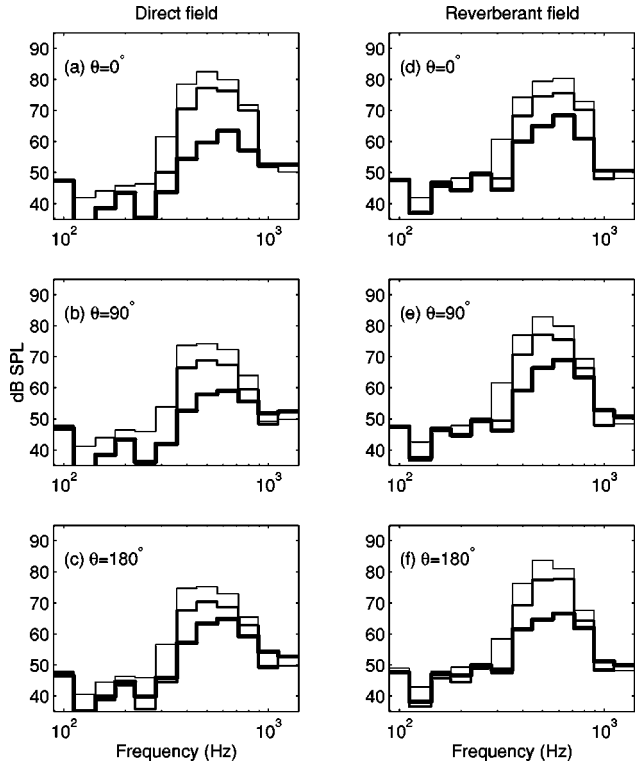


FIG. 7. Measured 1/3 octave power spectral density of the error microphone signal before control (thin line), after control with the analog feedback controller (normal line), and after control with the combined systems (thick line), in the direct field condition, with the head at positions (see Fig. 3) (a) θ=0°; (b) θ=90°; (c) θ=180°; and in the reverberant field condition, with the head at positions (d) θ=0°; (e) θ=90°; (f) θ=180°. A noise source of bandwidth 400–700 Hz was used.

illustrated by Tseng *et al.*[15] for local active sound control in a room, where it was shown that the performance of the feedforward system, unlike that of the feedback system, was highly dependent on position relative to the primary source.

D. Performance with a narrower-bandwidth noise

The experiment described above was repeated in this section, but this time with a noise source with a narrower bandwidth of 400–700 Hz, compared to 200–900 Hz above. The results are presented in Figs. 7 and 8 and Table II, in a similar way to the corresponding Figs. 4 and 5 and Table I.

Figure 7 shows that results are similar to those obtained with a broadband noise, i.e., the performance of the analog feedback controller is similar in all cases, while the performance of the digital feedforward controller is similar only in the reverberant field conditions, and performance is better at θ=0° compared to θ=180° in the direct field conditions. Nevertheless, the digital controllers seem to perform better at a head position of θ=180° compared to the broadband noise case above. These results are further confirmed by the attenuation curves in Fig. 8. Note that the attenuation values of the controllers at frequencies lower than 400 Hz cannot be regarded as correct due to the low signal levels and the poor signal-to-measurement noise ratio at this frequency range. The overall attenuation values in Table II confirm that the digital feedforward controller is performing better than in the broadband case, in particular under the direct field conditions, while the analog feedback controller is performing slightly worse than in the broadband noise experiment. The latter can be explained by the fact that although the attenuation values of the analog controller at every frequency are the same, the noise level in this case is more significant at the higher frequencies, where the analog controller has poorer attenuation value, and so the overall attenuation of the noise is worse than that achieved for broadband noise.

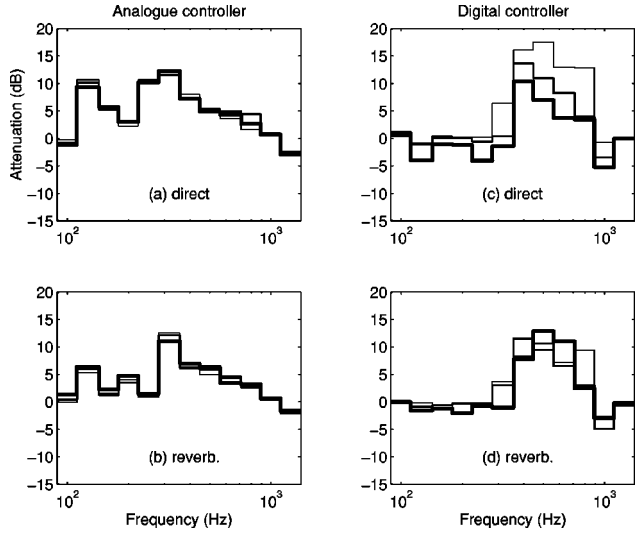


FIG. 8. Measured 1/3-octave attenuation at the error microphone for the head at positions (see Fig. 3) θ=0° (thin line), θ=90° (normal line), and θ=180° (thick line), for the analog feedback controller under (a) direct sound field; (b) reverberant sound field, and for the digital feedforward controller (excluding the attenuation of the analog controller) under (c) direct sound field; (d) reverberant sound field. A noise source of bandwidth 400–700 Hz was used.

TABLE II. Overall attenuation levels in decibels, corresponding to the attenuation curves presented in Fig. 8, of the analog feedback controller (analog) and the digital feedforward controller (digital, excluding the contribution of the analog controller) in the combined system, for various values of θ under direct field and reverberant field conditions, with the noise source of bandwidth 400–700 Hz.

Sound field	Analog	Digital
Direct $\theta=0°$	4.9	14.1
Direct $\theta=90°$	5.7	8.6
Direct $\theta=180°$	5.2	5.5
Reverberant $\theta=0°$	4.7	8.0
Reverberant $\theta=90°$	5.4	8.1
Reverberant $\theta=180°$	5.2	10.5

These improved results of the digital feedforward controller can be explained by the fact that narrow-band signals can be predicted more easily, and so good performance can be achieved even when the electrical delay is longer than the acoustic delay. The performance of the digital feedforward controller with narrower band noise will therefore be less affected by the acoustic delays and the surrounding sound field, with the performance in attenuating periodic noise being unaffected by these factors.

The performance of the digital feedforward controller in practice will therefore depend on the electronic delay of the control system, the acoustic delay affected by the surrounding sound field, and the bandwidth or predictability of the noise.

E. Performance with the head in motion

The experiments presented above studied performance when the subject was stationary in various positions, so that the feedforward controller had time to adapt, and clearly demonstrated the effect of the surrounding sound field. The experiment presented in this section investigated the performance under more realistic conditions when the subject's head is not stationary. The experimental procedure follows that described in Sec. III B, except that here the subject's head is turning repeatedly from side to side, starting from a position of $\theta=0°$ (as in Fig. 3), through position $\theta=90°$ to position $\theta=180°$, and then back to $\theta=90°$ and $\theta=0°$. The speed of head rotation was about 90° per 1 second. The analog feedback controller was operating throughout this experiment, while the digital feedforward control filter was initialized to zero at the beginning of the experiment, and allowed to converge throughout the experiment.

Figure 9 shows the mean-squared error signal at the internal microphone, as measured over about 10 s from the start of the experiment, and then averaged over 100-ms-long blocks of data. The error signal was also bandpass filtered in the range 100–1000 Hz to remove any signal components outside the bandwidth of the generated disturbance. Results taken in both the reverberant conditions and the direct field conditions are presented. An error signal with a higher level is observed in the first second or so, where the head is kept at position 0° and only the feedback controller was operating (the adaptation of the feedforward controller and the head motion started after about 1 s from the start of the measurement). After the feedforward system was switched on, the feedforward control filter converged within about 200 ms, and greater reduction in the error was observed with the combined system operating.

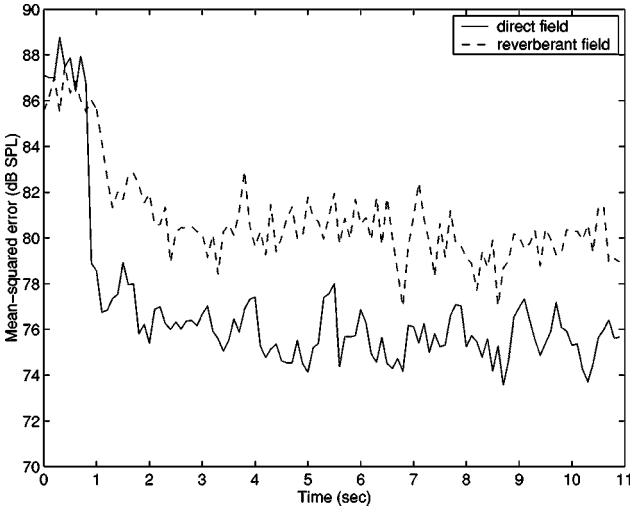


FIG. 9. Mean-squared error signal measured with the combined system switched on after about 1 s, and the subject's head turning repeatedly from a position of $\theta=0°$ to $\theta=180°$ and back, at a rate of about 90° per s, under both reverberant conditions and direct field conditions.

Figure 9 shows that higher error levels are observed in the reverberant conditions compared to the direct field conditions, which confirms the experiments with a stationary head, while both error signals show a reasonably steady level. It is interesting to note that although the latter is expected in the reverberant field conditions, it is less expected in the direct field conditions where significant differences in attenuation were demonstrated for various stationary head positions. However, in this experiment the head is moving and the feedforward controller adapting continuously, and so the overall level of the error signal could depend on several factors, such as convergence speed, the level of the noise in the room at the different head positions, and the attenuation performance of the feedforward controller as a function of time. In this example the effect of all these factors seemed to ''average out'' to produce a relatively steady error level, although this could be different for other conditions and systems.

IV. FORWARD PATH DELAY AND PERFORMANCE

As discussed by Nelson and Elliott,[13] good broadband performance in feedforward control requires that the acoustic delay between the reference and error microphones is not smaller than the corresponding electrical delay in the control path. The results of the experiments as described above showed that the performance of the digital feedforward controller in the direct sound field was dependent on position, which was explained by the reduced amount of prior information in the reference microphone, caused by the reduced acoustic delay between the two microphones in the head position of $\theta=180°$. The delay between the two microphones is therefore further investigated in this section, showing that the headset dynamics responsible for the passive sound transmission increases the delay, compared to the delay of sound propagating in open space.

The electrical delay of the digital controller is estimated first. This delay, which is equivalent to the delay of the plant G in the combined control system, as in Eq. (1), includes about one sample delay of the digital system and about 2.2 samples equivalent delay for each 8th-order low-pass filter,[13] adding to an overall delay of about 5.4 samples, which is equal to about 0.45 ms at the sampling rate of 12 kHz. Furthermore, the group delay of the transducers and the acoustic delay from the loudspeaker to the microphone will also increase the effective delay of the plant G.

On the other hand, the acoustic delay between the two microphones is estimated from the cross correlation between the two microphone signals, by comparing the position of the peak in the correlation curves calculated from experimental data. The experiment used the active headset system as described in Sec. II, with the headset fitted to a wooden block of rectangular base of dimensions 10×12 cm, and an inverted trapezoid face 14 cm high, 10-cm-long lower base, and 14-cm-long upper base. The reason for using a wooden block and not a human head is to improve the control over the headset fit and seal. The headset, fitted to the wooden block, was located about 25 cm from a loudspeaker generating 200–900-Hz bandlimited noise as described in Sec. III B, and was therefore experiencing a direct sound field. The experiment was conducted in a $6.0\times2.1\times2.1$-m semianechoic enclosure, with wooden surfaces and various absorbing materials, which also assisted in minimizing the effect of reflections.

Two fittings of the headset were investigated. One tight fitting, where it was ensured that the headset cushion scaled the gap between headset shell and the wooden block, was called ''earshell closed.'' In the other fitting the earshell cushion was displaced 1 cm away from the wooden block surface, called the ''earshell open'' fitting. Both fittings were studied with the headset positioned at $\theta=0°$ (reference microphone facing the loudspeaker) and $\theta=180°$ (reference microphone facing away from the loudspeaker). By comparing the earshell-closed fitting to the earshell-open fitting in the two positions, one could study the influence of the passive sound transmission through the headset on the delay and how this changed with headset positions.

The signals from the two microphones were acquired by a spectrum analyzer (Advantest R9211C), and the normalized cross correlation between the two microphone signals calculated using MATLAB. The results are presented in Fig. 10. Figure 10(a) shows that the peak of the correlation for the earshell-open case is at about 0.3 ms, corresponding to a spacing of about 10 cm, which is a reasonable approximation of the actual spacing. In this case, the acoustic delay is lower than the electrical delay approximated above, and broadband performance is expected to be poor, which contradicts the measured results. However, in practice the headset is fitted tightly to the head, as in the earshell-closed condition, and so the sound transmission path between the two microphones passes through the passive mechanism of the headset, by vibrating the earshell mass.[16] The figure shows that in this case the delay between the two microphones has increased to about 0.7 ms, which is now longer than the electrical delay and can explain the good measured attenuation in this case.

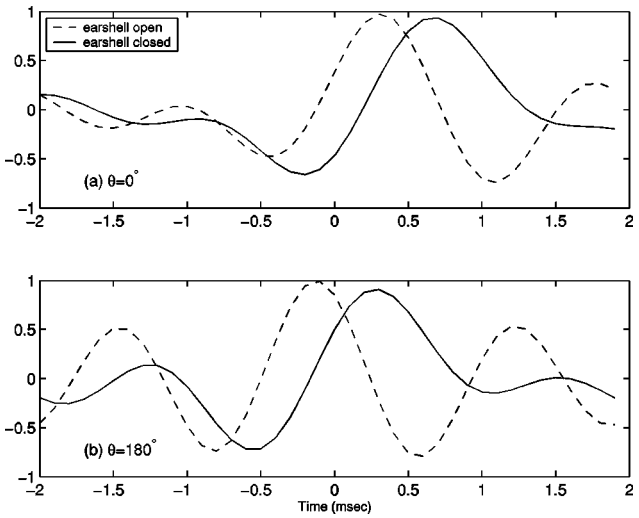


FIG. 10. Measured cross correlation between the external reference microphone and the internal error microphone, with the headset fitted tightly to the head (earshell closed), and fitted to allow a 1-cm-wide gap (earshell open), with a headset position of (a) $\theta=0°$ and (b) $\theta=180°$ relative to the loudspeaker.

Figure 10(b) shows the correlation results with the headset at a position of $\theta=180°$. The delay in the earshell-open condition is -0.1 ms, while the delay in the earshell-closed position is 0.3 ms. The negative delay value confirms that the sound from the loudspeaker reaches the internal error microphone before it reaches the reference microphone, although the difference is smaller than in the $\theta=0°$ position. A possible explanation is the additional delay into the internal error microphone introduced by the response into the earshell, which although not completely sealed might still introduce phase lag. However, once the headset is fitted tightly as in the earshell-closed condition, the delay is increased to 0.3 ms. This means that the loudspeaker signal now reaches the reference microphone first, although it is physically further away from the loudspeaker compared with the internal error microphone. Again, this is evidence of the increased delay due to the passive sound transmission response. Notice that in both positions [Figs. 10(a) and (b)], the increase in delay due to the passive sound transmission is about 0.4 ms.

The results presented above suggest that, provided the electrical delay is sufficiently small (e.g., less than 0.3 ms in this case), the reference microphone could provide advance information regardless of the surrounding sound field, and so some broadband performance could be achieved at all positions. This requires, however, a tightly fitted headset. A more comprehensive study of the relation between the effect of imperfect fit (leaks), noise bandwidth, and electrical delay on performance is suggested for future work.

Figure 11 shows the measured magnitude and phase response of the sound transmission through the headset, calculated as the transfer function between the external reference microphone and the internal error microphone, with the headset fitted tightly on the wooden block as described above, and the reference microphone facing the source (position $\theta=0°$). Of particular importance is the phase response, showing clearly the phase lag responsible for the group delay as discussed above.

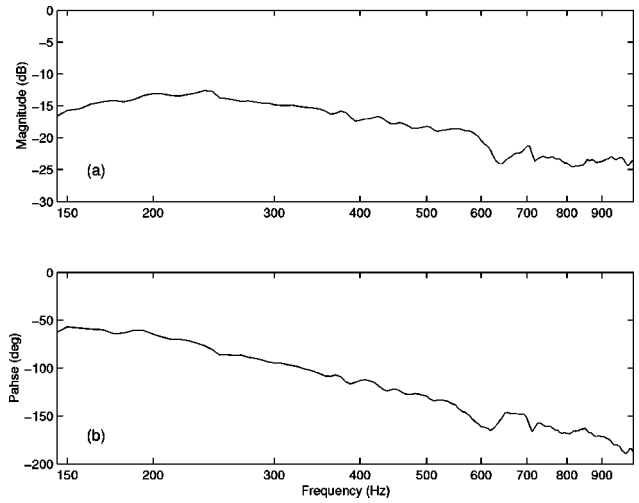


FIG. 11. Measured (a) magnitude and (b) phase response of the sound transmission through the headset described in Sec. II, calculated as the transfer function between the external reference microphone and the internal error microphone.

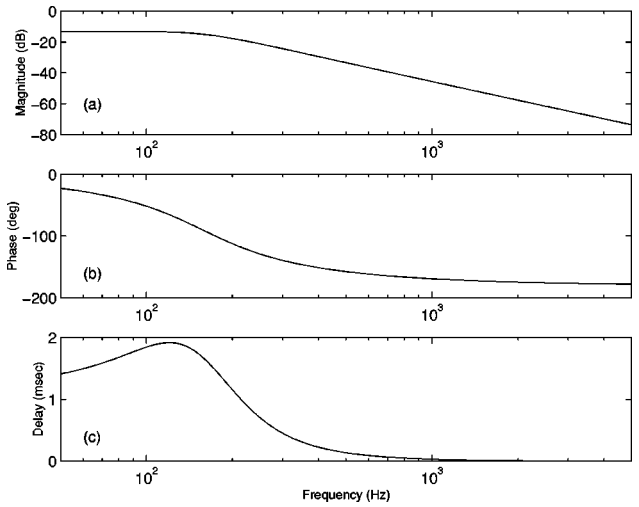


FIG. 12. Simulated (a) magnitude and (b) phase response of the sound transmission through a passive noise-reducing headset, calculated using Eq. (2). The group delay, calculated using Eq. (3), is shown in (c).

The explanation that the passive mechanisms increase the acoustic delay can be supported by theoretical analysis. The sound transmission at low frequencies through a passive noise-reducing headset, from the external pressure P_0 to the internal pressure P_1, can be written as a function of the headset parameters as[16]

$$\frac{P_1}{P_0} = \frac{K_v}{K_v + K + j\omega R - \omega^2 M}, \tag{2}$$

where K and R are the stiffness and damping of the earshell cushion, M is the earshell mass, and K_v is the stiffness of the volume inside the earshell. The sound transmission has the response of a second-order system, and the group delay of this response can be calculated from the phase response as

$$\tau_g = -\frac{d\phi(\omega)}{d\omega}, \tag{3}$$

where ϕ is the phase of P_1/P_0. It should be noted that this theoretical model is only valid at low frequencies, since acoustic resonances inside the earshell and mechanical resonances of the earshell structure are neglected. It is therefore presented here to support the understanding of the phenomenon of increased delay and not as an accurate model of the experimental system used here. The latter would require a more comprehensive modeling and is beyond the scope of this work.

The group delay τ_g has been simulated using Eqs. (3) and (2) with the values of the parameters obtained in previous studies.[17,18] The values represent a commercial passive noise-reducing headset, JSP model KMO7236, and are given by $M = 77$ g, $K = 7 \cdot 10^4$ N/m, $R = 90$ Ns/m, and $K_v = 5 \cdot 10^3$ N/m.

Figures 12(a) and (b) show the magnitude and phase response of the sound transmission simulated by Eq. (2). A second-order system behavior is observed, with the corresponding phase shifting from 0° to −180°. This phase lag is then translated to group delay using Eq. (3), and presented in Fig. 12(c). The group delay takes values of about 0.04 to 1 ms in the disturbance frequency bandwidth of 200–900 Hz. This is within the same order of the delay values obtained experimentally, as shown in Fig. 10 above. This provides a theoretical support to understanding the cause of the increased delay. Nevertheless, a more comprehensive model is required to predict the delays measured in the experimental system.

The experiment and theoretical analysis presented in this section show that, due to the passive mechanism of the headset, the effective delay between the two microphones is increased, and good broadband performance can be readily achieved in spite of the small spacing between the two microphones.

V. CONCLUSIONS

This paper presented an experimental study of the broadband performance of a combined feedback–feedforward active noise-reducing headset. It was shown that good broadband performance can be achieved in a reverberant enclosure regardless of subject position as expected, while under direct sound field conditions broadband performance depends on subject position, with the best performance achieved when the external reference microphone is placed upstream of the propagating sound field. The effect of the surrounding sound field was shown to be reduced when the bandwidth of the noise is reduced, with no expected effect for periodic noise components. It was also shown that the passive mechanism of the sound passing through the headset earshell improves the performance of the feedforward system by increasing the delay between the two microphones. The results presented here can be used to enhance the performance of combined or feedforward active headsets in realistic acoustic environments.

ACKNOWLEDGMENTS

The authors would like to thank Lectret Precision Pte. Ltd. for supporting the research work.

[1] E. D. Simshauser and M. E. Hawley, ''The noise canceling headset—An active ear defender,'' J. Acoust. Soc. Am. **27**, 207 (1955).
[2] M. H. Hawley, ''Acoustic interference for noise control,'' Noise Control **2**, 61–63 (1956).
[3] W. F. Meeker, ''Component characteristics for an active ear defender,'' J. Acoust. Soc. Am. **29**, 1252 (1957).
[4] J. C. Casali and G. S. Robinson, ''Narrow-band digital active noise-reduction in a siren-canceling headset—Real-ear and acoustical mannequin insertion loss,'' Noise Control Eng. J. **42**(3), 101–115 (1994).
[5] Y. Shimada, T. Fujikawa, Y. Nishimura, T. Usagawa, and M. Ebata, ''Active control of the sound of siren of ambulance using active headset,'' Proceedings of the Inter-noise 99 Conference, 6–8 December 1999, Fort Lauderdale, FL, pp. 651–654 (1999).
[6] B. Rafaely, ''Feedback control of sound,'' Ph.D. thesis, University of Southampton, England (1997).
[7] M. Pawelczyk, ''Multirate signal processing for broadband noise cancellation in active ear defender,'' Proceedings of the Third International Symposium on Methods and Models in Automation and Robotics, September 1996, Miedzyzdroje, Poland, pp. 10–13 (1996).
[8] A. J. Brammer, G. J. Pan, and R. B. Crabtree, ''Adaptive feedforward active noise reduction headset for low-frequency noise,'' Proceedings of the ACTIVE 97 Conference, Budapest, Hungary, pp. 365–372 (1997).
[9] D. A. Cartes, L. R. Ray, and R. D. Collier, ''Experimental evaluation of leaky least-mean-square algorithms for active noise reduction in communication headsets,'' J. Acoust. Soc. Am. **111** (4), 1758–1771 (2002).
[10] W. Winberg, S. Johansson, T. Lago, and I. Claesson, ''A new passive/active hybrid for a helicopter application,'' Int. J. Acoust. Vib. **4**(2), 51–58 (1999).
[11] C. Carme, ''The third principle of active control: The feed forback,'' Proceedings of the ACTIVE 99 Conference, December, 1999, Ft. Lauderdale, FL, pp. 885–896 (1999).
[12] B. Widrow and S. D. Stearns, *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1985).
[13] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).
[14] ISO 3741, ''Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision methods for reverberation rooms,'' International Organization for Standardization (1999).
[15] W. K. Tseng, B. Rafaely, and S. J. Elliott, ''Combined feedback–feedforward active control of sound in a room,'' J. Acoust. Soc. Am. **104**, 3417–3425 (1998).
[16] E. A. G. Shaw and G. J. Thiessen, ''Acoustics of circumaural earphones,'' J. Acoust. Soc. Am. **34**, 1233–1246 (1962).
[17] B. Rafaely, J. Carrilho, and P. Gardonio, ''Novel active noise reducing headset using earshell vibration control'' J. Acoust. Soc. (in print).
[18] B. Rafaely, ''Active noise reducing headset—an overview,'' Proceedings of the Internoise 2001 Conference, August 2001, The Hague, Holland, pp. 589–598 (2001).

Broadband transmission noise reduction of smart panels featuring piezoelectric shunt circuits and sound-absorbing material

Jaehwan Kim[a)]
Department of Mechanical Engineering, Inha University, 253 Younghyun-Dong, Nam-Ku, Incheon 402-751, Korea

Joong-Kuen Lee[b)]
Asia Noise and Vibration Institute, 334-30 Yeonhee 3 Dong, Seodaemoon-Ku, Seoul 120-113, Korea

(Received 28 February 2001; revised 27 November 2001; accepted 1 June 2002)

The possibility of a broadband noise reduction of piezoelectric smart panels is experimentally studied. A piezoelectric smart panel is basically a plate structure on which piezoelectric patches with electrical shunt circuits are mounted and sound-absorbing material is bonded on the surface of the structure. Sound-absorbing material can absorb the sound transmitted at the midfrequency region effectively while the use of piezoelectric shunt damping can reduce the transmission at resonance frequencies of the panel structure. To be able to reduce the sound transmission at low panel resonance frequencies, piezoelectric damping using the measured electrical impedance model is adopted. A resonant shunt circuit for piezoelectric shunt damping is composed of resistor and inductor in series, and they are determined by maximizing the dissipated energy through the circuit. The transmitted noise-reduction performance of smart panels is tested in an acoustic tunnel. The tunnel is a square cross-sectional tube and a loudspeaker is mounted at one side of the tube as a sound source. Panels are mounted in the middle of the tunnel and the transmitted sound pressure across panels is measured. When an absorbing material is bonded on a single plate, a remarkable transmitted noise reduction in the midfrequency region is observed except for the fundamental resonance frequency of the plate. By enabling the piezoelectric shunt damping, noise reduction is achieved at the resonance frequency as well. Piezoelectric smart panels incorporating passive absorbing material and piezoelectric shunt damping is a promising technology for noise reduction over a broadband of frequencies.
[DOI: 10.1121/1.1497367]

PACS numbers: 43.50.Ki, 43.55.Rg [MRS]

I. INTRODUCTION

Noise reduction of vibrating structures is a crucial issue in many engineering systems including automobiles, airplanes, ships, and buildings. In controlling noise, there are two distinct approaches: passive and active methods. Passive methods are based on the design of material properties or shapes of the structures so as to minimize the radiated noise. Exploiting sound-absorbing layers in the structures is a good example in passive noise control.[1] However, since an increasing amount of material is required for effective noise reduction at low frequencies, the passive approach is impractical for low frequency applications. Smart materials or structures have emerged as promising active techniques to reduce the radiated sound field. In such structures, piezoelectric materials are widely used as active devices on the structures. Successful noise reductions have been obtained by using piezoelectric sensors and actuators such that active structural acoustic control has been achieved.[2] However, this approach becomes infeasible at high frequencies due to the increased complexity of controller to take into account many radiating modes of the structure. Moving away from the limited frequency band of noise control, it is natural to combine passive and active approaches. It is based on the idea of utilizing a passive effect at high frequencies and an active effect at low frequencies. An example of this idea is smart foam that uses polyurethane foam for high frequency noise reduction and PVDF actuators with a controller for low frequency noise suppression.[3] Kim *et al.* proposed to combine the piezoelectric smart structure technology with passive sound-absorbing material to suppress the transmitted noise of smart panels over a broadband of frequencies.[4] Instead of using a complicated controller for active structural acoustic control for low frequencies, the piezoelectric smart panel featuring piezoelectric shunt damping and passive sound-absorbing material is proposed in this paper.

Piezoelectric materials have been introduced as a passive damping material to overcome the trade-off in width and height limits of viscoelastic materials. Passive damping schemes using piezoelectric devices were discussed by Forward and have been experimentally demonstrated in an optical system.[5] A fundamental scheme of piezo-damping is the fact that vibration suppression can be achieved by using piezoelectric materials with passive electric network which has inductor or resistor or both. This passive electrical network is

[a)]Electronic mail: jaehwan@inha.ac.kr
[b)]Electronic mail: anvi@anvi.co.kr

called a shunt circuit. Hagood and von Flotow have investigated the possibility of dissipating mechanical energy with piezoelectric material shunted with passive electrical circuits.[6] They optimally tuned the electrical resonance of a shunting circuit to structural resonance in a manner analogous to a mechanical vibration absorber for a selected mode. The use of piezo-damping at selectively chosen modes of a structure would be very effective in reducing the structure-borne noise.

However, there are some difficulties when this approach is applied to vibration suppression for multiple modes because the shunt circuits tuned for individual modes may interact each other. Ho *et al.* have used electrical circuits with piezoelectric elements to suppress the vibration of a beam passively.[7] An electromechanical circuit containing a transformer was modeled by a simple equivalent electrical circuit by using a mathematical equation. Resistive shunt was investigated in terms of coupling coefficient. Hollkamp has expanded the theory of piezoelectric shunting for single mode so that a single piezoelectric element can be used to suppress two modes by optimally designing the shunting parameters.[8] However, the modeling of piezoelectric smart structure applications in the mechanical domain is too complicated to optimally design the shunting parameters. A method for passive piezoelectric shunt damping of multiple vibration modes using a single piezoelectric transducer has been reported.[9] This method is to employ a blocking circuit in series with each parallel shunt circuit. However, since this method requires too many inductors, implementation of the blocking and shunt circuits is not so easy. A method has been presented for implementing a simulated impedance required for effective shunt damping of a resonant structure.[10] The arbitrary impedance has also created the opportunity for more advanced passive controllers that adapt to structural variations, i.e., adaptive controllers. A new tuning method for passive piezoelectric damping, which is based on the electrical impedance model and maximum dissipated energy, has been proposed and its validity has been investigated.[11] In contrast to the conventional tuning method associated with the mechanical vibration absorber, the proposed method uses an electrical impedance model of which the coefficients of the model are found from the measured impedance data. In the tuning process, the optimal shunt parameters are found for maximizing the dissipated energy at the shunt circuit. Since this method is based on the measured electrical impedance data, it can be easily applied to arbitrary shape or size of piezoelectric structures once the electrical impedance is measurable. It is also applicable to multimode piezoelectric damping as discussed in Ref. 11. In the present article, the approach is used to implement single-mode piezoelectric damping.

II. SMART PANELS

Figure 1 illustrates the concept of smart panels. The proposed smart panel is basically a plate structure on which the piezoelectric patch with a shunt circuit is connected (single panel). In addition, passive sound-absorbing material can be bonded on one side of the aluminum plate to enhance the noise reduction in the midfrequency and high frequency regions (single panel with absorbing material). The connected shunt circuit effectively dissipates out the electrical energy produced from the piezoelectric patch as heat energy at the low frequency region. To verify the proposed concept, two smart panels—single and single with absorbing material panels are manufactured. A 300 mm×300 mm×0.8 mm square aluminum plate is used for the host structure and 15-mm-thick polyurethane foam is glued on the plate. A piezoceramic patch (PZT-5) is located at the center of the plate, to take into account the first and second symmetric modes. An electronic shunt circuit composed of a resistor and inductor in series is connected to the piezoceramic patch for resonant shunt.

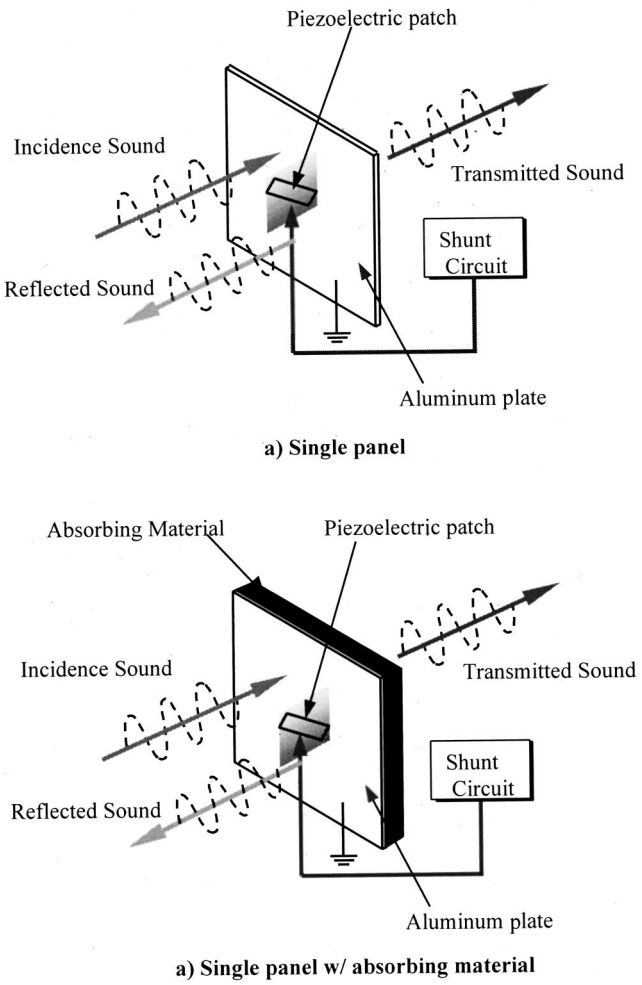


FIG. 1. Schematic diagram of proposed piezoelectric smart panels.

III. PIEZOELECTRIC SHUNT DAMPING

Piezoelectric damping is an energy dissipation mechanism. Open circuited piezoelectric patches store some portion of induced mechanical strain energy as electrical energy, and when a certain resistor is connected across the electrodes, the stored energy begins to flow through the circuit so as to dissipate the energy as heat. In this connection, an inductor is added to cancel the reactive component of the piezoelectric material. The electrical resonance in the shunt circuit with piezoelectric material of which the resonance frequency is corresponding to that of mechanical system,

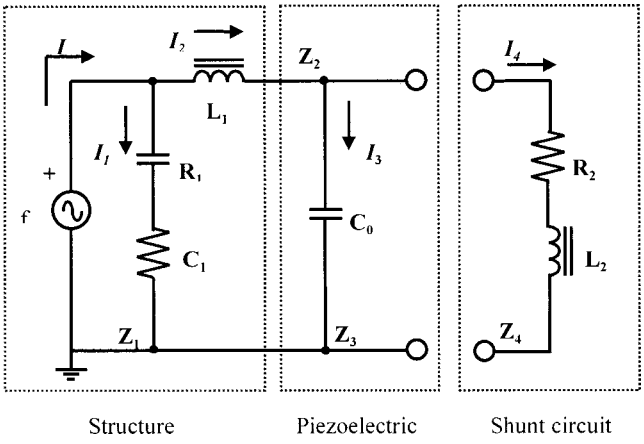


FIG. 2. Equivalent electrical circuit.

helps more charge flow through an adequately chosen resistor. This is the so-called resonant shunt. Since the maximum energy dissipation does not happen at the resonance frequency of the system, the choice of the optimal inductance and resistance is the key. In this paper, the new parameter tuning method based on an electrical impedance model and the maximum dissipation energy criterion is adopted and briefly summarized.[11]

A. Electrical impedance modeling

Piezoelectric materials can be approximately represented as equivalent electric circuits at resonance. There are two resonant models, namely, complex and Van Dyke circuit models. Generally, resonant circuit models are represented in terms of electrical characteristics such as dielectric loss (R_0) and inherent capacitance (C_0), and mechanical characteristics such as mass, damping, and stiffness coefficients (L_1,R_1,C_1). If we assume that the dielectric loss is very small, the Van Dyke model can be represented in terms of four components of which the values are real. An alternative circuit model, the complex model, is based on the lossless resonator model. This model contains only three circuit components, C_0, C_1, L_1, and they are complex. This circuit now has six parameters and some unique features make it an ideal model for representing the electrical characteristics of an unloaded piezoelectric resonator. Near the resonance of the structural system, the system can be represented using the Van Dyke circuit model (Fig. 2). Since the diagram is an impedance representation, the impedance values between the nodes can be written as

$$Z_1=\frac{C_1}{j\omega}+R_1, \quad Z_2=j\omega m_1, \qquad Z_3=\frac{1}{j\omega C_0}, \quad Z_4=j\omega L_2+R_2. \tag{1}$$

Here, C_1, R_1, L_1 represent the inverse of stiffness, damping, and mass of the structure, respectively. Without loss of generality, the Kirchhoff's voltage and current laws are applied to the impedance circuit, and the total electrical impedance and the current at each branch can be written as

$$Z=\frac{Z_1\left[Z_2+\dfrac{Z_3Z_4}{Z_3+Z_4}\right]}{Z_1+\left[Z_2+\dfrac{Z_3Z_4}{Z_3+Z_4}\right]}, \tag{2}$$

$$I_1=\frac{Z}{Z_1}I, \quad I_2=\left(1-\frac{Z}{Z_1}\right)I, \qquad I_3=\frac{Z_4(Z_1-Z)}{Z_1(Z_3+Z_4)}I, \quad I_4=\frac{Z_3(Z_1-Z)}{Z_1(Z_3+Z_4)}I. \tag{3}$$

The transfer function between the current input generated by the piezoelectric element and the current passing through the shunt circuit can be written as

$$\frac{I_2}{I}=1-\frac{Z}{Z_1}. \tag{4}$$

This transfer function is analogous to the transfer function of a mechanical vibration absorber.[6] To use the proposed electrical impedance model of piezoelectric structures, the coefficients of the Van Dyke model should be determined. This can be solved by measuring the electrical impedance of piezoelectric patches bonded on the sturcture directly. When electrical impedance is measured using the impedance analyzer (HP4192A), the coefficients of the Van Dyke or complex circuit model can be determined by using PRAP (piezoelectric resonance analysis program) software.[12] This software allows not only the complex elastic, piezoelectric, and dielectric properties but also the coefficients for Van Dyke's circuit model for piezoelectric or electrostrictive resonators to be determined.

B. Parameter tuning

Parameter tuning is essential to maximize the piezoelectric damping performance. For a resonant shunt, the optimal inductance should be determined to coincide with the resonant frequency of the shunt circuit of the piezoelectric structure. The optimal resistance of the shunt circuit should also be found for maximizing the dissipated energy. Instead of tuning the transfer function geometrically as used in the conventional tuning method, the new parameter tuning method associated with maximum dissipated energy is adopted. From the equivalent electrical impedance model shown in Fig. 2, the induced energy in the structure associated with external excitation is

$$P_{\text{in}}=\tfrac{1}{2}|VI|=\tfrac{1}{2}|(ZI)I^*|=\tfrac{1}{2}|Z||I|^2. \tag{5}$$

Since this energy is continuously generated, it can be referred to as an input energy for the shunt circuit. The dissipated energy at the shunt circuit can be written in terms of the voltage and electrical impedance across the resonant circuit,

$$P_D=\frac{1}{2}\text{Re}[(Z_4I_4)I_4^*]=\frac{1}{2}\text{Re}\left[\frac{Z_3Z_4}{(Z_3+Z_4)}\cdot\frac{(Z_1-Z)}{Z_1}\right]|I|^2. \tag{6}$$

The ratio of the dissipated energy to the input energy can be written at a specific frequency as

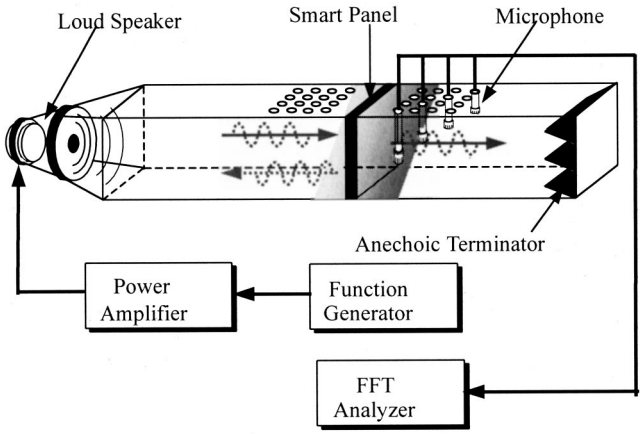


FIG. 3. Experimental apparatus.

$$J=\frac{P_D}{P_{\text{in}}}=\text{Re}\left[\frac{Z_3Z_4}{(Z_3+Z_4)}\frac{(Z_1-Z)}{Z_1}\right]\Big/|Z|. \tag{7}$$

In the tuning process, this should be maximized by optimally changing the shunting parameters. However, the ratio J varies with respect to the frequency. Thus, the objective function in optimization is taken as the averaged J at a certain frequency band near the targeted resonance frequency. The optimal design variables, L^*, R^* are found by maximizing the objective function:

$$[L^*,R^*]=\max_{\substack{L\\R}}\left[\frac{1}{n}\sum_{k=1}^{n}J_k\right], \tag{8}$$

where n is the number of frequency points in the frequency band.

Once an arbitrary piezoelectric smart structure is represented to the electrical impedance model, the model coefficients can be found from the measured electrical impedance and by the use of PRAP software. And the optimal inductance and resistance of the shunt circuit can be found according to Eq. (8).

IV. ACOUSTIC TEST SETUP

To test the noise reduction performance of smart panels, the transmission measurement from low to high frequencies should be available. For most panel materials the transmission loss has been measured under strict control. The experimental procedure most often used follows the ASTM standard E 90 (Standard Recommended Practice for Laboratory Measurement of Airborne Sound Transmission Loss of Building Partitions).[13] Since this test facility is too expensive a simple acoustic tunnel is used instead. The tunnel is a square tube of 300 mm×300 mm and 4 m long (Fig. 3). It is divided into two sections—upper and lower sections in equal length. A loudspeaker is attached at the end of the upper section and an anechoic termination is made with wedges installed at the other end of the lower section. A specially designed flange is provided where two sections are met such that smart panels can be mounted in between. The four edges of the smart panel are clamped using bolts and two sections are tightly connected so as to secure the pressure leak. The tunnel is made of 4.5-mm-thick steel plates and 6-mm-thick polyurethane foam is coated inside of the tunnel to mutilate the side modes in the tunnel. Sixteen holes are made on lower and upper sections of the tunnel to measure the sound pressures across the cross section of the tunnel at four locations. From the repeated measurement of sound pressure distribution at a cross section, the pressure distribution is found to be nearly plane below 800 Hz. Figure 4 represents the incidence pressure curves measured at eight different sec

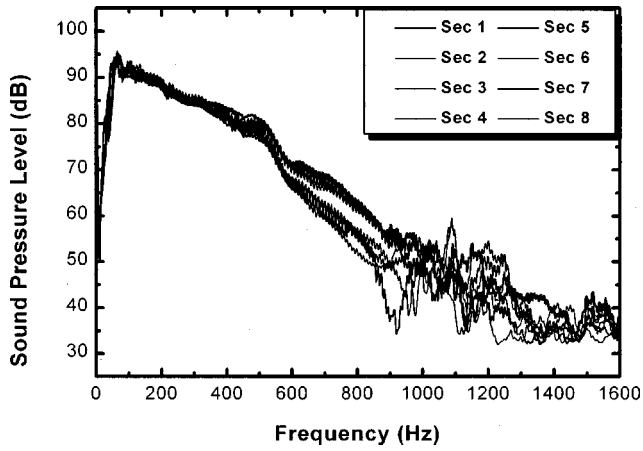


FIG. 4. Incidence sound pressure level at each section.

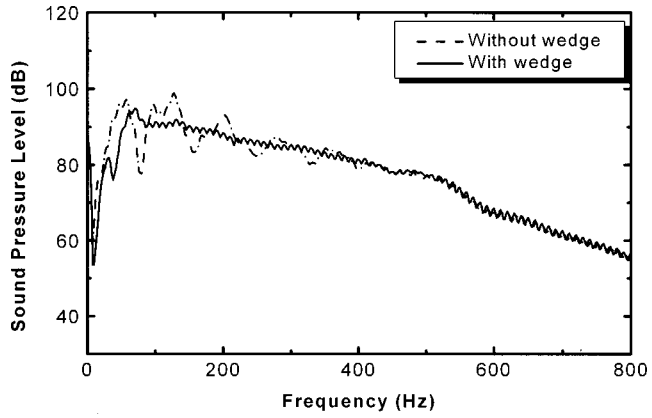


FIG. 5. The effect of anechoic termination in acoustic tunnel.

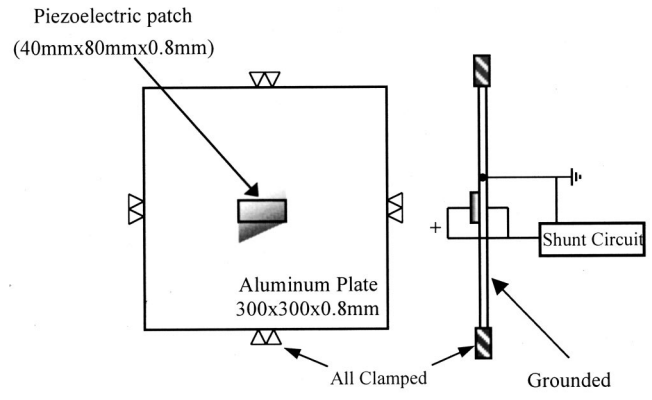


FIG. 6. A schematic diagram of proposed smart panel.

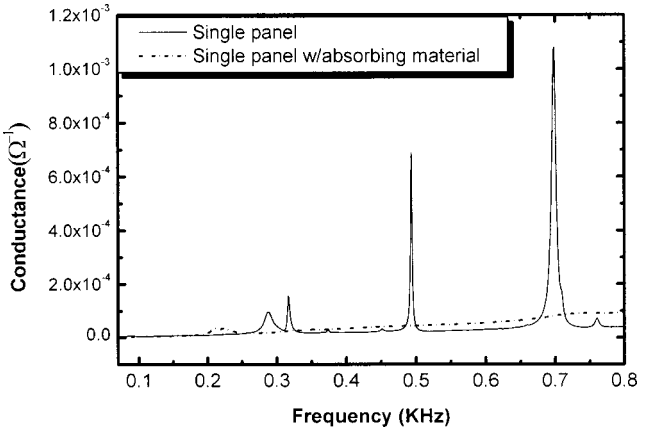
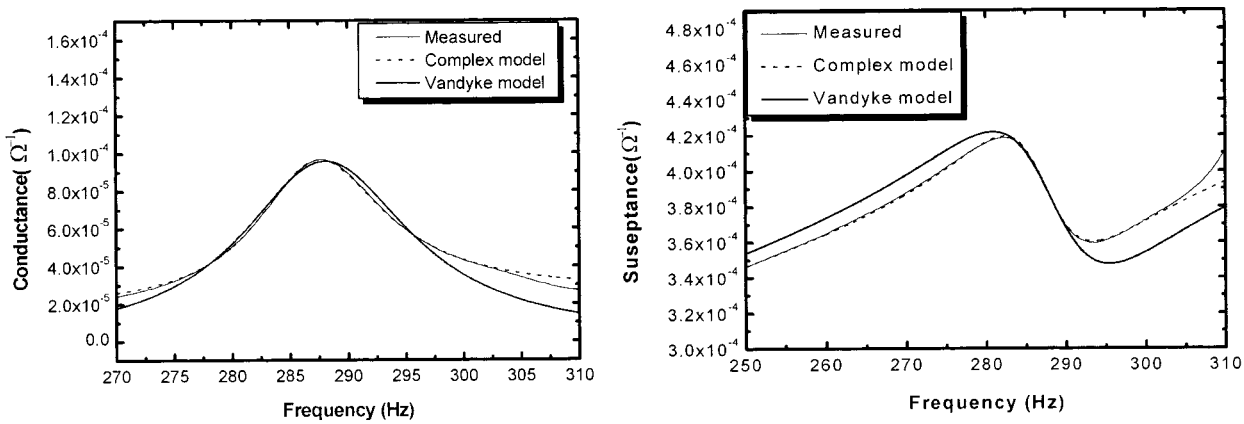


FIG. 7. Measured impedance of smart panels: single panels.

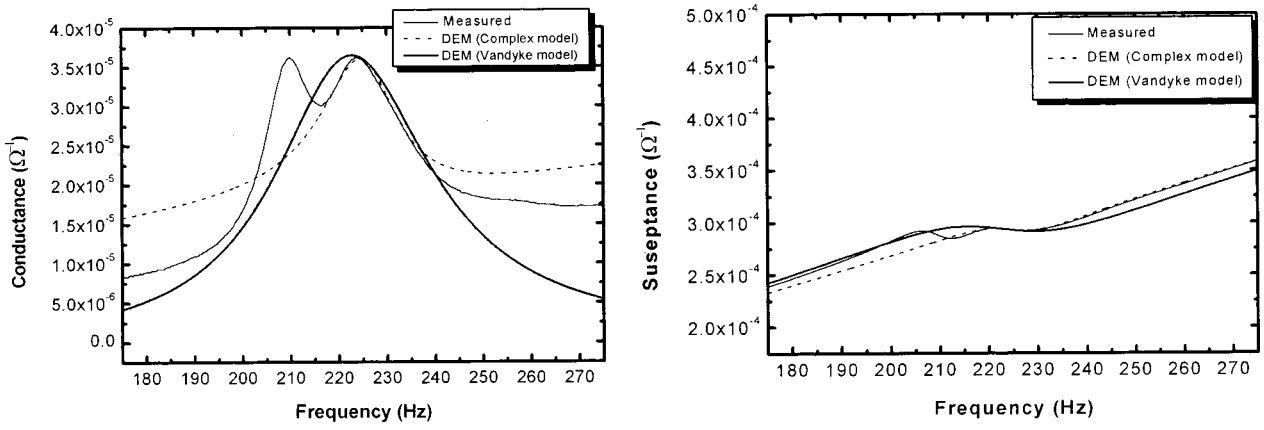
TABLE I. Measured electrical impedance parameters for the smart panels.

	Freq (Hz)	Parameter	Van Dyke	Complex
Single panel	287	C_0(F)	$2.125e-7$	$2.125e-7-j1.177e-8$
		C_1(F)	$3.288e-9$	$1.958e-9+j2.318e-10$
		$R_1(\Omega)$	$1.045e4$	
		L_1(H)	92.82	$153.7-j25.58$
Single panel w/ absorbing material	226	C_0(F)	$2.097e-7$	$2.097e-7-j1.322e-8$
		C_1(F)	$4.609e-9$	$9.931e-10-j3.543e-10$
		$R_1(\Omega)$	$2.74e4$	
		L_1(H)	110.7	$455.6+j120.5$



a) Conductance b) Susceptance

(a) Single panel



a) Conductance b) Susceptance

(b) Single panel w/ absorbing material

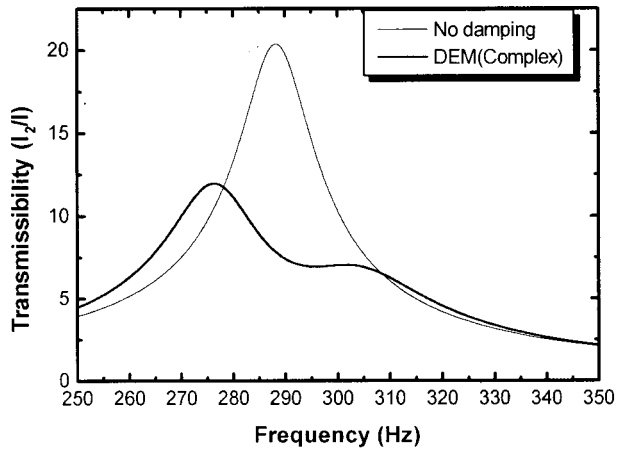
FIG. 8. Simulation result of impedance.

TABLE II. Optimal and measured parameters of the shunt circuit (DEM method).

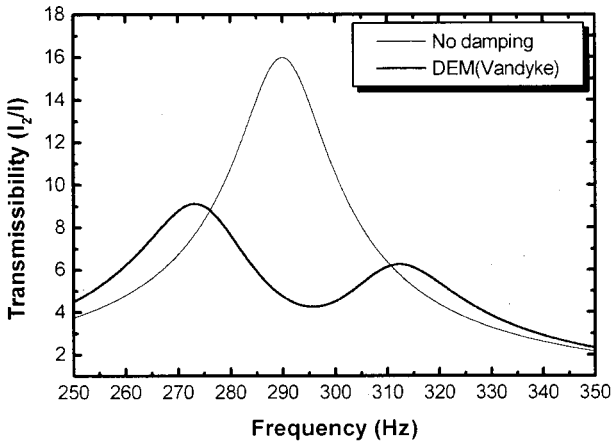
	Freq (Hz)	Parameter	Van Dyke	Complex	Measured
Single panel	287	$R_2(\Omega)$	255	145	48
		L_2(H)	1.355	1.39	1.28
Single panel w/ absorbing material	226	$R_2(\Omega)$	350	128	115
		L_2(H)	2.38	2.33	2.36

tions of the tunnel. Sections 1–4 in Fig. 4 mean the locations in the upper section while Secs. 5–8 indicate the locations in the lower section. In the transmission measurement, four microphones along the diagonal of the lower section are used as shown in Fig. 3. According to the duct noise measurement standard, the average value of four microphone pressures is used to calculate the transmission power. The same idea is adopted to determine the transmission power. From the accumulated measurements, no difference between the sound pressures measured at four microphones is observed at low frequencies but fluctuation happens above 800 Hz. This is due to the pressure variation across the cross section in the tunnel. Thus, incidence wave was assumed to be a plane wave below 800 Hz. Under this limitation, we can use the impedance tube technique or two-microphone measurement technique to measure the reflection coefficient of the panels.

Figure 5 represents the measured incidence pressure inside of the acoustic tunnel with and without the anechoic termination. The cutoff frequency of the termination is 60

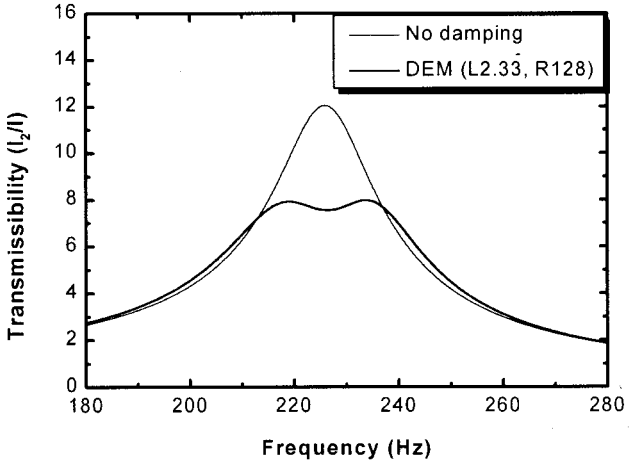


a) Complex model

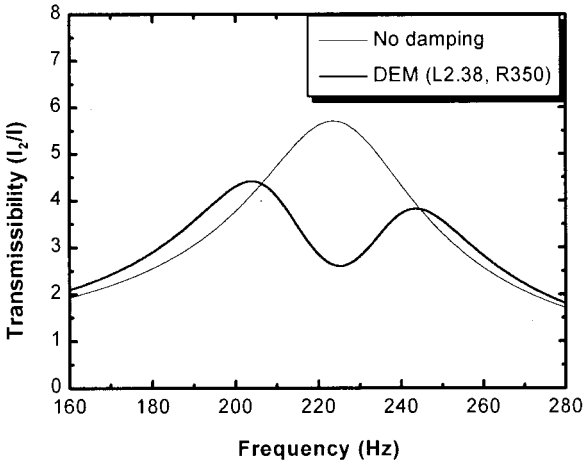


b) Van Dyke model

(a) Single panel



a) Complex model



b) Van Dyke model

b) Single panel w/ absorbing material

FIG. 9. Simulation result of optimal shunt parameter tuning.

Hz. When the termination was not installed, there are peaks and dips in the sound pressure that occur at the longitudinal modes of the tunnel. By installing the termination, they disappear.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Shunt parameter tuning

In the maximum dissipated energy method, the parameters for the electric impedance model for the structure and piezoelectric material are needed. As shown in Fig. 6, a 300 mm×300 mm×0.8 mm square aluminum plate is used for the host structure and a 40 mm×80 mm×0.8 mm piezoceramic patch (PZT-5) is bonded at the center of the plate. This system is termed as single panel. A 15-mm-thick polyurethane foam is bonded on one side of the single panel to enhance the noise reduction performance at midfrequency and high frequency regions. All edges of these panels are clamped into the acoustic test tunnel by using flanges. Figure 7 shows the conductance curves for single panels with/without sound-absorbing material measured by the HP4192A impedance analyzer. Conductance is the real part of admittance, which is the inverse of impedance. The second resonance frequencies for the single and smart panels appear at 296 and 226 Hz, respectively. The first resonance frequencies for the two panels cannot be seen in the conductance curves. This is due to the electrical and mechanical noises coming through the acoustic tunnel. The measured admittance is exported to a personal computer and PRAP software is used to determine the parameters of the equivalent electric impedance model. Table I shows the extracted Van Dyke and complex equivalent models for the first resonance frequency of the panels. Figure 8 represents the simulated admittance curves for the smart panels. Compared with the measured and simulated conductance curves, the complex model fits better than the Van Dyke model. From Fig. 8, it is proved that the established equivalent electrical models precisely describe the characteristics of the panels near the valid resonance frequency. Finally, the optimal parameters of the $R-L$ shunt circuit were obtained by using the dissipated energy method (Table II). Although the results of complex model fit better than the Van Dyke results, the Van Dyke model was used in the optimal parameter tuning process, since the complex model is unproven in the mechanical analogy which needs to be investigated. By using the optimal parameters for the shunt circuit and the equivalent electrical model parameters, the velocity transfer functions for the single and smart panels are simulated (Fig. 9). These simulation results prove that the magnitude of velocity transfer is reduced appropriately by optimally tuning the shunt parameters.

B. The passive method

Before testing the smart panel with piezoelectric shunt damping, sound transmission characteristics of the host plate and the plate with sound-absorbing material are tested (Fig. 10). For the host plate, the first and second resonance frequencies are 87 and 296 Hz, respectively. When the absorbing material is bonded on the plate, the first and second modes are shifted to 86 and 226 Hz, respectively, due to the added mass effect. Through the test of sound characteristic, it is evident that the sound reduction in a midfrequency region is obtained by adding absorbing material, except at the first resonance frequency.

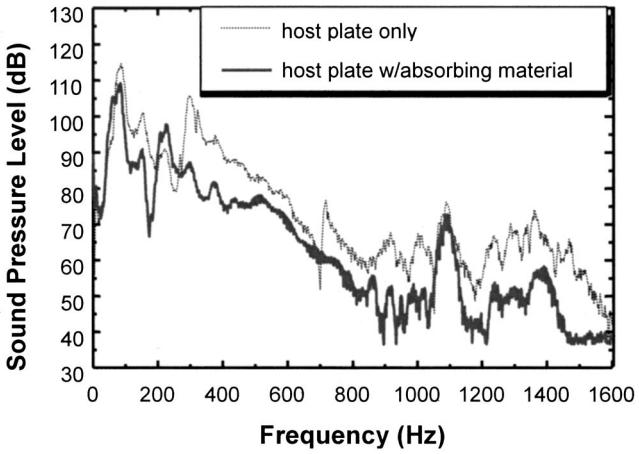


FIG. 10. Transmitted sound pressure level of host plate with absorbing material.

C. Piezoelectric shunt damping

To investigate the effectiveness of the piezoelectric damping for transmitted noise reduction, two smart panels—single panel and single panel with absorbing material—are mounted in the acoustic tunnel. The single panel is a plate structure on which a piezoelectric patch is bonded and an electrical shunt circuit is connected. The selected frequencies are 287 Hz for the single panel and 226 Hz for the single with absorbing material panel. The optimal inductance found is not practically feasible; thus, a synthetic inductor was used.[14] Figure 11 shows the circuit diagram of the synthetic inductor. The synthetic inductor is lightweight and can generate various inductance values, but it also has a resistance proportional to the inductance which is not desirable for designing the optimal resistance in the shunt circuit.[8] The inductance value of the synthetic inductor was measured at the resonance frequency by using the impedance analyzer (Table II). Similar values for the shunt circuit parameters were observed.

Figure 12 shows the sound pressure levels for the single panel on which sound-absorbing material is bonded. When the shunt damping is provided at 287 Hz, the sound pressure

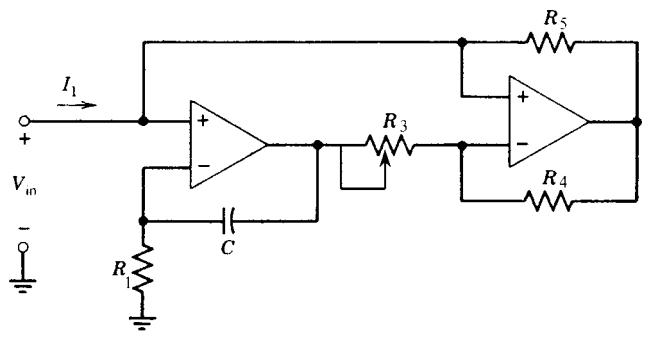


FIG. 11. A schematic diagram of synthetic inductor.

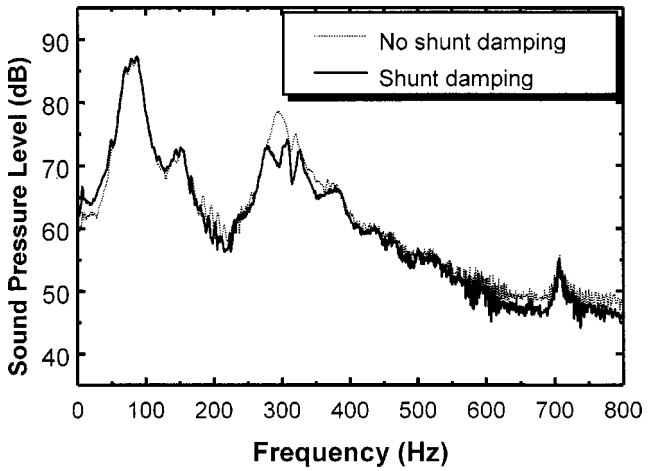


FIG. 12. Experimental result of smart panel: single panel.

level is reduced by 5 dB. Figure 13 represents the comparison of the results in Fig. 12 with the sound pressure level of the host plate only. Since the use of additional sound-absorbing material already reduced the sound pressure level 8 dB down over the second resonant peak, the total reduction at the second peak is 13 dB. In conclusion, the hybrid approach combining the passive shunt damping and sound-absorbing material is a promising technique for noise reduction over a broad frequency band.

The noise reduction level of smart panels may be improved by using high coupling coefficient materials, e.g., IDT piezoelectric patches,[15] single crystal piezoelectric materials,[16] and by optimizing the size of the piezoelectric patches. The merits of this approach are to tune the damping at a specific frequency band and to be able to apply arbitrary shaped structures. This research is at a proof-of-concept stage, and further research efforts are being undertaken to increase the piezoelectric shunt damping. In addition, other configurations of smart panels such as smart double panels should be investigated.

VI. CONCLUSIONS

To reduce the transmission noise over a broad frequency band, the concept of smart panels featuring a piezoelectric shunt circuit and sound-absorbing material was proposed and an experimental study was performed. It is based on the idea of utilizing a passive effect at midfrequency and high frequency and a piezoelectric damping effect at low frequency. In parameter tuning of piezoelectric shunt circuit, the new parameter tuning method associated with the electrical impedance model and maximum dissipated energy was used. To test the noise reduction performance of smart panels, an acoustic tunnel was designed such that it can measure the transmission and reflection of panels. The frequency limits of the facility were examined.

When a sound-absorbing material is bonded on the host plate, a remarkable noise reduction in the midfrequency region was observed, except at the first resonance frequency. Piezoelectric shunt damping was implemented at the second resonance frequency. When the single and smart panels were tested in the acoustic tunnel, 8 and 5 dB of transmission noise reduction were observed.

This approach can be easily applied to arbitrary structures on which piezoelectric patches are bonded and multimode damping with single piezoelectric patch is possible. This is a proof-of-concept stage and additional efforts to increase the piezoelectric damping as well as various configurations of smart panels should be made.

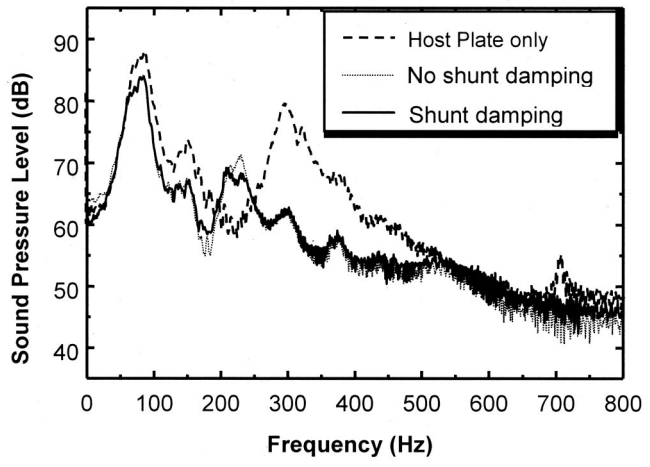


FIG. 13. Experimental result of smart panel: single panel with absorbing material.

ACKNOWLEDGMENT

This work was supported by Cooperative Research under the Korean–American Cooperative Science Program (No. 304-05-2).

[1] J. S. Bolton and N.-M. Shiau, ''Random incidence transmission loss of lined, finite double panel systems,'' 12th AIAA Aeroacoustics Conference, AIAA-89-1048, San Antonio, TX, 1989.
[2] C. R. Fuller, ''Active control of sound transmission/radiation from elastic plates by vibration inputs. I. Analysis,'' J. Sound Vib. **136**, 1–15 (1990).
[3] C. A. Gentry, C. Guigo, and C. R. Fuller, ''Smart foam for applications in passive-active noise radiation control,'' J. Acoust. Soc. Am. **101**, 1771–1778 (1997).
[4] J. Kim, J. Lee, B. Im, and C. Cheong, ''Modeling of piezoelectric smart structures including absorbing materials for cabin noise problems,'' *Sixth Annual Symposium on Smart Structures and Materials, Newport Beach, CA* [Proc. SPIE **3667**, 524–529 (1999)].
[5] R. L. Forward, ''Electric damping of vibrations in optical structures,'' J. Appl. Opt. **18**, 690–697 (1979).
[6] N. W. Hagood and A. von Flotow, ''Damping of structural vibrations with piezoelectric materials and passive electrical networks,'' J. Sound Vib. **146**, 243–268 (1991).
[7] S. T. Ho, H. Matsuhisa, and Y. Hoada, ''Passive vibration suppression of beam with piezoelectric elements,'' JSME Int. J., Ser. C **43**, 740–747 (2000).
[8] J. J. Hollkamp, ''Multimodal passive vibration suppression with piezoelectric materials and resonant shunts,'' J. Intell. Mater. Syst. Struct. **5**, 49–57 (1994).
[9] S. Y. Wu, ''Method for multiple mode shunt damping of structural vibration using a single PZT transducer,'' *Fifth Annual Symposium on Smart Structures and Materials, San Diego, CA* [Proc. SPIE **3327**, 159–168 (1998)].
[10] A. J. Fleming, S. Behrens, and S. O. R. Moheimani, ''Synthetic impedance for implementation of piezoelectric shunt-damping circuits,'' Electron. Lett. **36**, 1525–1526 (2000).
[11] J. Kim, Y.-H. Ryu, and S.-B. Choi, ''New shunting parameter tuning method for piezoelectric damping based on measured electrical impedance,'' Smart Mater. Struct. **9**, 868–877 (2000).

[12] PRAP (piezoelectric resonance analysis program), TASI technical Software, 174 Montreal Street, Kingston, Ontario K7K 3G4, Canada.
[13] L. H. Bell and D. H. Bell, *Industrial Noise Control* (Dekker, New York, 1994).
[14] W. K. Chen, *Passive and Active Filters* (Wiley, New York, 1986).
[15] S. Yoshikawa, M. Farrell, D. Warkentin, R. Jacques, and E. Saarmaa, ''Monolithic piezoelectric actuators and vibration dampers with interdigital electrodes,'' *Sixth Annual Symposium on Smart Structures and Materials, Newport Beach, CA* [Proc. SPIE **3668**, 578–585 (1999)].
[16] S. E. Park and T. R. Shrout, ''Ultrahigh strain and piezoelectric behavior in relaxor-based ferroelectric single crystals,'' J. Appl. Phys. **82**, 1804–1811 (1997).

Auditory brainstem responses in adult budgerigars (*Melopsittacus undulatus*)

Elizabeth F. Brittan-Powell and Robert J. Dooling
Department of Psychology, University of Maryland, College Park, Maryland 20742

Otto Gleich
ENT Department, University of Regensburg, 93042 Regensburg, Germany

(Received 24 September 2001; revised 15 May 2002; accepted 23 May 2002)

The auditory brainstem response (ABR) was recorded in adult budgerigars (*Melopsittacus undulatus*) in response to clicks and tones. The typical budgerigar ABR waveform showed two prominent peaks occurring within 4 ms of the stimulus onset. As sound-pressure levels increased, ABR peak latency decreased, and peak amplitude increased for all waves while interwave interval remained relatively constant. While ABR thresholds were about 30 dB higher than behavioral thresholds, the shape of the budgerigar audiogram derived from the ABR closely paralleled that of the behavioral audiogram. Based on the ABR, budgerigars hear best between 1000 and 5700 Hz with best sensitivity at 2860 Hz—the frequency corresponding to the peak frequency in budgerigar vocalizations. The latency of ABR peaks increased and amplitude decreased with increasing repetition rate. This rate-dependent latency increase is greater for wave 2 as indicated by the latency increase in the interwave interval. Generally, changes in the ABR to stimulation intensity, frequency, and repetition rate are comparable to what has been found in other vertebrates. [DOI: 10.1121/1.1494807]

PACS numbers: 43.64.Ri, 43.64.Tk [WWA]

I. INTRODUCTION

Much is known about hearing in birds from behavioral studies (see the recent review in Dooling *et al.*, 2000). In general, birds hear best between 1000 and 5000 Hz, with absolute sensitivity approaching 0–10 dB SPL in the frequency of best hearing, which is typically around 2000–3000 Hz (see the review in Dooling *et al.*, 2000). However, behavioral testing is not always possible in all birds (e.g., young birds or birds in the field). Over the last few decades, auditory brainstem responses (ABR) have been used as a tool for studying the functionality of the auditory system in a wide variety of mammals (e.g., Burkard *et al.*, 1996a; Burkard and Voigt, 1989; Donaldson and Rubel, 1990; Jewett *et al.*, 1970; McFadden *et al.*, 1996; Mills *et al.*, 1990; Liu and Mark, 2001) and in other nonmammalian vertebrates (e.g., Corwin *et al.*, 1982; Kenyon *et al.*, 1998; Higgs *et al.*, 2002). Overall, these studies show that brain stem responses are similar across most vertebrate classes (e.g., Corwin *et al.*, 1982; Walsh *et al.*, 1992).

To date, ABR studies have only been conducted on a limited number of avian species. While complete reports on the effects of intensity and rate on latency and amplitude of the ABR waves or ABR-generated audiograms have been examined in depth in precocial birds (e.g., Dmitrieva and Gottlieb, 1992; Saunders *et al.*, 1973), much less in known about these measures in altricial birds (Aleksandrov and Dmitrieva, 1992; Moiseff *et al.*, 1996; Woolley and Rubel, 1999; Wooley *et al.*, 2001). To this end, we examined the ABR as a measure of auditory sensitivity in the adult budgerigar. The budgerigar is a small, altricial Australian parrot that is the most widely studied parrot species in the field of acoustic communication (see the review in Farabaugh and Dooling, 1996; Heaton and Brauth, 1999; Hile and Striedter, 2000). This study consisted of two experiments that focused on ABR responses elicited by clicks and tone-burst stimuli in adult budgerigars. The response characteristics examined in experiment 1 were ABR threshold and the effects of intensity on latency and amplitude of the ABR waveform. Experiment 2 examined the effects of presentation rate on wave latency and amplitude in adult budgerigars. These data served as a baseline for assessing the development of hearing in nestling budgerigars.

II. METHODS

Adult budgerigars (1–4 years of age) served as subjects in these experiments. The birds, bought through local pet stores or bred within our colony, were housed in an avian vivarium at the University of Maryland and kept on a photoperiod correlated with the season. They were given free access to food and water. All birds were sedated with an intramuscular injection of ketamine (50 mg/kg) and diazepam (2 mg/kg) prior to electrode placement. The animals remained relatively motionless for up to 75 min. Body temperature was maintained at $41 \pm 0.5\,^{\circ}\mathrm{C}$ with a heating pad and monitored with a thermistor probe placed under the wing (Frederick Haer and Co., model 40-90, Bowdoinham, ME). After data collection was completed, the bird was placed in a heated therapy unit and allowed to recover from sedation. Once the animal was fully revived and eating, it was returned to the vivarium. The Animal Care and Use Committee at the University of Maryland approved all animal use.

The bird was positioned so that the speaker (KEF SP 3235, model 60S, frequency range 100 Hz to 20 kHz, KEF Electronics of America, Inc., Holliston, MA) was 30 cm

from the bird's right ear (90° azimuth relative to the bird's beak; 0° elevation relative to the bird's right ear). Standard platinum alloy, subdermal needle electrodes (Grass F-E2; West Warwick, RI) were placed just under the skin high at the vertex (active), directly behind the right ear canal (the ear ipsilateral to the speaker, reference), and behind the canal of the ear contralateral to stimulation (ground). Shielded electrode leads were twisted together to reduce electrical noise through common-mode rejection. The stimulus presentation, ABR acquisition, equipment control, and data management were coordinated using a Tucker-Davis Technologies (TDT, Gainesville, FL) modular rack-mount system controlled by an optical cable-linked 350-MHz Pentium PC containing a TDT AP2 digital signal process board and running TDT BIOSIG software. Sound stimuli were generated using TDT SIGGEN software and fed through a DA1 digital-analog converter, a PA4 programmable attenuator, and a power amplifier (HB6) which directly drove the speaker. The electrodes were connected to the TDT HS4 Headstage that amplifies and digitizes the signal before being relayed over fiberoptic cables to the TDT DB4 digital biological amplifier. This amplifier allows additional filtering and gain to be added. A TDT TG6 timing generator synchronized the A/D and D/A-conversion.

Stimulus intensities were calibrated in the free field by placing the $\frac{1}{2}$-in. microphone of a sound-level meter (System 824; Larson Davis, Inc., Provo, UT) at the approximate position of the bird's right ear. Continuous tones were generated using the TDT BIOSIG program and measured using the fast-weighting A scale on the sound-level meter (dB SPL). To determine the intensity of the click, we used the peak equivalent SPL of the click. This was determined using an oscilloscope and noting the peak-to-peak voltage of the click. A test tone, e.g., a 1000-Hz tone, was played and adjusted until the peak-to-peak voltage was the same as it was for the click. The SPL required to match the amplitude of the click, as indicated by the sound-level meter, was the peak equivalent SPL (dB pSPL) of the click stimulus.

For all experiments, only the first two wave components, designated by sequential Arabic numerals, were described by their amplitude and latency characteristics (Fig. 1). Positive evoked potential peaks were identified manually by cursor control, and associated latencies and amplitudes were automatically stored by the computer. Latencies to wave 1 and wave 2 were corrected for conduction delays between the sound source and the entrance of the ear canal of the animal (0.88 ms). The latency of the interwave interval (referred to as 1–2 interval) was calculated as the difference in latency from wave 1 to wave 2. ABR wave amplitudes were measured using baseline-to-peak for wave 1 and peak-to-peak (preceding trough) amplitude for wave 2.

III. EXPERIMENT 1: EFFECTS OF INTENSITY AND FREQUENCY ON THE ABR OF BUDGERIGARS

A. Introduction

For all animals tested to date, increasing stimulus level resulted in decreases in ABR response latency and increases in response amplitude. Frequency-specific ABRs were generated using brief tones. Experiment 1 evaluates the budgerigar ABR across stimulus intensities to determine whether the intensity dependence of evoked potentials in budgerigars is similar to those found in other animals.

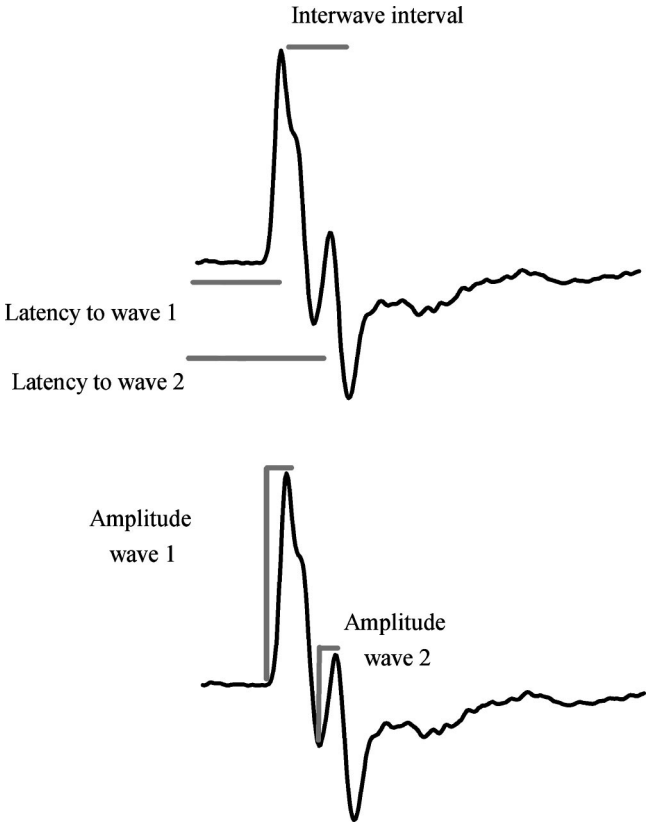


FIG. 1. Schematic showing how latency and amplitude measurements were taken for waves 1 and 2.

B. Stimuli

Ten adult budgerigars were presented with multiple intensity stimulus trains (e.g., Mitchell *et al.*, 1996, 1999) that varied in frequency and intensity. Each train consisted of nine single clicks or frequency tone bursts that increased in intensity and were presented at a rate of 4/s. The rectangular-pulse broadband clicks were 0.1 ms (100 μs) in duration with 25-ms interstimulus interval (ISI). Each individual tone burst was 5 ms in duration (1-ms rise/fall COS^2) with a 20-ms ISI. The tone bursts used were 500, 1000, 1500, 2000, 2860, 4000, and 5700 Hz. The tone bursts were played through the speaker and sampled at 40 kHz into the A/D module of the TDT rack. Spectra were generated using 1024-pt fast Fourier transform (FFT). Spectral analysis shows all second and third harmonics were at least 30 dB down from the peak of the frequency of interest, except for the first harmonic of the 500-Hz stimulus, which was 18 dB down.

For frequencies below 2000 and above 2860 Hz, stimulus intensities began at 45–50 dB and increased in 5-dB steps to a maximum of 85–90 dB. Tone bursts of 2000 and 2860 Hz, as well as the click, were initially presented at 35–40 dB and increased in 5-dB steps to 65–75 dB; the remaining intensities were in 10-dB steps to a maximum of

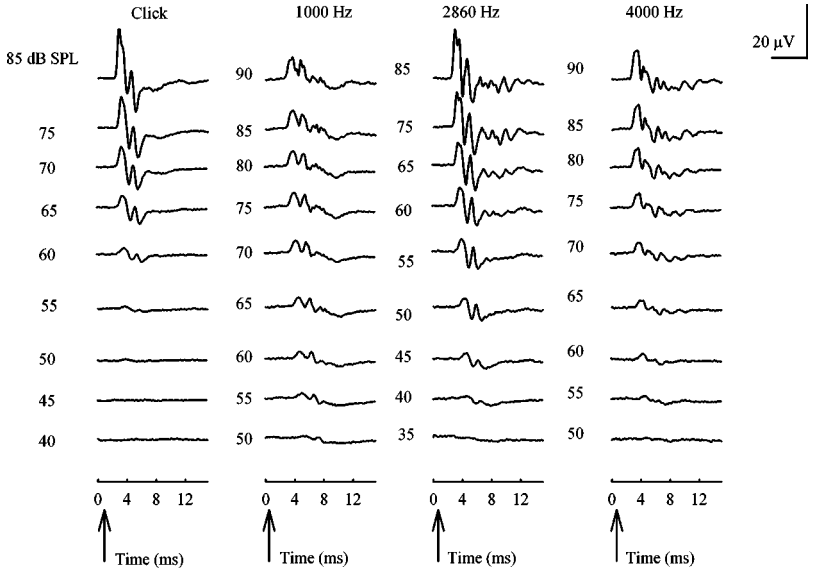


FIG. 2. Waveforms for a single bird for the click, 1000, 2860, and 4000 Hz as a function of SPL. Amplitude decreases and latency increases with decreasing SPL. Arrow denotes time at which stimulus reaches outer ear.

85 dB. Each ABR represents the average response of 1000 stimulus presentations (500 averages for each polarity/phase were added together to cancel the cochlear microphonic), sampled at 20 kHz for 235 ms following onset of the stimulus (allows for 25-ms recording time for each stimulus). The biological signal was amplified (×100 K) and notch filtered at 60 Hz with the DB4 digital biological amplifier during collection. The signal was bandpass filtered below 30 Hz and above 3000 Hz after collection using the BIOSIG program.

C. Analysis

ABR waveforms produced in response to high intensities were examined visually to determine which peaks would be used to measure latencies, amplitudes, and thresholds. A response was expected between 1 ms after the onset of the stimulus (travel time from speaker to the ear) and 10 ms because the response latency tends to increase at low SPLs. Using this time window, the wave components were described by their latency and amplitude characteristics.

ABR thresholds were estimated using several methods. First, thresholds were estimated using the visual detection method, i.e., the lowest intensity at which a response could be visually detected on the trace regardless of wave (Boettcher *et al.*, 1993) or 5 dB below the lowest intensity that elicited a measurable response (Walsh *et al.*, 1986). Second, a baseline-to-peak amplitude measure for wave 1 was taken for all intensities and stimuli (clicks and tone bursts). The amplitude-intensity functions generated from wave 1 amplitudes were used to estimate threshold (0-μV response) using linear regression. Last, thresholds were defined as the lowest stimulus intensity producing a response amplitude of 0.5 μV, which was at least 1 s.d. above the mean noise level (constant signal-to-noise ratio, S/N ratio of 0.5 μV). The average level of noise in the ABR (the average amplitude difference between the largest peak and trough in the 2–4-ms range) was determined for 37 traces obtained to inaudible stimuli. The noise level of these recordings was 0.28 μV ($\pm$0.16 μV s.d.). Thresholds estimated from these methods were compared with each other (one-way analysis of variance, ANOVA) and to thresholds estimated using behavioral procedures.

D. Results

Figure 2 shows typical ABR waveforms for an adult budgerigar click and three frequencies (1000, 2860, and 4000 Hz) as a function of intensity. Visual examination of the waveforms showed 2–3 prominent peaks that occurred within the first 5 ms after sound reaches the bird's external ear canal. Two major changes were evident for all waves. As the intensity of stimulation increased, latencies decreased and amplitudes increased.

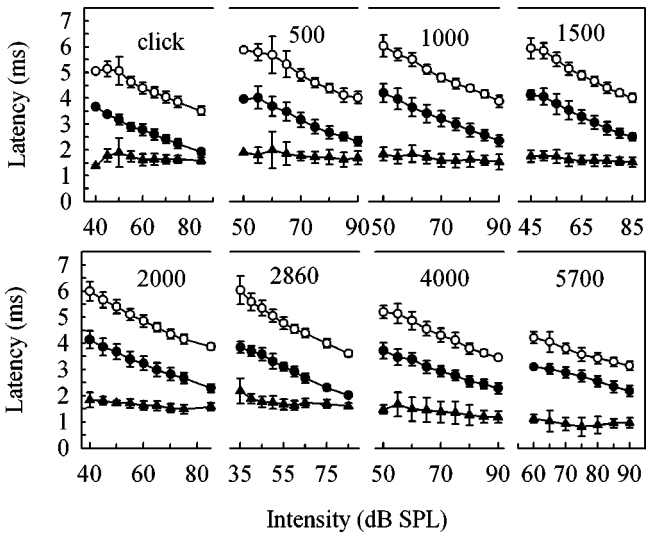


FIG. 3. Average latency as a function of intensity for the click and tone trains for ten adult budgerigars. The latency for wave 1 (closed circles) and wave 2 (open circles) increase as a function of decreasing level, while interwave interval (closed triangles) remains relatively constant. Error bars are standard deviations.

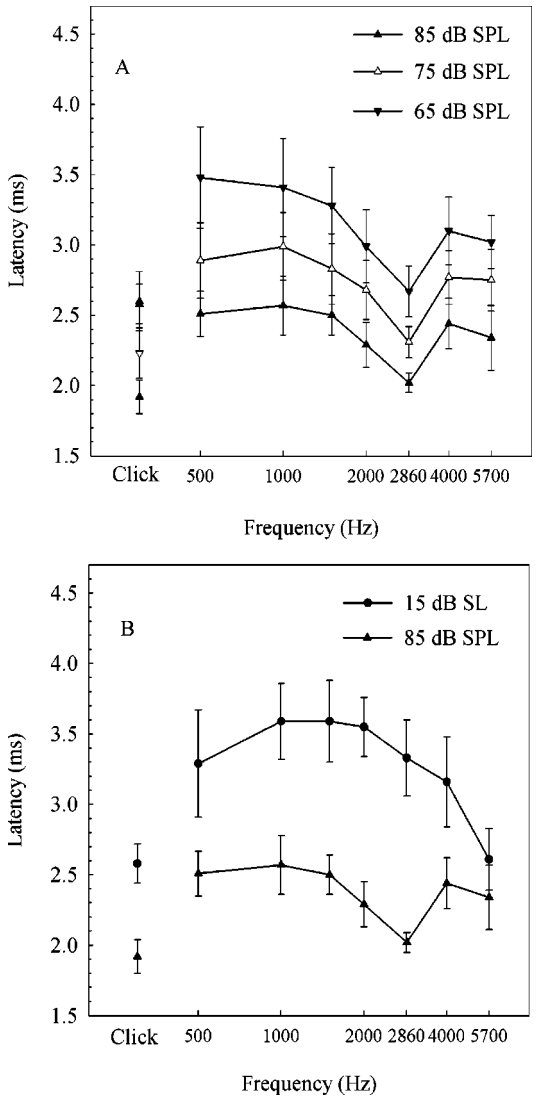


FIG. 4. (A) Average latency as a function of frequency for three constant SPL. There is not a consistent decrease in latency with increasing frequency. The lowest latencies are for 2860 Hz, which is the peak sensitivity of the behavioral audiogram as well as where most of the spectral energy of the contact call lies. (B) Average latency as a function of frequency using a constant SPL of 85 and 15 dB above threshold. Latency does decrease with increasing frequency if threshold is taken into account. Error bars are standard deviations.

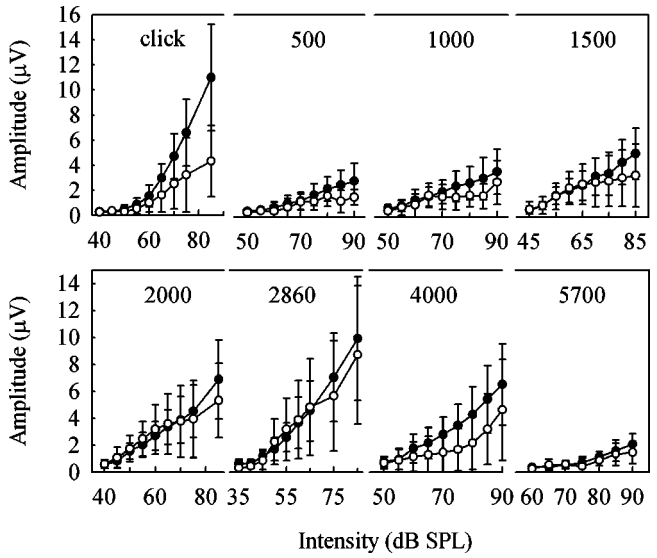


FIG. 5. Average amplitude as a function of intensity for the click and tone trains for ten adult budgerigars. Amplitudes for wave 1 (closed circles) and wave 2 (open circles) decrease rapidly with decreasing level. Error bars are standard deviations.

Latency decreased monotonically as a function of increasing intensity for waves 1 and 2, while the interwave interval remained relatively constant across intensities (Fig. 3). There was not a consistent decrease in latency as frequency increased for constant SPL [Fig. 4(a)]. The shortest latencies are for 2860 Hz, the region of best sensitivity in both the ABR and behavioral audiograms. When a constant ''sensation level'' of 15 dB (i.e., 15 dB above threshold) was used, higher frequencies resulted in shorter latencies [Fig. 4(b)]. Interestingly, the latency to wave 1 at 500 Hz was more like that to higher-frequency tone bursts (2860 and 4000 Hz) than lower-frequency tone bursts (1000 and 1500 Hz). Given the spectral characteristics of the 500-Hz stimuli, only the second harmonic (1000 Hz) could influence the response, and it was 18 dB below the peak. All other harmonics were at least 30 dB down from the peak; therefore, harmonics at higher frequencies cannot account for this latency discrepancy.

Figure 5 shows that the amplitude of waves 1 and 2 increased with increasing intensity. The slopes of the amplitude-intensity functions were frequency dependent, with clicks and midrange frequencies exhibiting steeper amplitude intensity functions than did low- or high-frequency tones. The highest average amplitudes were observed at frequencies in the budgerigar's range of best hearing (i.e., 2000–4000 Hz) and in response to the click stimulus. Amplitude increased monotonically for wave 1 across stimuli and did not show saturation at any SPL tested. Wave 2 amplitudes did not show a similar monotonic increase, but approached saturation at 500, 1000, 1500, and 5700 Hz.

The coefficient of variation (CV) was used to determine the relative amount of dispersion in ABR latency and amplitude measures across the adult budgerigars used in this study. Latency measures across subjects varied less than 10%, while amplitude measures varied between 40%–80% across subjects. Wave 2 amplitudes were more variable than wave 1 amplitudes (Fig. 6).

Figure 7 shows a comparison of the average ABR-derived audiogram based on the three ABR threshold methods and the behavioral audiogram for budgerigars. As shown in this figure, threshold estimates by various methods were relatively consistent for the click stimuli. But, threshold estimates for the tone pips varied by 10–13 dB across methods [$F(2,258)=13.18$; $p<0.001$]. The response criterion of 0.5 μV yielded significantly higher thresholds for the tone bursts than either the linear regression ($p<0.05$) or visual detection method ($p<0.05$). However, the response criterion and the visual detection methods more accurately paralleled the shape of the behavioral audiogram. The results from the regression method differed most from the U-shaped audiogram at the low-frequency end (500-2000 Hz). In general, all methods yielded threshold estimates that were 30–35 dB above behavioral estimates of hearing sensitivity.

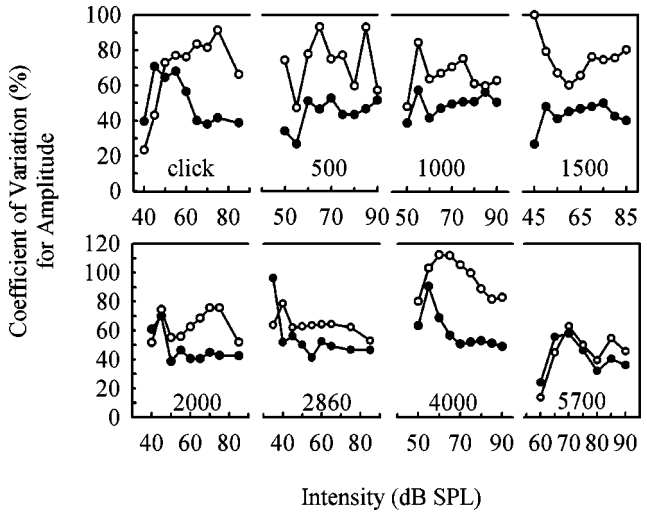


FIG. 6. Relative measure of amplitude dispersion across the ten adult budgerigars using the coefficient of variation (CV). Closed circles are for wave 1 measures. Open circles are for wave 2 measures. Overall, wave 2 amplitudes are more variable than wave 1 amplitudes.

IV. EXPERIMENT 2: THE RELATIONSHIP BETWEEN THE CM, CAP, AND ABR

A. Introduction

To date, there are no studies comparing the relationship between the cochlear microphonic (CM), compound action potential (CAP), and the ABR in the budgerigar. In this experiment, all three evoked potentials were recorded simultaneously and compared.

B. Stimuli

A single adult bird was presented with a multiple intensity stimulus train at 2860 Hz.

C. Analysis

The bird was anesthetized with IM injections of 20-mg/kg xylazine and 40-mg/kg ketamine. Additional doses of anesthetic (50% of initial dose) were supplemented as needed. The surgical procedures to gain access to the cochlea were described in detail for birds (Manley *et al.*, 1985). Briefly, feathers were removed from the head and around the external ear opening. An incision in the skin along the midline of the skull exposed the bone and it was cleared of connective tissue and dried. A small screw was cemented on the surface of the skull with dental cement to allow precise and stable placement of the head in a holding device and reproducible positioning of the ear canal opening in relation to the speaker.

For the CM and CAP potentials, the base of the cochlea was exposed by a dorsolateral approach. The core of a thin, Teflon-coated silver wire was exposed at the end and inserted through a tiny hole in the scala tympani to give direct electrical access to the perilymph. The Teflon insulation sealed the hole and prevented leaking of the perilymph. A subdermal electrode inserted into a neighboring neck muscle served as a reference. The electrodes for the ABR were placed as described in Sec. II, except the bird was 40 cm from the speaker. The stimulus presentation, data acquisition, equipment control, and data management were coordinated using the TDT setup described above, except where noted. Each ABR represents the average response of 600 stimulus presentations (300 averages for each), sampled at 20 kHz for 235 ms following onset of the stimulus (allows for 25-ms recording time for each stimulus). The biological signal was amplified ($\times$20 K) and notch filtered at 60 Hz with the DB4 digital biological amplifier during collection. The signal was bandpass filtered below 30 Hz and above 3000 Hz after collection using the BIOSIG program.

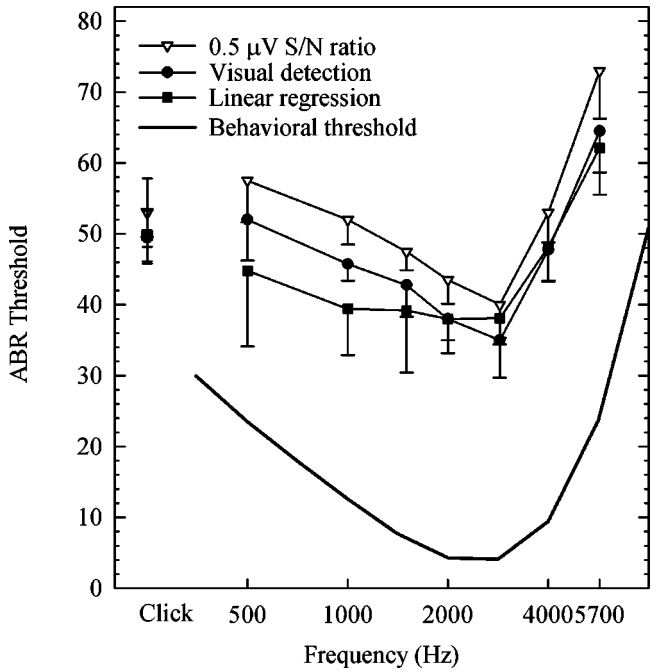


FIG. 7. Average ABR audiogram derived from the three methods: 0.5 μV S/N ratio (open triangle), visual detection method (closed circle), and the linear regression (closed square) compared to the behavioral audiogram (dark solid line). The curve estimated by all methods is 30 dB higher than the behavioral curve, but the visual detection and criterion methods more closely approximate the shape of the behavioral audiogram. Error bars are s.d.

The CAP response component was derived by adding the response traces obtained for the 90° and 270° stimuli (canceling the CM component) and scaling the resulting response amplitude by half. This derived neural response component was then subtracted from the response trace to the normal stimulus to derive the CM response. The ABR response component was derived by adding the two phases of the stimulus together (as in experiment 1).

D. Results

Simultaneous recording of CM, CAP, and ABR in response to a 2860-Hz tone played at 60 dB SPL for a single budgerigar are shown in Fig. 8. The onset of the CM occurs first (2.35 ms) and lasts the length of the stimulus. The first negative wave of the CAP occurs 1.55 ms after the onset of the CM and is followed closely (0.3 ms later) by the first positive deflection of the ABR. This comparison strongly suggests that wave 1 of the ABR is the auditory nerve.

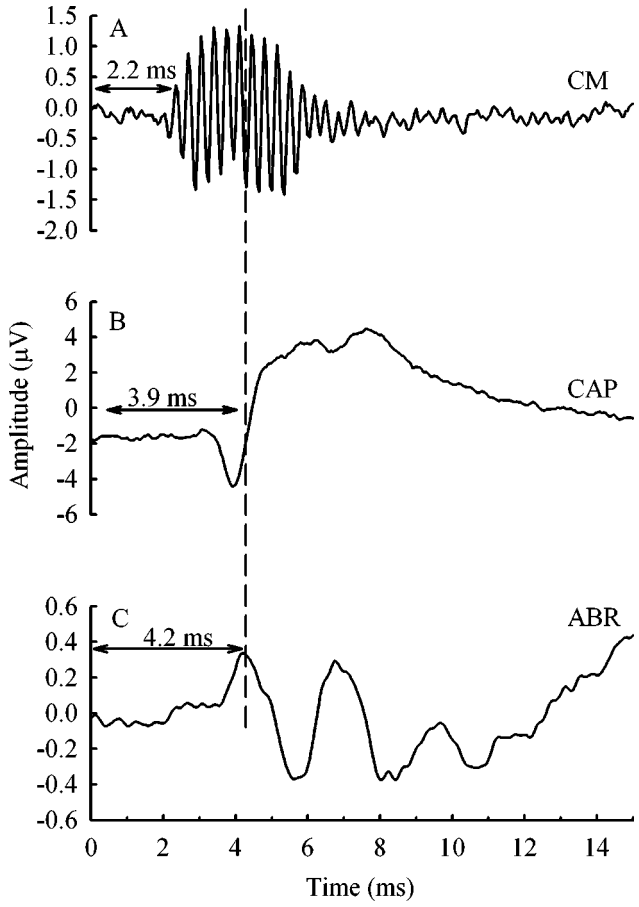


FIG. 8. Cochlear microphonic (A), compound action potential (B), and auditory brainstem response (C) recorded from a single budgerigar in response to a 2860-Hz tone played at 60 dB SPL. The negative deflection of the CAP corresponds well to the first positive deflection in the ABR.

V. EXPERIMENT 3: EFFECTS OF CLICK REPETITION RATE ON THE ABR OF BUDGERIGARS

A. Introduction

Studies of the ABR in humans and other mammals showed that higher click repetition rates caused increases in peak latency, decreases in peak amplitude, and alterations of wave morphology (e.g., Burkard and Voigt, 1989; Donaldson and Rubel, 1990; Jewett and Romano, 1972; and as reviewed in Hall, 1992). These changes may be a function of neural fatigue and adaptation (Hall, 1992). Knowing the baseline response of ABR latency and amplitudes to increasing presentation rate may allow investigators to use increased stimulation rate as a tool for detecting auditory neuropathologies.

The changes associated with increased presentation rate have been well studied in mammals. With the exception of the chick (Burkard *et al.*, 1994), few data exist for birds. However, like mammals, the chick shows similar rate-dependent changes. Experiment 2 extends the effects of click repetition rate on ABR latency and amplitude to the budgerigar.

The same ten birds were used in this experiment. All equipment and procedures were the same as in experiment 1, except where noted below.

B. Stimuli

Click level was held constant at 100 dB pSPL. Short duration, broadband clicks (0.1 ms) were presented at five rates: 5, 10, 30, 60, and 90 per second (Hz). Each ABR represents the average response of 1000 stimulus presentations (500 averages for each polarity were added together to cancel the cochlear microphonic), sampled at 20 kHz for 10 ms following onset of the stimulus. As with experiment 1, the biological signal was amplified (×100 K) and notch filtered at 60 Hz with the DB4 digital biological amplifier during collection. The signal was bandpass filtered below 30 Hz and above 3000 Hz after collection using the BIOSIG program.

C. Analysis

Latency and amplitude measures were calculated for waves 1 and 2 for all repetition rates as described earlier. As a further comparison of the effects of increasing rate on the amplitude of the waves, the ratio of wave 1 to wave 2 was calculated.

D. Results

Figure 9 shows a budgerigar's ABR waveforms to all click rates used in this study. As click repetition rate increased, latency increased and amplitude decreased to all waves. The waves also broadened, especially at rates higher than 10 Hz. Figures 10(a)–(c) show that with increasing repetition rate, the average latency to individual ABR waves increased. The latency of wave 1, wave 2, and the 1–2 interval were evaluated with a two-way ANOVA. Table I summarizes the results. Overall, as the repetition rate increased from 5 to 90 Hz, latency increased for wave 1 and wave 2 with a greater increase for wave 2, as shown by the increase in the wave 1–2 interval [Fig. 10(c)]. Pairwise comparisons using Student's t tests showed that latencies to waves 1 and 2 for repetition rates 30 Hz and above were significantly longer than for lower rates (10 Hz and below). The 1–2 interval increased significantly for repetition rates above 30 Hz.

The effects of increasing repetition rate from 5 to 90 Hz on amplitude are summarized in Figs. 10(d)–(f). Absolute amplitude decreased dramatically for both waves. The mean wave 1 to 2 amplitude ratio [Fig. 10(f)] shows that the amplitude of wave 1 was larger than the amplitude of wave 2, resulting in mean amplitude ratios greater than 1, but these ratios remained relatively constant across repetition rate. The amplitude of wave 1, wave 2, and the amplitude ratio of wave 1 to 2 were evaluated with a two-way ANOVA, and the results are summarized in Table I. Pairwise comparisons using Student's t tests showed that amplitudes of waves 1 and 2 were significantly larger for lower repetition rates (10 Hz and below) than for higher rates (30 Hz and above). The ratio of wave 1 to wave 2 amplitudes increased slightly, but this increase was not significant ($p > 0.05$), suggesting that both waves were equally affected by increasing presentation rate.

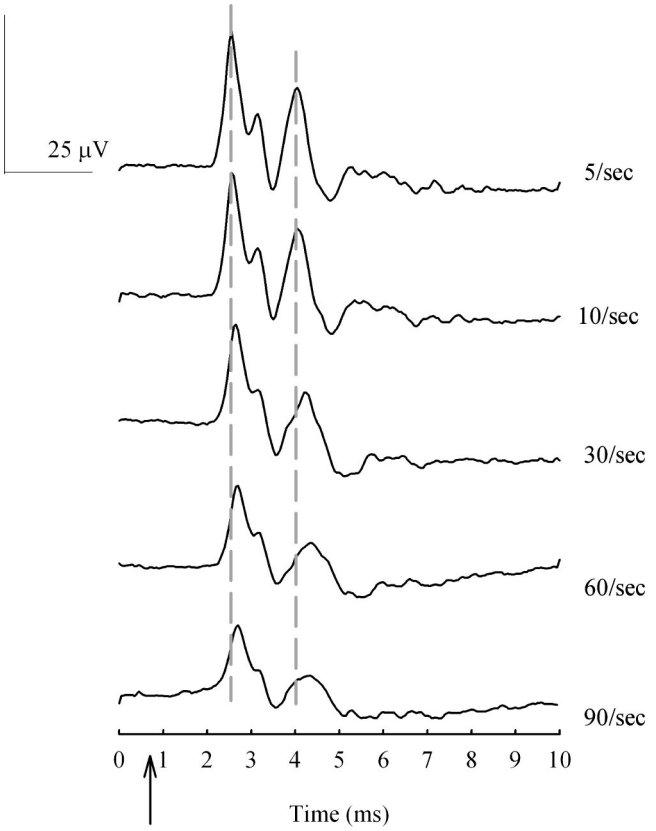


FIG. 9. Typical ABR trace for each repetition rate for an adult budgerigar. Increasing rate causes increased latencies and decreased amplitudes, with the biggest changes occurring above 10 Hz. Arrow denotes time at which stimulus reaches outer ear.

VI. DISCUSSION

A. Surface ABRs as an approach to the measurement of hearing in birds

The main purpose of this experiment was to measure hearing sensitivity in an adult bird using the ABR. Previous ABR studies have shown that while frequency-dependent ABRs are good predictors of audiogram shape, they are not necessarily predictive of absolute auditory sensitivity (e.g., Borg and Engström, 1983; Stapells and Oates, 1997; Wenstrup, 1984). ABR audiograms generated for the budgerigar were similar in shape to behavioral audiograms, showing a similar range of best hearing and a peak sensitivity at 2860 Hz (see the review in Dooling *et al.*, 2000). While the ABR audiogram was 30 dB less sensitive than the behavioral audiogram, much of this difference can be attributed to auditory temporal integration (Dooling *et al.*, 2000).

Figure 11 shows ABR audiograms for several species of vertebrates. Each audiogram reflects the bandwidth and shape of behavioral audiograms in the tested species, and all show a range of best sensitivity in the 2000–4000-Hz region. ABR thresholds from the current study are similar to those reported from Woolley and Rubel (1999) and Woolley *et al.* (2001) on Bengalese finches. Interestingly, thresholds for birds are higher than those of the mammalian species shown. Studies of tuning curve thresholds in birds show a larger spread of neural threshold at a given CF (review in Gleich and Manley, 2000; Manley *et al.*, 1985). Since the ABR is a synchronized response, detecting a response near threshold requires synchronous activation of a sufficient number of fibers—on average this may be easier in mammals, where the proportion of fibers activated at threshold is likely greater

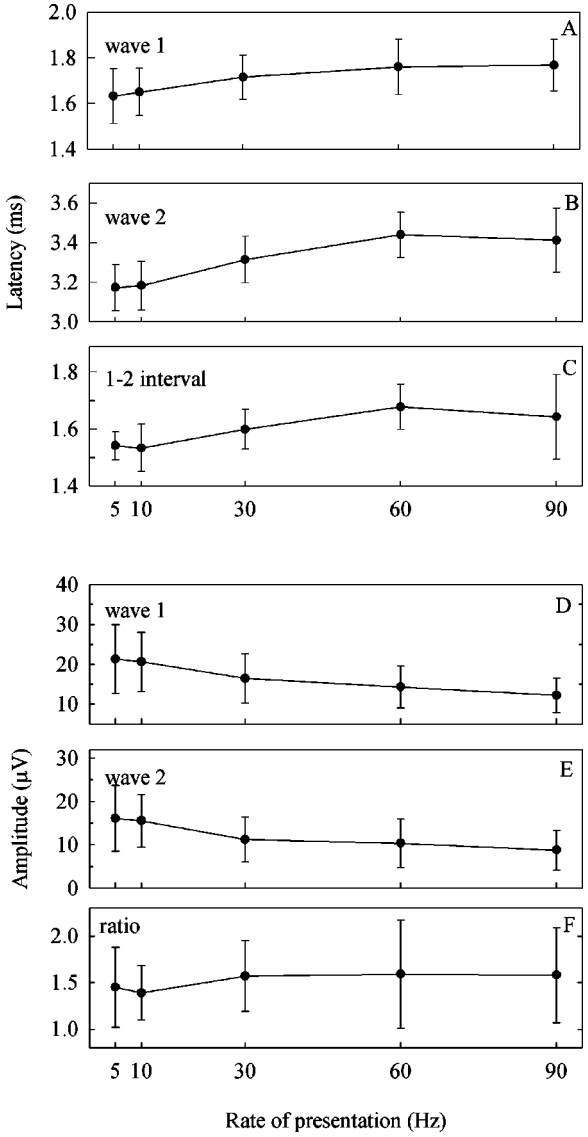


FIG. 10. Average latency of wave 1 (A), wave 2 (B), and the inter wave interval (C) plotted as a function of repetition rate for ten adult budgerigars. Latencies for all waves increased as rate increased, with greater increases seen in later waves (increase in interval). Average amplitude of wave 1 (D), wave 2 (E), and the ratio of wave 1 to 2 (F) plotted as a function of repetition rate for ten adult budgerigars. Amplitudes of all waves decreased as a function of increased rate. This decrease was similar for both waves 1 and 2. Error bars are standard deviations.

TABLE I. Results of ANOVA for latency and amplitude as a function of repetition rate.

Dependent variable	*N*	*F* ratio	Probability
Wave 1 latency	10	19.07	$p<0.0001$
Wave 2 latency	10	24.08	0.001
Wave 1-2 Interval	10	7.40	0.0002
Wave 1 amplitude	10	24.73	0.0001
Wave 2 amplitude	10	13.36	0.0001
Wave 1/2 Ratio	10	1.13	NS

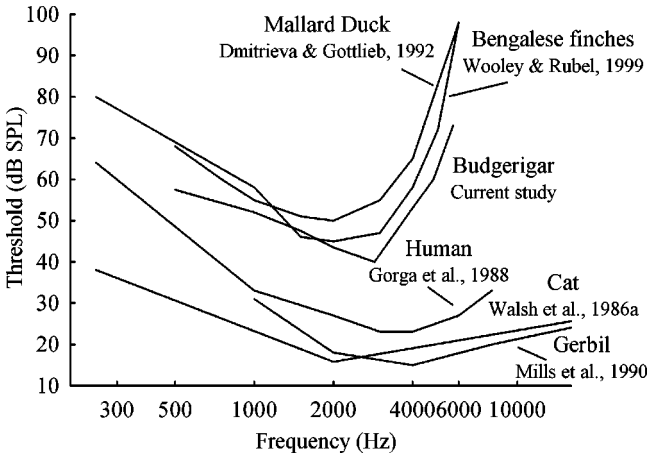


FIG. 11. ABR audiograms for several species of birds and mammals. All of the bird audiograms have similar shapes and thresholds. The bird audiograms are up to 30 dB higher than those of the mammals.

than it is in birds. Also, in absolute terms, the signal may be improved because the absolute number of fibers is higher in mammals (e.g., cats 50 000, Gacek and Rasmussen, 1961; see the review in Ryugo, 1992) than in birds (e.g., budgerigar 9800, Manley *et al.*, 1993).

B. Effects of intensity on latency and amplitude for clicks and tone bursts

Typically, our budgerigar ABR waveform showed two measurable peaks that occurred within the first 4 ms after stimulation. This is similar to what was seen by Umemoto *et al.* (1993). A third peak was evident in some but not all traces, and was not examined further. In general, latency increased with decreasing SPL, and these increases occurred at the same rate for waves 1 and 2, such that the interwave interval remained constant across SPL. Similar results have been found in mammals (e.g., Burkard and Voigt, 1989; Burkard *et al.*, 1996a; Huang and Buchwald, 1978). There was not a consistent dependency of latency on frequency at constant SPL levels for budgerigars as has been shown in other studies (e.g., Gorga *et al.*, 1988). In fact, latencies were shortest for 2860 Hz, which is where the behavioral audiogram shows its most sensitive point and where most of the spectral energy of the budgerigar contact call is located. The amplitude of all peaks also decreased with decreasing SPL, but not at the same rate for wave 1 and 2. Overall, amplitude measures are more variable across subjects (relative dispersion between 40%–80%), while latency measures show little variation (less than 10%) across subjects (see Fig. 6). This is a typical finding in ABR studies.

There are no anatomical data to indicate why the largest amplitudes and shortest latencies were reported for the click and frequencies between 2000 and 4000 Hz, the region of greatest sensitivity in the budgerigar's audiogram. The basilar papilla contains about 5400 hair cells and is innervated by an estimated 9800 fibers (Manley *et al.*, 1993). Morphological features of the budgerigar papilla, such as hair-cell dimension, change gradually along the length and width, which does not provide an obvious basis for a dramatic change in function along the papilla. Likewise, the bases for other known psychophysical specializations (such as increased sensitivity at 2860 Hz in the critical ratio; Dooling *et al.*, 2000) are also not evident at the level of the hair cells. As of yet, there are no detailed reports of the innervation pattern or innervation density of hair cells of different regions in the budgerigar's basilar papilla as compared to other species. Köppl *et al.* (2000) note that the budgerigar's fiber number/hair-cell number ratio is similar to that of the barn owl and suggest looking for regional increases in cochlear innervation densities in this species.

C. Effects of click presentation rate on latency and amplitudes

With increasing click repetition rate, human, mammalian, and chick ABRs show increases in peak latencies and decreases in peak amplitudes (Burkard *et al.*, 1994; Burkard and Voigt, 1989; Burkard *et al.*, 1996a, b; Donaldson and Rubel, 1990; Lasky, 1997). The same effect was observed in budgerigars (Fig. 9). The decrease in amplitude occurred at the same rate for both waves [Fig. 9(f)].

As rate increased, there was a greater latency change for wave 2, which resulted in increasing interwave intervals with increasing repetition rate [Fig. 9(c)]. This has also been reported in humans (e.g., Burkard and Hecox, 1987), gerbils (Burkard and Voigt, 1989), cats (Burkard *et al.*, 1996a), and chicks (Burkard *et al.*, 1994). For humans, as rate increased from 10 to 90 Hz there was a mean I–V interval increase of 0.31 ms, from 3.71 to 4.02 ms (Burkard and Hecox, 1987). For gerbils, there was a 0.244-ms increase from 15 to 90 Hz for the i–v interval (Burkard and Voigt, 1989) and a 0.3-ms increase in the i–iv interval as rate increased from 10 to 100 Hz in cats (Burkard *et al.*, 1996a). As presentation rate increased from 10–90 Hz, the interval between wave 2 and 3 increased by approximately 0.10 ms in chicks and by 0.11 ms in budgerigars (from wave 1 to 2). As in other species tested, there was a greater latency change for later ABR peaks in the adult budgerigar. The constant ratio of amplitudes suggests that wave 1 and 2 were affected in the same way across the range of repetition rates used in the current experiment. Overall, these data suggest that the neural generators responsible for wave 1 and wave 2 in the ABR of budgerigars show similar rate-dependent ABR changes as mammals and other birds that have been tested (i.e., adaptation increases with increasing presentation rate, especially above 10 Hz).

D. Neural correlates of ABR waves in budgerigars

While, to our knowledge, there have been no studies of ABR generators in budgerigars, it is possible to suggest possible neural generators for waves 1 and 2 based on latency data from other animals (e.g., chickens and cats). Wave 1 in animals is consistently attributed to auditory nerve. Wave 1 latencies were 1.3–1.4 ms for 103-dB SPL clicks (presented at 3.3 Hz) in chickens (Katayama, 1985) and 1.65 ms for 90-dB SPL clicks (presented at 5 Hz) in cats (Burkard *et al.*, 1996a). For the adult budgerigar, 100-dB SPL clicks presented at 5 Hz resulted in wave 1 latencies that are 1.6–1.7 ms. These latencies for budgerigars are similar to what has

been found for chickens and cats. Experiment 2 in the current study showed that the first negative deflection of the CAP corresponded well to the first positive deflection of the ABR waveform suggesting that the generators are the same. Therefore, we suggest that wave 1 of the budgerigar ABR is generated by the auditory nerve.

The source of wave 2 is somewhat more difficult to identify. Visual inspection of the chicken waveform suggests that Katayama's N4 might correspond to budgerigar wave 2. Nucleus laminaris is the suggested generator of N4 (latency 2.2–2.4 ms), and N4 in chickens has been suggested as equivalent to the third peak in the ABR of most mammals (Katayama, 1985; Burkard *et al.*, 1996a). In budgerigars, wave 2 latencies are 3.2–3.3 ms for a 100-dB click. Based on similarities in latency across these studies, wave 2 in the budgerigar may be generated by nucleus laminaris.

VII. CONCLUSIONS

Audiograms based on scalp-recorded ABRs of adult budgerigars were similar in shape to behavioral audiograms (Fig. 7). As predicted from the lack of temporal integration in ABRs and the number of fibers needed for an ABR response, the ABR audiogram was less sensitive and underestimated behavioral thresholds by about 30 dB. However, the region of greatest sensitivity in the ABR audiogram was in the 2000–4000-Hz region, which matches the behavioral audiogram. This is also the region of peak energy in the adult contact calls (Dooling, 1986).

The effects of intensity on latency and amplitude of the ABR waves were similar to other animals studied to date. As intensity increased, latency to all waves decreased and amplitude increased. Analysis of the coefficient of variation showed that latency measures in this sample were more consistent than amplitude measures across subjects (Fig. 6). Last, increasing repetition rate in adult budgerigars resulted in longer latencies and decreased amplitudes. The latency changes were greater for wave 2 than for wave 1 (Fig. 10), indicating that generators of later waves are more affected by neural adaptation than wave 1. Wave amplitudes were equally affected by increasing presentation rate.

Overall, this study shows that the ABR is useful for assessing peripheral auditory system function and for estimating hearing thresholds in adult budgerigars. These procedures open the door to other investigations such as the recovery of hearing following hair-cell regeneration (Woolley and Rubel, 1999; Woolley *et al.*, 2001), the phase response of the basilar papilla (see for example, Dau *et al.*, 2000), and the development of hearing (Brittan-Powell, 2002). In fact, the noninvasive aspect of this procedure allowing repeated testing of the same individual makes it an ideal tool for assessing hearing during development. Hearing development using the ABR has been done in several mammals that are considered altricial, but not yet in altricial birds. Because budgerigars use auditory feedback to learn and modify their calls throughout life, knowing how the auditory system develops and what and when the animal hears provides insights into the role that hearing plays in the development of different types of vocalizations.

ACKNOWLEDGMENTS

The authors would like to thank C. Carr, C. Moss, D. Yager, E. Walsh, K. Köppl, M. Leek, F. Kubke, M. Dent, D. Higgs, and J. Triblehorn for comments on earlier drafts and G. Manley and an anonymous reviewer for improvements made to this work. This work was supported in part by training grant No. DC-00046 from the National Institute of Deafness and Communicative Disorders of the National Institutes of Health to E.F.B.P. and National Institutes of Health Grant No. DC00198 to R.J.D. The work described here was in partial fulfillment of the requirement of a doctor of philosophy to E.F.B.P. (University of Maryland, College Park, 2002).

Aleksandrov, L. I., and Dmitrieva, L. P. (**1992**). ''Development of auditory sensitivity of altricial birds: absolute thresholds of the generation of evoked potentials,'' Neurosci. Behav. Physiol. **22**, 132–137.

Boettcher, F. A., Mills, J. H., Norton, B. L., and Schmiedt, R. A. (**1993**). ''Age-related changes in auditory evoked potentials of gerbils. II. Response latencies,'' Hear. Res. **71**, 145–156.

Borg, E., and Engström, B. (**1983**). ''Hearing thresholds in the rabbit,'' Acta Oto-Laryngol. **95**, 19–26.

Brittan-Powell, E. F. (**2002**). ''Development of auditory sensitivity in budgerigars,'' Department of Psychology, College Park, University of Maryland.

Burkard, R., and Hecox, K. E. (**1987**). ''The effect of broadband noise on the human brain-stem auditory evoked response. III. Anatomic locus,'' J. Acoust. Soc. Am. **81**, 1050–1063.

Burkard, R., Jones, S., and Jones, T. (**1994**). ''Conventional and cross-correlation brain-stem auditory evoked responses in the white leghorn chick: Rate manipulations,'' J. Acoust. Soc. Am. **95**, 2136–2144.

Burkard, R., McGee, J., and Walsh, E. J. (**1996a**). ''Effects of stimulus rate on feline brain-stem auditory evoked response during development. I. Peak latencies,'' J. Acoust. Soc. Am. **100**, 978–990.

Burkard, R., McGee, J. and Walsh, E. J. (**1996b**). ''Effects of stimulus rate on feline brain-stem auditory evoked response during development. II. Peak amplitudes,'' J. Acoust. Soc. Am. **100**, 991–1002.

Burkard, R., and Voigt, H. F. (**1989**). ''Stimulus dependencies of the gerbil brain-stem auditory-evoked response (BAER). I. Effects of click level, rate, and polarity,'' J. Acoust. Soc. Am. **85**, 2514–2525.

Corwin, J. T., Bullock, T. H., and Schweitzer, J. (**1982**). ''The auditory brain stem response in five vertebrate classes,'' Electroencephalogr. Clin. Neurophysiol. **54**, 629–641.

Dau, T., Wegner, O., Mellert, V., and Kollmeir, B. (**2000**). ''Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion,'' J. Acoust. Soc. Am. **107**, 1530–1540.

Dmitrieva, L. P., and Gottlieb, G. (**1992**). ''Development of brainstem auditory pathway in mallard duck embryos and hatchlings,'' J. Comp. Physiol., A **171**, 665–671.

Donaldson, G. S., and Rubel, E. W. (**1990**). ''Effects of stimulus repetition rate on ABR threshold, amplitude, and latency in neonatal and adult Mongolian gerbils,'' Electroencephalogr. Clin. Neurophysiol. **77**, 458–470.

Dooling, R. J. (**1986**). ''Perception of vocal signals by Budgerigars (*Melopsittacus undulatus*),'' Exp. Biol. **45**, 195–218.

Dooling, R. J., Lohr, B., and Dent, M. L. (**2000**). ''Hearing in birds and reptiles,'' in *Comparative Hearing: Birds and Reptiles*, edited by R. J. Dooling, A. N. Popper, and R. R. Fay (Springer-Verlag, NY), pp. 308–359.

Farabaugh, S. M., and Dooling, R. J. (**1996**). ''Ecology and evolution of acoustic communication in parrots: laboratory and field studies of budgerigars, Melopsittacus undulatus,'' in *Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca, NY), pp. 97–118.

Gacek, R. R., and Rasmussen, G. L. (**1961**). ''Fiber analysis of the statoacoustic nerve of guinea pig, cat and monkey,'' Anat. Rec. **139**, 455–463.

Gleich, O., and Manley, G. A. (**2000**). ''The hearing organ of birds and crocodilia,'' in *Comparative Hearing: Birds and Reptiles*, edited by R. J. Dooling, A. N. Popper, and R. R. Fay (Springer-Verlag, NY), pp. 70–138.

Gorga, M., Kaminski, J., Beauchaine, K., and Jesteadt, W. (**1988**). ''Auditory brain response to tone burst in normal hearing subjects,'' J. Speech Hear. Res. **31**, 89–97.
Hall, J. (**1992**). *Handbook of Auditory Evoked Responses* (Allyn and Bacon, Boston, MA).
Heaton, J. T., and Brauth, S. E. (**1999**). ''Effects of deafening on the development of nestling and juvenile vocalizations in budgerigars (*Melopsittacus undulatus*),'' J. Comp. Psychol. **113**, 314–320.
Higgs, D. M., Soares, D., Brittan-Powell, E. F., Souza, M. J., Carr, C. E., Dooling, R. J., and Popper, A. N. (**2002**) ''Amphibious auditory responses of the American Alligator (*Alligator mississipiensis*),'' J. Comp. Physiol., A **188**, 217–223.
Hile, A. G., and Striedter, G. F. (**2000**). ''Call convergence within groups of female budgerigars (*Melopsittacus undulatus*),'' Ethology **106**, 1105–1114.
Huang, C., and Buchwald, J. (**1978**). ''Factors that affect the amplitudes and latencies of vertex short latency acoustic responses in the cat,'' Electroencephalogr. Clin. Neurophysiol. **44**, 179–186.
Jewett, D., and Romano, M. (**1972**). ''Neonatal development of the auditory system: Potentials averaged from the scalp of rat and cat,'' Brain Res. **36**, 101–115.
Jewett, D., Romano, M., and Williston, J. (**1970**). ''Human auditory evoked potentials: Possible brainstem components detected on the scalp,'' Science **167**, 1517–1518.
Katayama, A. (**1985**). ''Postnatal development of auditory function in the chicken revealed by auditory brain-stem responses (ABRs),'' Electroencephalogr. Clin. Neurophysiol. **62**, 388–398.
Kenyon, T. N., Ladich, F., and Yan, H. Y. (**1998**). ''A comparative study of hearing ability in fishes: the auditory brainstem response approach,'' J. Comp. Physiol., A **182**, 307–318.
Köppl, C., Wegscheider, A., Gleich, O., and Manley, G. (**2000**). ''A qualitative study of cochlear afferent axons in birds,'' Hear. Res. **139**, 123–143.
Lasky, R. E. (**1997**). ''Rate and adaptation effects on the auditory evoked brainstem response in human newborns and adults,'' Hear. Res. **111**, 165–176.
Liu, G. B., and Mark, R. F. (**2001**). ''Functional development of the inferior colliculus (IC) and its relationship with the auditory brainstem response (ABR) in the tammar wallaby (*Macropus eugenii*),'' Hear. Res. **157**, 112–123.
Manley, G. A., Gleich, O., Leppelsack, H. J., and Oeckinghaus, H. (**1985**). ''Activity patterns of cochlear ganglion neurons in the starling,'' J. Comp. Physiol., A **157**, 161–181.
Manley, G. A., Schabedissen, G., and Gleich, O. (**1993**). ''Morphology of the basilar papilla of the budgerigar, *Melopsittacus undulatus*,'' J. Morphol. **218**, 153–165.
McFadden, S. L., Walsh, E. J., and McGee, J. (**1996**). ''Onset and development of auditory brainstem responses in the Mongolian gerbil (*Meriones unguiculatus*),'' Hear. Res. **100**, 68–79.
Mills, J. J., Schmiedt, R. A., and Kulish, L. F. (**1990**). ''Age-related changes in auditory potentials of Mongolian gerbil,'' Hear. Res. **46**, 210–210.
Mitchell, C., Kempton, J., Creedon, T., and Trune, D. (**1996**). ''Rapid acquisition of auditory brain stem responses with multiple frequency and intensity tone bursts,'' Hear. Res. **99**, 38–46.
Mitchell, C., Kempton, J., Creedon, T., and Trune, D. (**1999**). ''The Use of a 56-stimulus train for the rapid acquisition of auditory brainstem responses,'' Audiol. Neuro-Otol. **4**, 80–87.
Moiseff, A., Haresign, T., and Wang, J. (**1996**). ''Sound localization from binaural cues by the barn owl auditory system,'' in *Neuroethological Studies of Cognitive and Perceptual Processes*, edited by C. F. Moss and S. J. Shettleworth (Westview Press, Boulder, CO), pp. 305–323.
Ryugo, D. K. (**1992**). ''The auditory nerve: Peripheral innervation, cell body morphology, and central projections,'' in *The Mammalian Auditory Pathway: Neuroanatomy*, edited by D. B. Webster, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 23–65.
Saunders, J. C., Coles, R. B., and Gates, G. R. (**1973**). ''The development of auditory evoked responses in the cochlea and cochlear nuclei of the chick,'' Brain Res. **63**, 59–74.
Stapells, D. R., and Oates, P. (**1997**). ''Estimation of pure-tone audiogram by the auditory brainstem response: A review,'' Audiol. Neuro-Otol. **2**, 257–280.
Umemoto, M., Sakagami, M., Ashida, K., Fukazawa, K., Matsunaga, T., Senda, T., and Fujita, H. (**1993**). ''The ultrastructure of the basilar papilla of the budgerigar's inner ear,'' Acta Oto-Laryngol. **501**, 66–71.
Walsh, E. J., McGee, J., and Javel, E. (**1986**). ''Development of auditory-evoked potentials in the cat. I. Onset of response and development of sensitivity,'' J. Acoust. Soc. Am. **79**, 712–724.
Wenstrup, J. J. (**1984**). ''Auditory sensitivity in the fish-catching bat, *Noctilio leporinus*,'' J. Comp. Physiol., A **155**, 91–101.
Woolley, S. M. N., and Rubel, E. (**1999**). ''High-frequency auditory feedback is not required for adult song maintenance in Bengalese Finches,'' J. Neurosci. **19**, 358–371.
Woolley, S. M. N., Wissman, A. M., and Rubel, E. (**2001**). ''Hair cell regeneration and recovery of auditory thresholds following aminoglycoside ototoxicity in Bengalese Finches,'' Hear. Res. **153**, 181–195.

Limitations on rate discrimination

Robert P. Carlyon and John M. Deeks
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, England

(Received 5 June 2001; revised 29 May 2002; accepted 31 May 2002)

We investigated the limits of temporal pitch processing under conditions where the place and rate of stimulation on the basilar membrane were independent. Stimuli were harmonic complexes passed through a fixed bandpass filter and resembled filtered pulse trains. The task was to detect a difference in $F0$. When the harmonics were filtered between 3900–5400 Hz, presented monaurally, and summed in sine phase, subjects could perform the task at all $F0$s studied. However, when the pulse rate was doubled by summing components in alternating phase, thresholds increased with increasing $F0$ until the task was impossible at $F0=300$ Hz (pulse rate=600 pps). Thresholds improved again at higher $F0$s, presumably because some harmonics became resolved. The $F0$ at which this breakdown occurred decreased when the complexes were filtered into a lower frequency region, and increased when they were filtered into a higher region. In the highest region tested (7800–10 800 Hz), all listeners could detect an increase of less than about 20% *re*: a pulse rate of 600 pps for alternating-phase complexes. Presenting a copy of the standard (lower-$F0$) stimulus to the contralateral ear during all intervals of a forced-choice trial improved performance markedly under conditions where monaural rate discrimination was very poor. This showed that temporal information is present in the auditory nerve that is unavailable to the temporal pitch mechanism, but which is accessible when a binaural cue is available. The results are compared to the inability of most cochlear implantees to detect increases in the rate of electrical pulse trains above about 300 pps. It is concluded that this inability is unlikely to result entirely from a central pitch limitation, because, with analogous acoustic stimulation, normal listeners can perform the task at substantially higher rates. [DOI: 10.1121/1.1496766]

PACS numbers: 43.66.Fe, 43.66.Hg, 43.66.Mk, 43.66.Ts [LRB]

I. INTRODUCTION

The experiments described here, although performed using acoustic stimuli, were inspired by an apparent paradox in the literature concerning the auditory temporal coding of electrical stimulation. Although physiological measurements of auditory-nerve responses to trains of electric impulses can show very fine phase locking (van den Honert and Stypulkowski, 1987; Shepherd and Javel, 1997), human discrimination of electrical pulse rate is, with the exception of a very few subjects, very poor once the baseline rate exceeds about 300 pps (Shannon, 1983a; Tong and Clark, 1985; Townshend *et al.*, 1987; McKay *et al.*, 2000). A similar deterioration occurs for sinusoidal electrical stimulation (Shannon, 1983a). Rate discrimination by implantees also differs markedly from the exquisite ability of normal listeners to detect changes of less than 1% in the frequencies of pure tones, at least at low to moderate frequencies. Although frequency shifts in pure tones produce changes in place of excitation, there is mounting evidence that, for frequencies up to about 2000 Hz, the detection of these changes is mediated by phase-locking cues. For example, the changes produced in the excitation pattern cannot account for the small frequency DLs observed in normal listeners (Moore, 1973), or the way in which they vary with baseline frequency (Sek and Moore, 1995), hearing impairment (Tyler *et al.*, 1983; Moore and Peters, 1992) or duration (Micheyl *et al.*, 1998). Hence, it seems that the ability of cochlear implant listeners to detect rate or frequency changes is inferior not only to that predicted from physiological data, but also to that of normal listeners to use phase-locking information.

A number of authors have noted that the limitations on temporal processing by cochlear implantees resembles that shown by normal listeners required to encode the rate of amplitude modulation (AM) applied to a noise carrier (Shannon, 1983b; Blamey *et al.*, 1984; Shannon, 1992; McKay *et al.*, 2000). In each case, discrimination thresholds are higher than for acoustic pure-tone frequency discrimination, and deteriorate at high rates. It is possible that either or both of these two features share a common basis for cochlear implant users and normal listeners. For example, Loeb *et al.* (1983) have proposed a model of pitch processing in which it is important for phase locking to occur in neurons innervating the appropriate places on the basilar membrane. The general idea that the frequency of phase locking and the place of excitation need to be consistent is also embodied in the ''average localized synchronized rate (ALSR)'' measure proposed by Sachs *et al.* (1983). According to this general idea, there is a mechanism that can accurately process the frequencies of acoustic pure tones, and do so up to high overall frequencies, but which could not act either on electrical pulse trains or amplitude-modulated noise.

An alternative explanation is that the patterns of performance observed with modulated noise in normal listeners and electric pulse trains in implantees, although similar, have quite different causes. Although physiological experiments do reveal highly accurate phase locking to electrical stimu-

lation, two observations suggest caution in generalizing this finding to hearing by human users of cochlear implants. First, physiological measures are usually obtained at current levels necessary to record a response in a single neuron, and these levels may be substantially higher than those used clinically, where many nerves are excited by each electrode channel. Second, the electrical recordings are usually obtained on very recently deafened animals; Shepherd and Javel (1997) have shown that phase locking deteriorates with increasing length of deafness, presumably because of the degeneration of peripheral processes. These two factors suggest that phase locking to electrical stimulation in human implantees may well be worse than is usually measured in animal recordings, and could provide a basis for their poor rate-discrimination abilities, especially at high pulse rates.

Conversely, the deterioration in rate discrimination of AM noise in normal listeners may result from two other factors. One of these stems from the stochastic properties of the noise carriers, which will result in some modulations at the output of the auditory filter bank. The rate of these ''inherent'' modulations (determined by auditory filter bandwidth) will typically be much higher than the slowest rates of the AM typically imposed on the stimuli (Formby and Muir, 1988; Hanna, 1992), but will be more similar to the highest AM rates used. It is possible that the inherent modulations will interfere with the processing of these higher rates of imposed AM (cf. Houtgast, 1989). At the same time, auditory filtering will reduce the depth of any imposed AM for rates higher than half the auditory filter bandwidth. This may account for the finding that the increase in AM-rate discrimination thresholds at high rates is greater when the noise is filtered into a low-frequency region (where auditory filters are relatively narrow) than when it contains high-frequency energy (Hanna, 1992).

The present study measured the ability of normal listeners to detect changes in the fundamental frequency ($F0$) of bandpass-filtered harmonic complexes. When the $F0$ is low relative to the frequency region into which the stimuli are filtered, changes in the $F0$ of the complexes, which resemble filtered pulse trains, are encoded by purely temporal factors (Hoekstra, 1979; Shackleton and Carlyon, 1994). Bandpass-filtered harmonic complexes can, therefore, provide a measure of purely temporal rate coding under conditions, where, like AM noise, the rate and place of excitation are unrelated, but without the random fluctuations inherent in noise stimuli. Furthermore, the combinations of $F0$ and filter settings under which pitch is derived from purely temporal codes have been extensively measured (Shackleton and Carylon, 1994; Carlyon, 1996, 1998; Carlyon *et al.*, 2000; Grimault *et al.*, 2000). We used these sounds with the aim of providing new information on what, if any, limitations are imposed by the central auditory system when the frequency of phase locking is not tied to the place of excitation.

A potential complicating factor in the interpretation of the results arises from the action of cochlear filtering on our stimulus. This could affect performance in one of two ways, depending on the particular cue to which subjects attend. At high $F0$s, the components of the harmonic complex will become resolved by the peripheral auditory system, thereby limiting the highest rates at which discrimination can be attributed to purely temporal processes. For example, Hoekstra (1979) showed that, as $F0$ was increased above about a fifth of the filter center frequency (CF), thresholds dropped markedly and were similar to those for pure tones. Hence, it is possible that at the highest $F0$, subjects could attend to changes in the frequencies of individual, resolved, components. To address this, we presented some of our complexes in alternating phase, which approximately doubles the perceived temporal pitch (by adding an extra pulse midway through each period) but which leaves the *long-term* spectrum unchanged (Shackleton and Carlyon, 1994). The effectiveness of this solution is qualified by the existence of differences between the *short-term* spectra of sine- and alternating-phase complexes, an issue that will be addressed further in Sec. IV. A second consequence of cochlear filtering is that it may smooth the temporal representation of the stimuli. When subjects perform temporal processing on the output of a fixed set of auditory filters, this could elevate thresholds at high rates and could be mistaken for a more central limitation. By measuring $F0$ difference limens (''$F0$DLs'') as a function of rate for complexes filtered into three different frequency regions, we examined whether any high-rate threshold elevation resulted from auditory filter ringing or from more central factors. If it were caused by ringing, then the rate at which the deterioration occurs should be increased as the stimulus is filtered into higher frequency regions. (This prediction stems from the facts that the duration of any ringing will vary inversely with auditory filter bandwidth, and that these bandwidths are greatest at high CFs.)

II. EXPERIMENT 1

A. Rationale and overview

Experiment 1 measured $F0$ difference limens ($F0$DLs) for harmonic complexes bandpass filtered between 3900–5400 Hz. Shackleton and Carlyon (1994) have shown that in this ''HIGH'' frequency region, the harmonics of complexes having $F0$s no greater than 250 Hz are unresolved by the peripheral auditory system. Their definition of an unresolved complex was one that fulfilled three criteria: (a) summing the components in alternating as opposed to sine phase doubled the pitch; (b) at least three components interacted within the 10-dB-down bandwidth of an auditory filter centered on the complex; and (c) the $F0$DL was higher than all of the ''resolved'' complexes they tested. Although they did not study $F0$s higher than 250 Hz, their second criterion, combined with the observation that the 10-dB bandwidth of an auditory filter centered on the geometric mean of the passband (4590 Hz) is 936 Hz (Glasberg and Moore, 1990), suggests that $F0$s up to about 312 Hz will be unresolved in this region.

As discussed in the Introduction, we measured $F0$DLs as a function of $F0$ not only for complexes summed in sine phase, but also for those summed in alternating phase. If the use of purely temporal cues were limited to relatively slow rates, then we might expect to see a deterioration at a given $F0$ in the alternating-phase but not in the sine-phase condition. A specific prediction was that as the $F0$ of alternating

phase complexes increased to some value, the $F0$DLs would increase, and, as further $F0$ increases led to the spectral components becoming resolvable, the $F0$DL would drop again. In contrast, no such deterioration might be observed in the sine-phase condition, where there may be no $F0$ sufficiently high for the pulse rate to be too fast to be accurately processed, without the spectral components becoming resolvable. This latter prediction is in line with the data of Hoekstra (1979) and of Cullen and Long (1986), who observed no rate at which the rate discrimination of bandpass-filtered pulse trains (effectively cosine-phase harmonic complexes) broke down. The issue of whether any deterioration observed in the alternating-phase condition resulted from auditory filter ringing, or from more central factors, will be investigated in experiment 2.

B. Method

The method of stimulus generation was similar to that described by Micheyl and Carlyon (1998). All stimuli were generated at a sampling rate of 40 kHz by summing a large number of harmonics (sufficient to fill the skirts of the bandpass filter down to an attenuation of 27 dB). The components of each complex were summed either in sine phase or in ''alternating phase'' (odd harmonics in sine phase, even harmonics shifted by 90°). The harmonic complexes were played out of a CED1401 16-bit DAC, antialiased (Kemo VBF25.01, 3-dB-down point=16 600 Hz, attenuation=100 dB/octave), passed through a low- and a high-pass Kemo VBF25.03 filter (48 dB/octave attenuation, cutoffs 5400 and 3900 Hz) in series, and attenuated. The overall level of each complex was on average 66 dB SPL, but this level was roved over a 6-dB range from presentation to presentation in order to discourage the use of any loudness differences that may have co-occurred with changes in $F0$ (cf. Henning, 1966). The complexes were mixed with a continuous pink noise having a passband extending from dc to 20 kHz, and a spectrum level of 21 dB SPL at 4 kHz, before being presented to one earpiece of a Sennheiser HD414 headset.

$F0$DLs were measured using a 2I-2AFC adaptive procedure (Levitt, 1971). On each trial the $F0$s of the signal and standard stimuli were geometrically centered on the baseline value, which was 75, 106, 150, 212, 300, 424, or 600 Hz. The subject's task was to indicate whether the signal (higher $F0$) was presented before or after the standard. Correct-answer feedback was provided visually after each trial. The signal duration was 200 ms (including 20-ms raised-cosine ramps) and the interstimulus interval was 500 ms. At the start of each adaptive run the difference between the standard and signal $F0$s (the ''ΔF0'') was 20%. This difference was decreased after every two consecutive correct responses and increased after every incorrect response, with any change from decreasing to increasing difference or vice versa, constituting a ''turnpoint.'' Each adaptive run continued until 16 turnpoints had been reached, with threshold for that run being estimated from the geometric mean of the last 12. The step size (by which the $\Delta F0$ was adjusted) corresponded to a factor of 2.0 for the first four turnpoints and a factor of 1.19 thereafter. If an adaptive run converged on a $\Delta F0$ higher than the starting $\Delta F0$, the threshold for that run was set to an arbitrary value of 40%. Thresholds presented here are the geometric means of at least three runs per condition (four for listener GN), except when more than one run was assigned the arbitrary 40% value, in which case a value of 40% was plotted. The rules for dealing with these instances of very poor performance were the same as in the study by Carlyon *et al.* (2000, experiment 3).

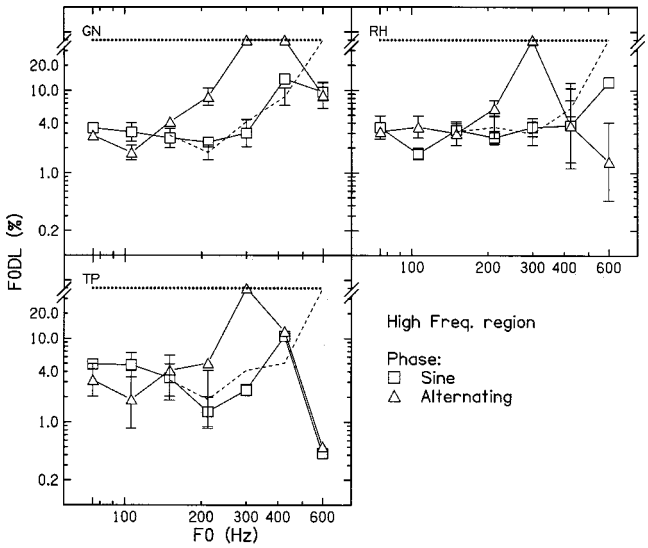


FIG. 1. $F0$ discrimination thresholds ($F0$DLs) plotted as a function of baseline $F0$ for harmonic complexes filtered into the HIGH region and summed in sine phase (squares) or in alternating phase (triangles). Cases where an $F0$DL could not be tracked are plotted at an arbitrary value of 40% (also indicated by a horizontal dotted line). The dashed line with no symbols shows the alternating-phase data shifted to twice the $F0$ tested, so that the pulse rate corresponds to that shown on the abscissa. The alternating-phase data obtained at the highest $F0$s have been omitted from this part of the plot, in order to fit onto the plot area.

Three listeners took part, aged 18, 21, and 55. Listeners TP and RH had thresholds within 15 dB of the 1969 ANSI standard at all audiometric frequencies. This was also true for GN (the oldest listener), except for thresholds of 21- and 28 dB HL at 4000 and 8000 Hz, respectively.

C. Results

$F0$DLs for each listener are shown in the three panels of Fig. 1. For the alternating-phase stimuli (triangles), thresholds are roughly constant at 2%–3% for $F0$s up to 150 Hz, and then increase dramatically. At an $F0$ of 300 Hz, it was not possible to track a threshold for any listener, and the data are plotted at an arbitrary value of 40%. This was also true for listener GN at an $F0$ of 424 Hz. As the $F0$ was increased further the F0DLs dropped again, consistent with the spectral components of the complex becoming at least partly resolvable. Note, however, that only for listener TP do the $F0$DLs drop below those observed at the lowest F0s. This may have been because even at the highest $F0$s the spectral components were not *completely* resolved.

The breakdown in performance at 300 Hz in the alternating-phase conditions reflects an inability of the auditory system to process pulse rates as high as 600 pps. No such deterioration is observed in the sine-phase condition

(squares), presumably because, at an $F0$ of 300 Hz, the pulse rate is 300 (rather than 600) pps. However, a smaller deterioration *is* observed in the data of GN and TP at an $F0$ of 424 Hz, and of listener RH at an $F0$ of 600 Hz. This deterioration may have occurred because at these higher $F0$s the pulse rate was too fast, even for sine-phase stimuli, to be accurately encoded by temporal pitch mechanisms, and because spectral cues were not sufficiently strong to maintain performance. The reasons why such a deterioration has not been observed previously will be addressed in Sec. VI.

The idea that performance for unresolved complexes is determined by pulse rate rather than $F0$ is supported by a comparison between the data obtained in sine- and alternating phase, when the latter are shifted rightwards along the abscissa to twice their $F0$ (dashed lines). When transformed in this way, the values labeling the abscissa refer to pulse rate, which is equal to $F0$ for the sine-phase stimuli and to $2F0$ for the alternating-phase stimuli. It can be seen that the two sets of data are very similar for pulse rates up to about 424 pps, diverging only at 600 pps, where components may become resolved for the sine-phase stimuli, whose $F0$ is 600 Hz, but not for the alternating-phase stimuli, whose $F0$ is 300 Hz.

Finally, we should comment on the large error bars for three points of RH's data. In each case these resulted from the combination of one or two very low thresholds ($\leqslant$1%) with one or two very high thresholds ($>$9%). This may be related to Cullen and Long's (1986) observation of a bimodal distribution of thresholds for some listeners and conditions, a finding consistent with subjects using weak spectral cues on some trials but not on others.

III. EXPERIMENT 2: DIFFERENT FREQUENCY REGIONS

A. Rationale and method

The results of experiment 1 indicate that, for complexes filtered between 3900–5400 Hz, temporal pitch mechanisms cannot derive an accurate pitch estimate from pulse rates as high as 600 pps. What it does not show is whether this reflects a central limitation in the ability to process high pulse rates, or whether it results from peripheral factors such as temporal smoothing introduced by ringing of the auditory filters responding to the complex. To differentiate between these two explanations, experiment 2 measured $F0$DLs for alternating-phase stimuli in three different frequency regions. Experiment 2a tested the ''MID'' frequency region (1375–1875 Hz) described by Shackleton and Carlyon (1994) and the 3900–5400 ''HIGH'' region used in experiment 1. The $F0$DLs tested were chosen to span the values where, according to a filter-ringing hypothesis and the results of experiment 1, a deterioration in performance might be expected. These were 90, 106, 126, 150, 178, 212, and 252 Hz in the MID region, and 178, 212, 252, 300, 356, 424, and 503 Hz in the HIGH region. Experiment 2b tested a new ''very high (VHIGH)'' region, in which the complexes were filtered between 7800–10 800 Hz. In this region, baseline $F0$s of 252, 300, 356, 424, 503, 600, 714, and 850 Hz were used. The overall level of the stimuli and the method of stimulus generation were the same as in experiment 1.

In a supplementary experiment, we measured the detection thresholds for pure tones in the presence of the pink-noise background. This was done to gain some insight into the effects of a decrease in absolute sensitivity at very high frequencies on the detectability of various parts of the complexes when presented in noise. We used a two-down, one-up adaptive tracking procedure (Levitt, 1971), and pure tones having frequencies of 4589 and 9178 Hz (the geometric means of the HIGH and VHIGH passbands), of 6490 Hz (20.2% away from both the upper cutoff of the HIGH band and the lower cutoff of the VHIGH band), and 12 982 Hz (20.2% above the upper cutoff of the VHIGH band). Subsequently, we also measured the detection thresholds for complex tones summed in alternating phase, having an $F0$ of 300 Hz in all three regions and, additionally, for an $F0$ of 150 Hz in the MID region. Only two listeners (JD and AO) participated in this last part of the experiment, in which Sennheiser HD250 headphones were used.

B. Listeners and procedure

Six listeners took part in experiment 2a, and four took part in experiment 2b. In total, four new subjects participated, all of whom were aged under 30 and had thresholds within 15 dB of the 1969 ANSI standard at all audiometric frequencies. Experiment 2b took place more than 3 years after experiments 1 and 2a. Listener GN was initially tested in experiment 2b, but his chance performance at all baseline $F0$s in the VHIGH region led us to retest his detection thresholds in quiet. This retest revealed that his threshold at 8000 Hz had increased to 53 dB SPL (40 dB HL), and so we chose to exclude him from this experiment and from all further measurements in the VHIGH region. His threshold at 4000 Hz was 33.6 dB SPL (24.6 dB HL), only 4 dB higher than measured originally, and so he was tested in the HIGH region in experiment 3.[1]

The procedure was the same as in experiment 1, except that the $\Delta F0$ at the start of each run in the HIGH region of experiment 2a was increased to 32% for the 300-Hz $F0$. This higher initial starting value was used in all conditions of experiment 2b.

C. Results

1. Experiment 2a

Each panel of Fig. 2(a) shows the data from one listener in experiment 2a. The $F0$DLs obtained in the HIGH frequency region, shown by open triangles, follow a similar pattern to that observed in experiment 1. All six listeners show a breakdown in performance ($F0$DLs plotted as 40%) at an $F0$ of 300 Hz, with this deterioration being extended to adjacent $F0$s for all except listener DN. The data from the three listeners who took part in experiment 1 are generally in good quantitative agreement with those earlier results. Minor exceptions come from the data of RH and GN at the highest $F0$ of 503 Hz, where $F0$DLs were lower than in the 600-Hz alternating-phase condition of experiment 1. Because $F0$DLs

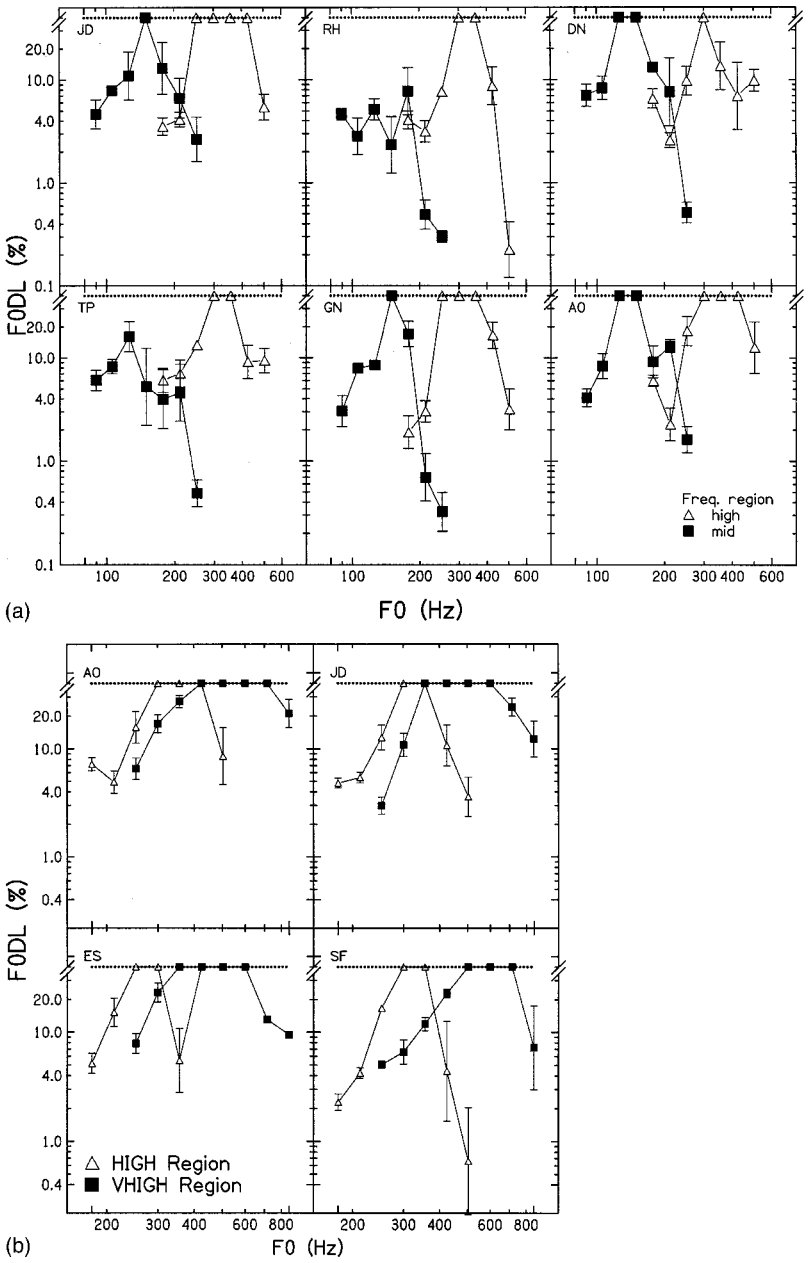


FIG. 2. (a) $F0$DLs for alternating-phase complexes filtered into the MID (filled squares) or HIGH (open triangles) region. (b) $F0$DLs for alternating-phase complexes filtered into the HIGH (open triangles) or VHIGH (filled squares) region. Cases where an $F0$DL could not be tracked are plotted at an arbitrary value of 40% (also indicated by a horizontal dotted line).

tend to *increase* as $F0$ is reduced below 600 Hz, this difference is unlikely to result from the slightly different $F0$s studied. Rather, it is consistent with these two subjects having learned to take advantage of subtle spectral cues in experiment 2a.

The data from the MID region are shown by the solid squares. Four out of the six listeners could not converge on a threshold for an $F0$ of 150 Hz, and two of these, DN and AO, could not do so at an $F0$ of 126 Hz either. The $F0$DLs for two listeners, RH and TP, are below 20% at all $F0$s studied. However, listener TP shows large standard errors at $F0$s of 126 and 150 Hz, reflecting an inability to track a threshold below the starting value of 20% on a subset of trials. Listener RH was able to track a threshold at all $F0$s, but also showed larger standard errors at intermediate $F0$s.

2. Experiment 2b

Each panel of Fig. 2(b) shows data from one listener in experiment 2b. The results for stimuli filtered into the HIGH region (open triangles) are similar to those in experiments 1 and 2a, with no listeners being able to track a threshold once the $F0$ reached 300 Hz. A new listener, ES, was also unable to do so at an $F0$ of 252 Hz. He also showed highly variable and nonmonotonic $F0$DLs at the highest $F0$s studied in the HIGH region, perhaps reflecting an inability to consistently use whatever place-of-excitation cues were available in those conditions (cf. Cullen and Long, 1986). When the stimuli were filtered into the VHIGH region, all listeners showed an increase in the ''breakdown point,'' which we define as the lowest $F0$ at which a threshold cannot be tracked. Importantly, they could all track a threshold of less than about 20% at an $F0$ of 300 Hz, which corresponds to a pulse rate of 600 pps. (Note that if subjects had been using residual spectral cues in this condition, we would have expected to see performance *improve* as $F0$ was increased further to 424 Hz; instead, it deteriorated.) This upper limit of discriminability is greater than that observed with the majority of cochlear implant listeners (Shannon, 1983a; Tong and Clark, 1985;

TABLE I. (a) Pure-tone detection thresholds (in dB SPL) of three listeners in the presence of the continuous pink noise used in the experiments. Mean thresholds are shown in bold type, with the standard error of the three runs per condition shown in italics. (b) Masked thresholds for alternating-phase harmonic complexes. Data are shown for an $F0$ of 150 Hz in the MID region, and for an $F0$ of 300 Hz in the MID, HIGH, and VHIGH regions of experiment 2.

(a)	JD		SF		AO	
4589 Hz	**41.7**	*1.2*	**40.8**	*0.3*	**42.2**	*0.7*
6490 Hz	**43.4**	*0.2*	**37.3**	*2.4*	**41.1**	*0.3*
9178 Hz	**48.7**	*2.0*	**45.2**	*1.5*	**41.7**	*0.7*
12982 Hz	**43.6**	*1.3*	**44.8**	*0.7*	**75.5**	*0.7*

(b)	JD		AO	
150 Hz, MID	**44.44**	*0.55*	**47.06**	*1.51*
300 Hz, MID	**52.94**	*0.44*	**45.90**	*0.30*
300 Hz, HIGH	**49.64**	*0.70*	**46.94**	*0.57*
300 Hz, VHIGH	**43.57**	*0.45*	**46.16**	*0.73*

Townshend *et al.*, 1987; McKay *et al.*, 2000). For example, McKay *et al.* (2000) measured the ability of four implant users to detect a 20% increase in a pulse rate of 500 Hz, and observed chance performance in all four cases. This in turn suggests that the limitations in temporal processing seen with implant users do not stem entirely from the fact that rate and place of stimulation are not usually matched in electric hearing, because the same was true for our acoustic stimuli.

3. Masked thresholds

Detection thresholds for pure tones masked by the continuous pink noise of experiment 2 are shown in Table I(a). Thresholds are roughly constant across all frequencies tested, with the exception of an elevated threshold for listener AO at 12 982 Hz. The implications of these results will be discussed in Sec. VII A.

Masked threshold for alternating-phase harmonic complexes are shown in Table I(b). For an $F0$ of 300 Hz, thresholds are independent of frequency region for listener AO, and drop slightly at higher regions for listener JD. This suggests that differences in sensation level across frequency region did not play a very large role in the pattern of results obtained in experiment 2. The average threshold was about 48 dB SPL, so our stimuli (which had a level of 66 dB SPL) were presented at a sensation level of approximately 18 dB. For listener JD, masked threshold in the MID region was lower when the $F0$ was 150 Hz than when it was 300 Hz; however, the same was not true for listener AO.

4. Summary

The results of experiment 2 indicated that rate discrimination by normal listeners may be limited, at least in part, by auditory filter ringing. This conclusion is supported by the general observation that the nonmonotonic functions observed in experiment 2 shift towards higher $F0$s as frequency region is increased. However, this does not demonstrate that filter ringing is the *only* limitation, and this issue will be addressed further in later sections.

IV. EFFECTS OF COCHLEAR FILTERING AND SHORT-TERM EXCITATION-PATTERN CUES

A. Disruptive excitation-pattern cues?

So far, we have discussed the breakdown in performance for alternating-phase stimuli in terms of a limitation in temporal, rather than spectral, processing. This interpretation receives some support from the finding that the results differ for sine-phase stimuli, whose *long-term* spectra are identical to those of the alternating-phase stimuli. However, Krumbholz *et al.* (2000), in an implementation of Cohen *et al.*'s (1995) SPINET model of pitch perception, have shown that the *short-term* excitation patterns of sine- and alternating-phase complexes can differ, as a result of the interaction between the ringing response of the auditory filter to adjacent ''pulses'' in the stimulus waveforms. At low $F0$s, the pulses in both the sine- and alternating-phase complexes are sufficiently spaced in time to preclude such interactions, and no excitation-pattern cues are present. As $F0$ is raised, the filter responses to adjacent pulses in the alternating-phase stimuli start to overlap. Because these pulses are phase-shifted versions of each other, this does not produce a short-term spectrum containing peaks at integer multiples of $F0$. Instead, the peaks in the spectrum and excitation pattern are spaced at intervals of $2F0$, and, in the output of the SPINET model, oscillate throughout the stimulus. In contrast, sine-phase stimuli provide no short-term excitation pattern cues at all until twice the $F0$ at which they start to appear for alternating-phase stimuli (because the pulses are twice as far apart).

The existence of such short-term excitation-pattern cues is important for the interpretation of our results, because they allow an account that does not depend solely on limitations in purely *temporal* processing. Specifically, it may be that performance breaks down for alternating-phase stimuli once a certain $F0$ is reached, not because of a limitation in temporal processing *per se*, but because of excitation-pattern cues that interfere with subjects' ability to selectively and optimally process the temporal information. Hence, performance with an $F0$ of 300 Hz and alternating phase could be better in the VHIGH than in the HIGH region because the shorter impulse responses of auditory filters prevent ''dis-

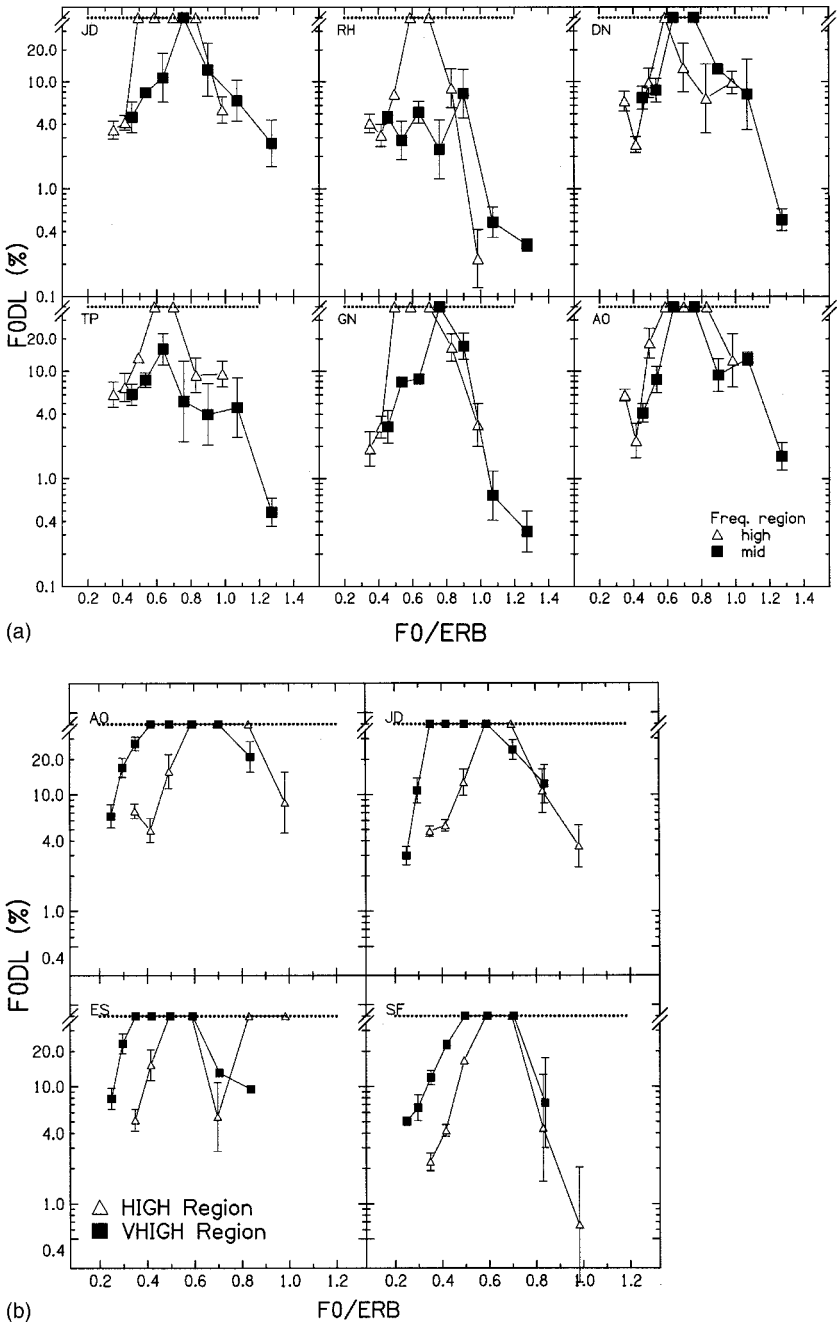


FIG. 3. DL$F0$s obtained in experiment 2, with the abscissa plotted as the $F0$ of the complex divided by the ERB of an auditory filter having a CF in the middle of the passband of that complex. (a) MID and HIGH regions (experiment 2a); (b) HIGH and VHIGH regions (experiment 2b).

turbing'' short-term excitation pattern cues from occurring until higher $F0$s.

It is worth stressing that the above argument, if true, would not contradict one important conclusion from experiment 2, which is that, in the VHIGH region, listeners can use purely temporal processing to detect changes in pulse rate *re*: a 600-pps standard. Indeed, it would argue that the limit on temporal processing is in fact higher, and that subjects performed poorly at 600 pps only because of disruptive features of the short-term excitation pattern. However, it could influence the interpretation of why the breakdown point varies with frequency region. If this *were* attributable to excitation-pattern cues, then one would expect to see the breakdown point to occur at the same point in each frequency region, when $F0$ is normalized to the auditory filter bandwidth in that region. The same prediction is made by the hypothesis that the limitation is *entirely* due to auditory filter ringing. To test this, Fig. 3 replots the data of Fig. 2 with the abscissa showing $F0$ divided by the equivalent rectangular bandwidth (ERB) of an auditory filter geometrically centered on the passband of the physical filters used in each region. The results show that this prediction is not upheld; there is some evidence that the deterioration occurs at a lower value of $F0$/ERB in the HIGH than in the MID region [Fig. 3(a), subjects JD and GN], and strong evidence that it occurs at an even lower $F0$/ERB in the VHIGH region [Fig. 3(b), all subjects].

Another way of evaluating whether performance was limited by disruptive short-term excitation-pattern cues is to vary the nature of the task. Experiments 1 and 2 used an adaptive procedure, which may have caused the excitation-pattern and temporal cues to vary from trial to trial, and may

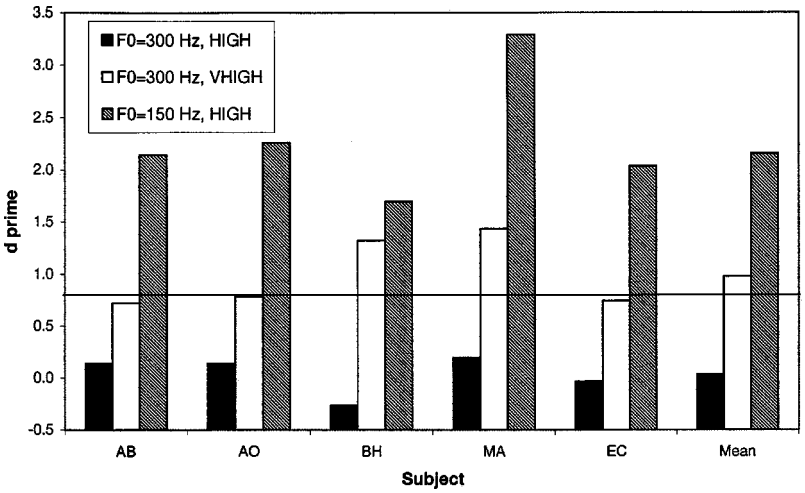


FIG. 4. Histogram showing the results of experiment 3 for alternating-phase complexes. The horizontal line indicates a value of $d' = 0.78$, which, on the assumption of no response bias, corresponds to the 71% point on the psychometric function (on which the adaptive procedures used in experiments 1 and 2 converged).

have forced subjects to explicitly extract the pitch of the standard and signal stimuli. We therefore performed an additional experiment in which the same stimuli were used within each block of trials. In addition, this experiment used a three-interval, two-alternative task which allows subjects to pick the ''odd man out'' in any way that gives the correct answer, without having to explicitly extract the pitches of the sounds. If excitation-pattern cues were sufficiently disruptive to interfere with the detection of temporal processing in experiment 2, then one might expect listeners to be able to use them in experiment 3, where the cue is the same on every trial and where they can use any perceptual dimension to perform the task.

B. Experiment 3: Method of constant stimuli

1. Method

Experiment 3 was similar to experiment 2b except that only three conditions were included, and that the method of constant stimuli was used. The three conditions, all of which used alternating-phase stimuli, consisted of nominal $F0$s of 150 and 300 Hz in the HIGH region, and of 300 Hz only in the VHIGH region. The difference in $F0$ between the standard and signal stimuli was kept fixed at 20%, geometrically centered on the nominal values. Each condition was tested in turn, in blocks of 40 trials, until the subject had completed 200 trials per condition. In addition, the task was changed to a three-interval, two-alternative forced choice, in which subjects were required only to identify whether the second or third sound in each trial differed from the other two. This was done so that subjects would not have to make any judgment about the pitch of the sound, but merely have to pick the odd man out—thereby further reducing the chances that a usable temporal cue would be disrupted by irrelevant excitation-pattern cues. In this paradigm, subjects could perform the task using any cue except level, which was roved as before. The only other changes were that Sennheiser HD250 headphones were used, and that the signal duration was increased to 400 ms (including 50-ms raised-cosine ramps).

2. Results

The results of experiment 3 are shown in Fig. 4, and, for all five listeners, show a pattern that is entirely consistent with that obtained from the adaptive procedures of experiments 1 and 2. Specifically, performance is near chance at an $F0$ of 300 Hz in the HIGH region, and improves either when the complexes are filtered into the VHIGH region or when the $F0$ is reduced to 150 Hz. The improvement caused by shifting to the VHIGH region is smaller than that caused by reducing $F0$, consistent with the fact that the average DL$F0$ in the former case is 12.9% [Fig. 2(b) filled squares], compared to 3.1% in the latter (Fig. 1, open triangles). A one-way repeated-measures ANOVA revealed a highly significant main effect of condition [$(F(2,8) = 53.3,\ p < 0.001$]; differences between all pairs of conditions were also significant ($p < 0.01$ in all cases).

In conclusion, the results of experiment 3 replicate the same general pattern of findings obtained in experiments 1 and 2. Because these results persisted in a paradigm with minimum uncertainty and where subjects could use virtually any cue, it appears that they represent a fundamental limitation in monaural processing of rate information.

V. EXPERIMENT 4: CHANGES IN PERFORMANCE PRODUCED BY A CONTRALATERAL STANDARD

A. Rationale

Experiment 4 was designed to further constrain explanations for the breakdown in performance observed in experiment 2 for high-rate alternating-phase stimuli. The general approach was to determine whether, under conditions where performance in experiments 1 and 2 was at chance, any information on the timing of the stimulus pulses was preserved at the output of peripheral auditory filters. To do this, it asked whether another form of temporal processing, carried out by the binaural system, can exploit temporal information in the auditory nerve under conditions where monaural rate discrimination was impossible. This seemed plausible because it is well known that the binaural system is exquisitely sensitive to interaural timing differences (ITDs). Normal listeners can detect ITDs of as little as 10 μs between filtered acoustic pulse trains (Hafter and Dye, 1983), and some cochlear implant users can detect electrical ITDs of the order of 50–200 μs (Long *et al.*, 1998; Lawson *et al.*, 1999).

We modified a technique previously used in a study on bilateral cochlear implants by van Hoesel and Clark (1997),

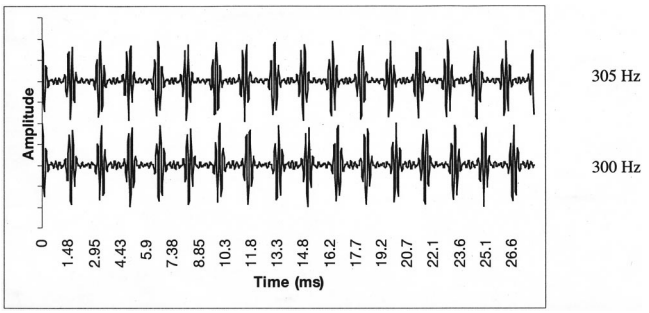


FIG. 5. Schematic representation of the pitch-pulse asynchronies that occur when alternating-phase complexes, having slightly different $F0$s, are presented to opposite ears. The waveforms shown are from two alternating-phase complexes, filtered into the HIGH region, and with $F0$s of 300 and 305 Hz (pulse rates of 600 and 610 Hz, respectively).

in which subjects discriminated between the rates of a standard and a (higher) signal pulse train in one ear, while a copy of the standard stimulus was presented, in both observation intervals, to the opposite ear. Although they did not discuss their data in this way, it seems likely that the contralateral standard could give rise to a cue based on either the diffuseness of the binaural image, or perhaps to the perception of spatial motion. In the standard interval the same stimulus is presented to both ears, and subjects should hear a centered image, whereas in the signal interval the pulses drift out of phase in the two ears. These ''pitch-pulse asynchronies'' are illustrated in Fig. 5, which shows our acoustic analog of the technique. Note that, although the contralateral standard provides no new information (being the same on every presentation), it could allow subjects to use binaural cues to process the temporal information in the ''target'' ear, possibly bypassing the temporal pitch mechanism. The prediction is that, if no information on the temporal structure of the stimuli is preserved in the auditory nerve, performance should be at chance both with and without a contralateral standard. If, however, some information is present which can be used by binaural mechanisms but not by pitch mechanisms, then the contralateral standard may improve performance. An additional advantage of this stimulus is that, because the initial pulses in the two ears are always simultaneous, it prevents subjects from using ''onset'' cues, as would have been possible if we had, for example, simply measured sensitivity to a delay in the waveform to one ear.[2]

B. Method

The method of stimulus generation was similar to that used in experiments 1 and 2. However, the use of the contralateral standard made it inappropriate to adopt the adaptive procedure of those experiments. This is because the binaural cue, which depends on the asynchrony between envelope maxima (''pulses'') in the two ears, may not vary monotonically with $\Delta F0$. As shown in Fig. 5, the $\Delta F0$ causes the pulses to drift out of phase, but eventually this asynchrony will rotate through 180° and the pulses will then gradually move back into phase. For a ''true'' pulse train the pulses will coincide once every $1/\Delta F0$s (where $\Delta F0$ is in Hz). For our alternating-phase stimuli, where there are two envelope peaks per period, a peak in one ear will coincide with one in the opposite ear every $0.5/\Delta F0$s. Experiment 4 used a signal duration of 400 ms (including 50-ms raised-cosine ramps). For this duration, any $\Delta F0$ greater than 1.25 Hz would cause the envelope maxima in the two ears to drift in and out of phase more than once. To avoid this, we measured four-point psychometric functions for values of $\Delta F0$ of 0.31, 0.63, 0.94, and 1.25 Hz. Performance was measured for alternating-phase stimuli at an $F0$ of 300 Hz, both in the HIGH and (except for listener GN) VHIGH regions of experiment 2, and both in the presence and absence of the 6-dB level rove used in that experiment. In all other respects (stimulus level, background noise, filter settings) the method of stimulus generation was the same as in experiment 2. The only difference in the trial structure was that a three-interval, two-alternative forced-choice task was used, in which the signal could occur in either the second or third interval of each trial. This allowed subjects to perform the task using any perceivable difference between the signal and standard, rather than being influenced by the experimenter's description of what the cue was. This was deemed advisable because we were not sure that all subjects would perceive any binaural cue in the same way.

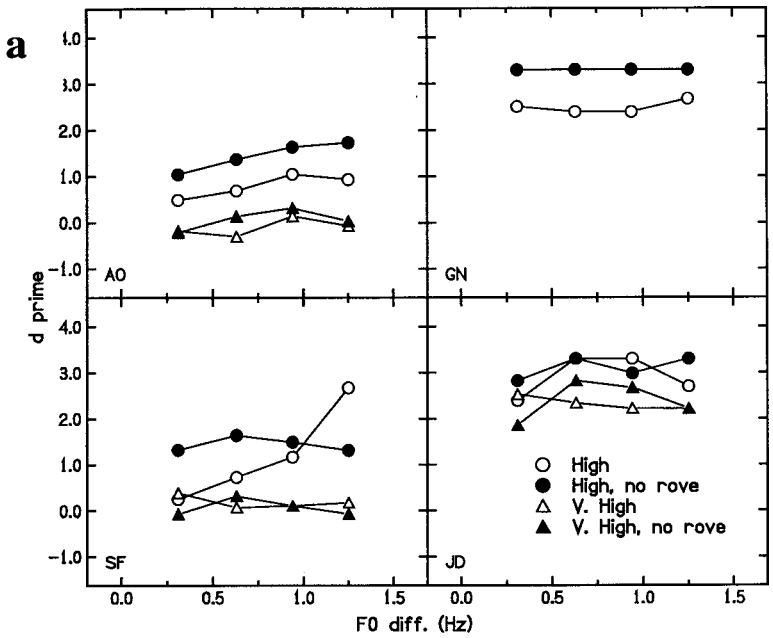
Because the largest $\Delta F0$ studied corresponded to a difference of less than 0.42% between the standard and signal, we expected performance to be at chance in the absence of binaural cues. We therefore initially measured psychometric functions only in the presence of a contralateral standard, using the method of constant stimuli. First, 20 trials at each point on the function were measured for one of the four conditions (HIGH or VHIGH region, presence or absence of level rove). This was done in sub-blocks, so that 20 trials at one value of $\Delta F0$ were run, preceded by two practice trials, before moving on to a different value of $\Delta F0$. The order of the $\Delta F0$s to be tested was selected at random. Then, the other three conditions were tested in this way, with the order of conditions selected at random. This whole procedure (with fresh randomizations of condition and $\Delta F0$ order) was repeated five times so that 100 trials per point were measured for each condition. The data were then averaged across the 100 presentations for each point and converted to the sensitivity index, d'. When performing ANOVAs on the data for each subject, we calculated d' for each sub-block of 20 trials and used the variation across the d' in these different repetitions to derive an error term.

After the main experiment was completed, we remeasured d' in the HIGH region, alternating blocks of trials with a control condition in which the contralateral standard was absent. This was done because we observed above-chance performance with dichotic stimulation, and because we wanted to check that this genuinely resulted from the use of binaural cues, rather than to the differences in procedure *re*: experiments 1 and 2. For this supplementary experiment only, Sennheiser HD250 headphones were used. The stimulus level was not roved.

C. Results

1. HIGH region

Each panel of Fig. 6(a) shows sensitivity (d') for one of the four subjects of experiment 4, plotted as a function of the



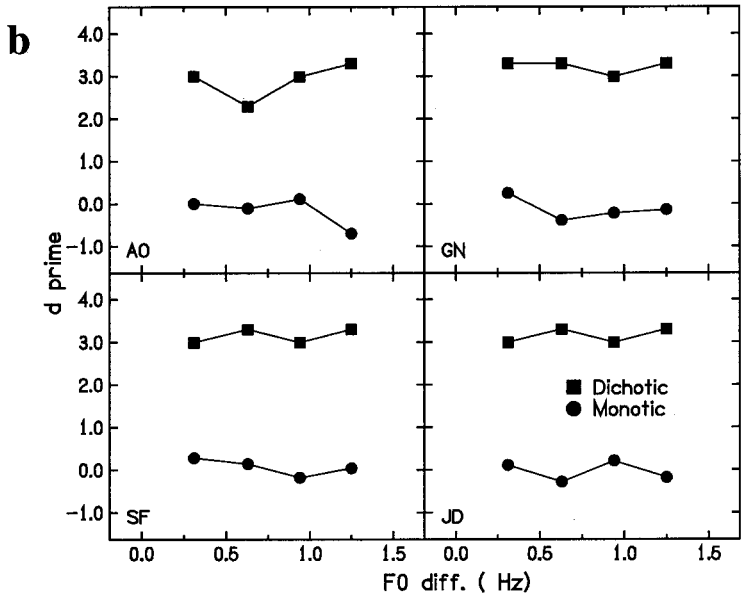


FIG. 6. (a) Sensitivity to $\Delta F0$s in the dichotic conditions of experiment 3 for stimuli filtered into the HIGH (circles) and VHIGH (triangles) regions. Data obtained with a level rove are shown by open symbols, and those without a level rove by filled symbols. (b) Data from the supplementary experiment in experiment 3 for the dichotic (squares) and monotic (circles) condition.

$F0$ difference between the standard and signal. Both with (open circles) and without (filled circles) the 6-dB level rove, all subjects achieved above-chance performance in the HIGH frequency region. Performance was at or close to ceiling for subjects JD and GN, and below ceiling (but above chance) for listeners AO and SF. The second part of the experiment [Fig. 6(b)] showed that performance in the HIGH region and with no level rove was superior with (squares) than without (circles) a contralateral standard. Hence, the binaural system is able to utilize some aspect of the rate difference between the signal and standard which is not accessible to the subjects when they have to judge pulse rate using monaural mechanisms. It is also worth noting that performance in the HIGH region was better in experiment 4b than in experiment 4a. This may have been due either to practice, or to the fact that, in the second part of the experiment, there were no blocks of harder dichotic conditions (VHIGH region, and HIGH region with level rove) mixed in with the condition under test (HIGH region, no rove).[3]

2. VHIGH region

Perhaps surprisingly, performance in the VHIGH region [Fig. 6(a), triangles] for the main (dichotic) part of the experiment was close to chance for all listeners except JD. Three-way ANOVAs (rate$\times\Delta F0\times$freq. region) revealed that performance was significantly worse in the VHIGH than in the HIGH region for listeners AO [$F(1,4)=94.6$, $p<0.001$], and SF [$F(1,4)=40.5$, $p<0.005$]. For listener JD, whose performance was close to ceiling in all conditions, this difference approached but did not reach significance [$F(1,4)=4.92$, $p=0.061$].[4]

The generally inferior performance observed in the VHIGH region may be related to the findings of Bernstein and Trahiotis (1994), who observed that sensitivity to ongoing ITDs decreased slightly as the center frequency (CF) of the stimuli was increased from 4 to 8 kHz, and deteriorated markedly at a CF of 12 kHz. They suggested that the deterioration may be related to an observation, which they reported informally, that cells in the central nervous system that are sensitive to ITDs are rarely sensitive to high spectral frequencies (see also Irvine, 1992). Whether or not this explanation is correct, it does seem that most subjects have difficulty in processing ITDs in very high spectral regions, at least for the stimuli studied here and by Bernstein and Trahiotis. Furthermore, when taken together with the results of experiment 2, they reveal an interesting dissociation between the central limitations involved in a pitch task and those involved in the use of binaural cues. Recall that in experiment 2b [Fig. 2(b)] all subjects could perform the monaural $\Delta F0$ task when the baseline $F0$ was 300 Hz in the VHIGH, but not in the HIGH, region. In contrast, Fig. 6 shows that performance in the binaural task is generally *worse* in the VHIGH than in the HIGH region.

To summarize, the results of experiment 4 allow us to draw two conclusions. First, although listeners cannot detect monaural differences in $\Delta F0$ for a 300-Hz alternating-phase complex in the HIGH region, sufficient temporal information is preserved in the auditory periphery to be usable by the binaural system. Second, subjects are often worse at processing interaural information in the VHIGH region than in the HIGH region, even for stimuli where their monaural rate discrimination is better in the VHIGH region. Taken together with the first conclusion, this shows that different forms of central limitation are present in the two tasks.

VI. COMPARISON WITH PREVIOUS STUDIES AND EXPERIMENT 5: INVESTIGATION OF DETERIORATION AT HIGH *F*0s WITH SINE-PHASE STIMULI

A. Previous research

Figure 7(a) compares the mean data obtained from the alternating-phase complexes of experiment 1 (diamonds) with those reported by Krumbholz *et al.* (2000, circles) for similar complexes, filtered between 3.2 and 3.8 kHz. The two sets of data are similar up to the highest $F0$, 256 Hz, studied by Krumbholz *et al.* (2000). The main new finding of experiment 1 is that, by increasing the $F0$s of alternating-phase complexes even further, it is possible to cause a complete breakdown in $F0$ discrimination performance.

Most previous studies have used either complexes summed in cosine phase, or almost equivalently, filtered pulse trains. The diamonds in Fig. 7(b) show the mean data obtained with sine-phase stimuli in experiment 1. They are compared to data obtained by Hoekstra (1979) for pulse trains passed through a 1/3rd octave filter centered on 4000 Hz, by Cullen and Long (1986) for pulse trains high-pass filtered at 5 kHz, and by Krumbholz *et al.* (2000) for cosine-phase harmonic complexes bandpass filtered between 3.2 and 3.8 kHz. For $F0$s up to 300 Hz, our results are very similar to those obtained by Krumbholz *et al.* (2000, circles),

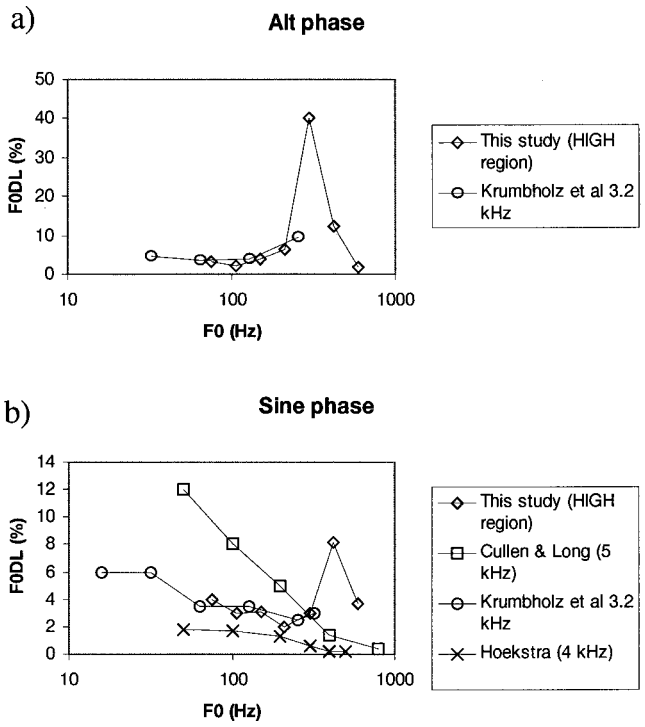


FIG. 7. (a) Mean $F0$DLs obtained in the HIGH region for alternating-phase stimuli, compared to relevant data from the literature. (b) same as part (a), but for sine-phase stimuli.

and show only a modest overall elevation in $F0$DLs relative to those obtained by Hoekstra (1979, crosses). All three studies show thresholds at low $F0$ that are smaller than those observed by Cullen and Long (1986).[5] The largest difference between the results of experiment 1 and those reported previously occurs at the highest $F0$s studied, where our thresholds increased as $F0$ was raised to 424 Hz, and then decreased again at 600 Hz. In the previous studies $F0$DLs decreased monotonically with increasing baseline $F0$, and so we performed two further experiments to investigate why we observed such a deterioration whereas others did not. Experiment 5a investigated whether our use of a level rove was important in reducing the effectiveness of residual place-of-excitation cues. Experiment 5b studied the effects of background noise.

B. General method

The general method was the same as in experiment 1, except that the initial $\Delta F0$ at the beginning of each experimental run was 32% (instead of 20%), and Sennheiser HD250 headphones were used. Baseline $F0$s differing in steps corresponding to a factor of 1.19 were used. The exact ranges of $F0$s tested were selected separately for each subject based on pilot experiments, in order to encompass the region of any deterioration. Each data point was calculated from the geometric mean of four adaptive runs.

C. Experiment 5a: Effect of level rove

Five subjects were tested with and without the 6-dB level rove used in experiment 1. Each subject did one run at all baseline $F0$s for one condition and then the other, until

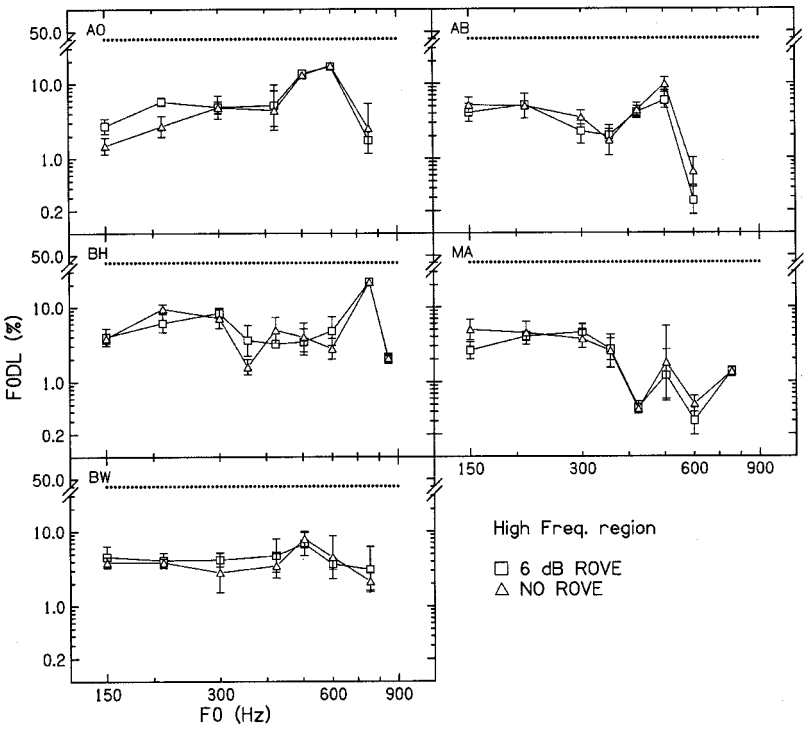


FIG. 8. Results of experiment 5a for the conditions with and without a level rove.

each had completed four runs for each $F0$ and condition. The results, shown in Fig. 8, reveal that the level rove had no effect. Note also that one of the subjects, MA, did not show a deterioration (defined here as a nonmonotonic function with a maximum DL$F0$ of 7% or greater) in this experiment. However, the error bars for her data at an $F0$ of 504 Hz are large, because she obtained a threshold of higher than 10% on one or two of the four runs in each condition. As already mentioned in Sec. II B 2, a bimodal distribution of thresholds was also reported for listener RH in experiment 1 and for the listeners in Cullen and Long's (1986) study. A similar finding occurred for some data points in the results of other subjects in this experiment. In addition, it is worth noting that, although the $F0$ at which the deterioration occurred was stable for a given subject, in that it was the same with and without the level rove, it did vary markedly across subjects. In addition, the size of the deterioration varied across listeners, being largest for AO and BH, and rather small for listener BW. This differs from the highly consistent effect observed with alternating-phase stimuli, and may reflect individual differences is subjects' abilities to use subtle temporal and spectral cues.

D. Experiment 5b: Effect of background noise

Initially, four subjects were tested with and without the continuous background pink noise of experiment 1. Each subject did one run at all baseline $F0$s for one condition and then the other, until each had completed four runs for each $F0$ and condition. These data are plotted in Fig. 9 as open

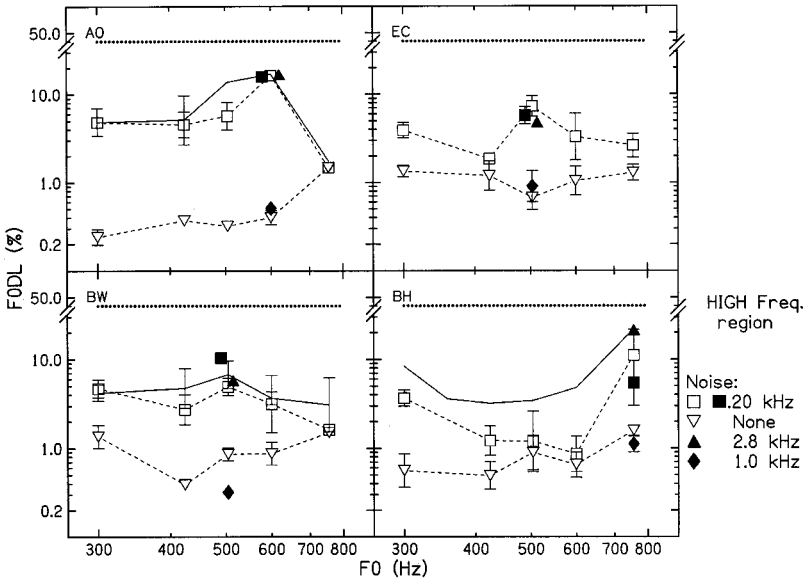


FIG. 9. Results of experiment 5b. Symbols connected by dashed lines show DL$F0$s in the presence and absence of pink noise extending from dc to 20 kHz. Unconnected filled symbols show $F0$DLs at a single $F0$ in the presence of pink noise low-pass filtered at 1000 Hz, 2585 Hz, and 20 kHz.

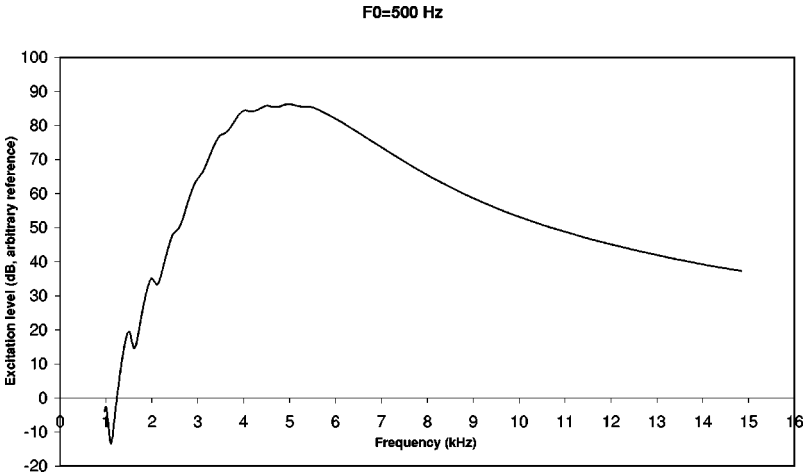


FIG. 10. Excitation pattern (Patterson *et al.*, 1988) of a sine-phase complex having an $F0$ of 500 Hz and filtered into the HIGH frequency region.

squares connected by dashed lines. All listeners showed a region where $F0$DLs increased with increasing $F0$, although this deterioration was very small for listener BW. Data from experiment 5a are replotted as solid lines for the three listeners who took part in both experiments; the results remained the same for two of these, and improved overall for listener BH. The highest $F0$DL for listener BH in experiment 5b was at 756 Hz—the highest $F0$ studied—but, as found in experiment 5a (Fig. 7), his $F0$DL decreased again when he was tested at the even higher $F0$ of 849 Hz.

When the noise was removed (open inverted triangles, dashed lines), $F0$DLS decreased overall and the deterioration disappeared for all listeners. To investigate this further, each listener was retested at the $F0$ giving the maximum $F0$DL in three conditions. One of these used a 20-kHz-wide pink noise, and was a replication of one of the points obtained in the first part of experiment 5b. In the two new conditions, the pink noise was low-pass filtered (Kemo VBF25.01; attenuation rate 100 dB/octave) at 1000 or 2758 Hz. (The latter frequency is half an octave below the lower cutoff of the filter through which the harmonic complexes were passed). The results, plotted as unconnected filled symbols, show that performance with a low-pass filter of 1000 Hz (diamonds) was similar to that obtained with the noise absent, indicating that good performance in the absence of noise did not result from the detection of distortion products having a frequency equal to $F0$ (Wiegrebe and Patterson, 1999; Pressnitzer and Patterson, 2001). However, when the noise was low-pass filtered at 2575 Hz (upright triangles), $F0$DLs were similar to those obtained with the full 20-kHz noise bandwidth (squares). This indicates that the poor performance observed at some $F0$DLs in the presence of broadband noise was not due to the noise ''filling in the valleys'' of the amplitude modulation present in the outputs of auditory filters having CFs within the passband of the complex. A more plausible explanation is suggested by Fig. 10, which shows the excitation pattern derived using the ''gammatone'' filter bank (Patterson *et al.*, 1988) in response to a sine-phase complex having an $F0$ of 500 Hz. It can be seen that, on the lower skirts of the excitation pattern, around about 2000 Hz, a few harmonics are partially resolved. It is possible that subjects were able to use this part of the excitation pattern to maintain performance when the noise was absent or low-pass filtered at 1000 Hz, but not when it was either unfiltered or filtered at 2575 Hz. An alternative explanation is that subjects used distortion products having frequencies on or just below the low-frequency skirt of the bandpass filter.

In conclusion, the existence of deterioration with sine-phase stimuli such as those used here and previously does not depend on the use of a level rove or on the detection of distortion products at $F0$. Rather, it may be attributed to a combination of other factors. First, Hoekstra (1979) did not use a background noise when studying pulse trains filtered into high-frequency regions, and so his subjects may have been able to maintain performance by listening to the lower skirt of the excitation pattern. Second, the highest $F0$ studied by Krumbholz *et al.* (2000) was 320 Hz, lower than the lowest $F0$ at which we observed a deterioration. Third, although Cullen and Long (1986) studied an $F0$ as high as 800 Hz, the next-lowest $F0$ was 400 Hz. As can be seen in Fig. 7(b), if we had used $F0$s separated by an octave it is likely that we would have missed the deterioration observed for most of our subjects.

VII. DISCUSSION AND CONCLUSIONS

A. Limitations on temporal processing in normal listeners

1. Simulations based on ringing of a single auditory filter

Experiments 1 and 2 showed that the functions relating $F0$DL to baseline $F0$ shift rightwards as the complex tones are filtered into progressively higher frequency regions. Because these functions are nonmonotonic, an increase in frequency region can result in either an increase or decrease in $F0$DL, depending on the baseline $F0$ at which it is measured. This in turn shows that the results cannot be attributed to differences in the overall efficiency of $F0$-encoding mechanisms between sounds presented in different regions, and is consistent with the idea that auditory filter ringing impairs performance. However, in Sec. IV we argued that the variation in breakdown point with frequency region was not consistent with the impairment stemming entirely from auditory filter ringing. An additional illustration of the inability of a filter-ringing hypothesis to completely account for the breakdown is shown in the left column of Fig. 11, which

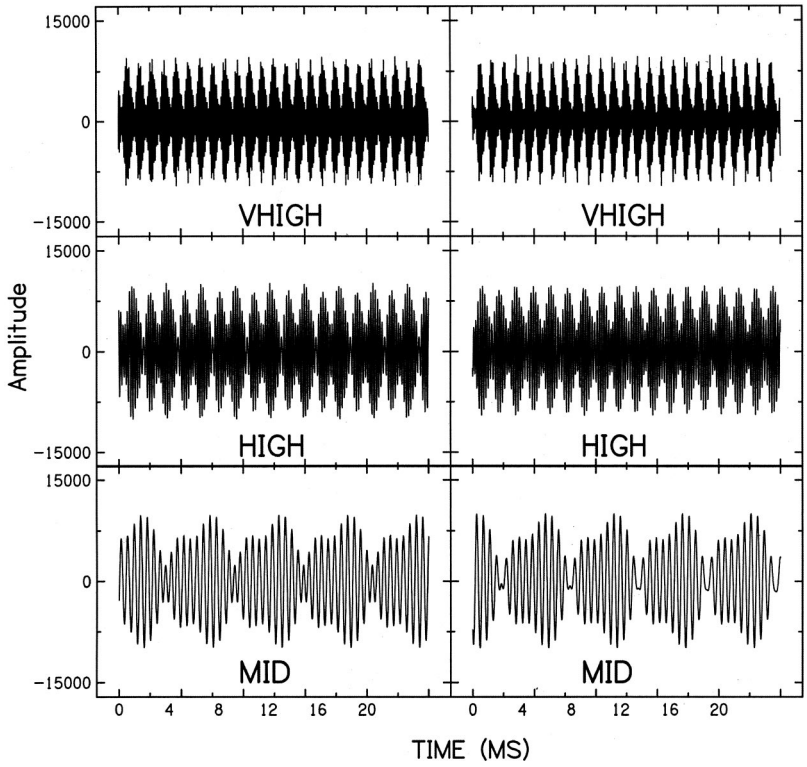


FIG. 11. The left column shows simulated outputs of a gammatone auditory filter in response to alternating-phase complexes filtered into the MID, HIGH, and VHIGH regions. The CFs of the filters were chosen to be at the midpoint of the passband in each region, on an ''ERB'' scale (Glasberg and Moore, 1990). The $F0$s of the complexes were 150 Hz in the MID region, 300 Hz in the HIGH region, and 424 Hz in the VHIGH region. The right column shows the corresponding outputs of the gammachirp filterbank described by Irino and Patterson (1997), using software provided by Irino. Parameters were $b=1.019$, $n=4$, $c=3.38-0.107$ Ps (for interpretation of these see Irino and Patterson, 1997) and the stimulus level at input was set for each complex to be the rms level used in experiment 2.

shows the outputs of ''gammatone'' auditory filters with CFs in the passbands of the MID, HIGH, and VHIGH regions. The $F0$s of the complexes were chosen to be close to the breakdown point measured in each region in experiment 2; these values were 150, 300, and 424 Hz in the MID, HIGH, and VHIGH regions, respectively. It can be seen that the filter output repeats at a rate of $2F0$ in the VHIGH and HIGH regions, and that the modulation depth is rather large—despite the fact that at these $F0$s, performance was at chance. In contrast, the output of an auditory filter tuned to the MID region shows a smaller modulation depth and repeats at a rate of $F0$, presumably because here the output is driven primarily by two components. Similar results are obtained using the ''gammachirp'' auditory filter model (Irino and Patterson, 1997), as shown in the right-hand column of Fig. 11.

To examine this idea in more detail, we compared the breakdown point in each of the three regions with the duration of filter ringing predicted from the assumption that the duration of ringing is inversely proportional to filter bandwidth. When the breakdown point was expressed in terms of $F0$/ERB and averaged across listeners, it decreased from 0.7 in the LOW region to 0.5 in the HIGH and 0.4 in the VHIGH regions. (For the MID region of experiment 2a we included only those four listeners who showed a clear breakdown in performance.) This suggests that filter ringing cannot account entirely for the breakdown in performance at high $F0$s, a conclusion supported by a further analysis of the outputs of gammatone filters centered on the HIGH and VHIGH regions of experiment 2b, in response to alternating-phase stimuli at the $F0$s used in that experiment. We observed that, in each frequency region, the amount of modulation in the filter output, as measured by the fourth central moment, decreased as a function of $F0$. Crucially, the value of the fourth moment in the HIGH region at an $F0$ of 300 Hz was equal to that in the VHIGH region at an $F0$ of 600 Hz, suggesting that, if filter ringing were entirely responsible for the deterioration, the breakdown point should have shifted by a factor of 2 between the HIGH and VHIGH regions. As noted above, this was not observed in the data.

In contrast, the $F0$ above which performance starts to improve again *is* fairly constant across frequency regions, with values of $F0$/ERB of 0.76, 0.69, and 0.65 for the MID, HIGH, and VHIGH regions, respectively. Hence, filter characteristics may completely explain the usability of place-of-excitation cues, but not the limitations on temporal processing. Finally, it is worth noting that the variations in the breakdown point across frequency region are unlikely to result from differences in the ability to maintain performance at high $F0$s by listening to the upper skirt of each complex, where the auditory filters are wider and ring for a shorter time. As shown in Table I, thresholds for pure tones half an octave above the HIGH and VHIGH regions (6490 and 19 882 Hz, respectively) were roughly similar for two out of the three listeners tested.[6]

2. Central limitations

Experiment 4 showed that a contralateral standard, by introducing a binaural cue, dramatically improved the detection of rate differences. Hence, with monaural presentation, it seems that there is a more central limitation that prevents optimal processing of the information present in the auditory nerve.

When considering what form a central limitation might take, it is important to consider three constraints imposed by the data presented here and elsewhere. First, the central limitation should be consistent with the rightward shift in the

functions relating $F0$DL to $F0$, observed as the frequency region is raised from MID to HIGH to VHIGH. If this shift at least *partly* results from variations in the ringing of auditory filters with CF, then any central limitation should not swamp the effects of such ringing. Second, any proposed temporal limitation should account for the fact that, for pure tones and resolved harmonics, subjects can use phase-locking cues at frequencies up to about 2000 Hz (see the Introduction). Third, it should allow the binaural system to detect increases in high pulse rates that are indiscriminable when all stimuli are presented monaurally.

One way in which this could occur would be if central pitch templates were less accurate at high $F0$s, perhaps because the $F0$s of most natural sounds are generally below about 500 Hz. Evidence in favor of this comes from the observation that the detection of deviations from an octave relationship, between the members of a tonal dyad, decreases markedly as the $F0$ increases above about 500 Hz (Demany *et al.*, 1991). This account would propose that, at low CFs filter ringing was the main limitation but that, as CF is raised further, allowing higher pulse rates to be represented in the peripheral auditory system, the locus of the limitation would shift gradually towards the less accurate templates that are present at high pitches. Note also that this explanation is not inconsistent with pitch mechanisms being able to use phase locking to quite high pure-tone frequencies. For example, phase locking to a 1000-Hz component might accurately be used by a central pitch mechanism when it is the fifth component of a 200-Hz $F0$, but not when it is the second harmonic of a 500-Hz $F0$. Because this proposed limitation is specific to the pitch mechanism, it would allow temporal cues to be exploited by the binaural system even when they cannot be used to obtain pitch judgments.

A second explanation, specific to groups of unresolved harmonics, is as follows. For such stimuli, all peripheral filters will be driven essentially by a series of pulses, and one would expect their responses to be approximately in synchrony. The only way in which a central mechanism could derive a pitch from this stimulus is to estimate the intervals between the neural responses to different pulses. Recent evidence, obtained both with normally hearing listeners and cochlear implant users (Kaernbach and Demany, 1998; Kaernbach and Bering, 2001; Carlyon *et al.*, 2002), suggests that subjects are sensitive only to the intervals between *adjacent* pulses in the stimulus, and not to higher-order intervals. It is likely that, as pulse rate is increased beyond a certain value, first-order intervals in the stimulus will no longer be represented as first-order *neural* intervals. For example, a 300-Hz complex whose harmonics is summed in alternating phase will contain a pulse every 1.7 ms, and refractory effects[7] make it unlikely that auditory-nerve fibers will fire on every pulse. Rather, as pulse rate is increased further, more and more of the first-order stimulus intervals will be ''missed,'' and be replaced by longer higher-order neural intervals. The result would be that pitch would no longer increase systematically with increases in pulse rate.

B. Limitations in rate processing—Cochlear implant users

The main implication of the present research for hearing by cochlear implant users is that the limitations in rate processing observed with electrical stimulation are unlikely to result entirely from a mismatch between place and rate of stimulation. This conclusion arises from the observation that, in the VHIGH region, all of our normal listeners could detect differences in pulse rate relative to a 600-pps baseline. This occurred despite the degradations in temporal envelope that will have been imposed by auditory-filter ringing, a factor which does not occur with electrical stimulation. It is also worth noting that the superior sensitivity of normal listeners to differences between high rates of stimulation, compared to that shown by cochlear implantees, is not restricted to this single condition. For example, in the HIGH region, the DL for sine-phase stimuli at an $F0$ of 300 Hz was on average only 3%. We can rule out the possibility that performance in this condition was aided by residual place-of-excitation cues, because the $F0$DL *increased* as $F0$ was raised further to 424 Hz. In contrast, although the rate DL at 300 Hz for electrical stimulation varies markedly across subjects, DLs are typically of the order of 10% or are unmeasurable, and it would be extremely unusual to observe three out of three implantees showing DLs of around 3%.

The above considerations are consistent with McKay *et al.*'s (2000) conclusion that, for a typical cochlear implant user, limitations on temporal processing are likely to have a peripheral component. However, it is also worth stressing that temporal processing of electrical pulse rate is likely to also be affected by whatever central factors limit temporal pitch processing in acoustic hearing. Hence, one would expect central processes to limit the highest rate at which a cochlear implant user can detect rate changes, albeit at higher pulse rates than the 300 pps typically encountered.

C. Summary

The results obtained with our alternating-phase complex tones reveal a sensitivity to $F0$ differences which, over a wide range of $F0$s, is mediated by temporal processing of the envelope repetition rate, rather than the use of place-of-excitation cues. The sensitivity to these rate changes in the VHIGH region and at high overall rates is superior to that seen with cochlear implant users and electric pulse trains. The fact that our listeners can detect changes relative to a pulse repetition rate of 600 pps, under conditions where the place and rate of cochlear simulation are unrelated, makes it unlikely that the inability of most implantees to detect rate changes above about 300 pps results entirely from this place-rate independence.

The envelope repetition rate at which performance breaks down for our alternating-phase complexes varies with the frequency region into which the complexes are filtered, in a manner consistent with performance being limited, *to some extent*, by auditory-filter ringing. However, this is not the only limitation and it is likely that more central processes limit performance at the highest pulse rates. Whatever the nature of these central limitations, they can be circumvented

by presenting a contralateral standard, thereby allowing subjects to perform the task using a binaural cue.

For alternating-phase stimuli, performance using the binaural cue when the envelope repetition rate is 600 pps ($F0 = 300$ Hz) is better in the HIGH than in the VHIGH region. This is consistent with previous reports of reduced sensitivity to ITDs in very high-frequency regions. This frequency dependence is opposite to that observed in the monaural version of the task, suggesting that different central limitations affect binaural mechanisms than those mechanisms responsible for monaural pitch perception. This latter conclusion is consistent with recent evidence that the temporal pitch mechanism operates in parallel to, and in competition with, binaural mechanisms, rather than receiving a direct input from them (Carlyon *et al.*, 2001).

ACKNOWLEDGMENTS

The experiments described here were originally motivated by a discussion with Colette McKay and Hugh McDermott. We are very grateful to Christopher Long, Colette McKay, and three reviewers for helpful comments on a previous version of this article. Finally, we thank Toshio Irino for providing his ''gammachirp'' filter bank software, and for advising us on its implementation.

[1]To further assess GN's sensitivity in the HIGH and VHIGH regions, we measured absolute thresholds (mean of three adaptive runs, taken from the last 12 of 16 turnpoints in each run) at the geometric centers of these two regions. The resulting absolute thresholds at 4589 and 9178 Hz were 36.5 and 74.7 dB SPL, respectively.

[2]Strictly speaking, because our stimuli were shaped with 50-ms onset and offset ramps, some small interaural time differences (ITDs) would have been present between the first *audible* pulse in each ear. However, these differences would have been considerably smaller than those present later in the stimulus.

[3]The difference is unlikely to have resulted from our switch from HD414 to HD250 headphones, which was necessitated when one earpiece of our last remaining HD414 headset broke. Although there were some differences in the frequency responses of the two headphone types, these would not have affected the sensation level of the tones, as they would have influenced the signals and background noise equally. We also checked the waveforms at the output of the two types of headset, using a B&K-type 4153 artificial ear and $\frac{1}{2}$-in. condenser microphone (model 4134), and did not observe any marked differences.

[4]Several other significant main effects and interactions were observed. Subject AO: Rove [$F(1,4) = 25.0$, $p < 0.01$], $\Delta F0$ [$F(1.6,6.3) = 9.3$, $p < 0.2$], Frequency region×rove [$F(1,4) = 11.5$, $p < 0.05$]. Subject SF: Frequency region×$\Delta F0$ [$F(3,12) = 9.1$, $p < 0.005$], Rove×$\Delta F0$ [$F(3,12) = 11.6$, $p < 0.001$], Rove×$\Delta F0$×Frequency region [$F(2.9,11.6) = 5.6$, $p < 0.05$]. Subject JD: $\Delta F0$ [$F(2.4,9.8) = 5.9$, $p < 0.05$]. Degrees of freedom have been adjusted according to the Huynh–Feldt sphericity correction.

[5]The fact that, at low $F0$s, the thresholds obtained by Cullen and Long (1986) are higher than in the other three studies may result from them requiring subjects to detect a stepwise change in $F0$ midway through a 1-s sound, rather than, as in the other studies, compare the $F0$s of two steady sounds. It is known that the detection of sinusoidal frequency modulation in the $F0$s of unresolved complex tones is poorer than the detection of steady $F0$ differences (Shackleton and Carlyon, 1994; Plack and Carlyon, 1995; Carlyon *et al.*, 2000), a finding which has been attributed to ''temporal sluggishness.'' A similar limitation may well have applied to the stepwise FM employed by Cullen and Long. This limitation is greatly reduced when resolved harmonics are present (Plack and Carlyon, 1995), which may account for the fact that the difference between Cullen and Long's data and those of other researchers is greatest at low $F0$s, where all harmonics would have been unresolved.

[6]Listener AO did show an increased threshold at the highest frequency tested, but revealed a similar pattern of results to the other two listeners in experiment 2b [Fig. 2(b)].

[7]The first author and his colleagues (Carlyon, 1997; Carlyon *et al.*, 2002) have argued that the responses from numerous auditory-nerve (AN) fibers must be combined before a temporal pitch estimate is derived. Hence, although some pulses will not be represented in the response of an individual AN fiber, they may be represented in the combined response of several fibers. The refractory effects described here may therefore reside in the output of neurons that sum the responses of many AN fibers.

ANSI (**1969**). ANSI S3.6-1969, ''Specifications for audiometers'' (American National Standards Institute, New York).

Bernstein, L. R., and Trahiotis, C. (**1994**). ''Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise,'' J. Acoust. Soc. Am. **95**, 3561–3567.

Blamey, P. J., Dowell, R. C., Tong, Y. C., and Clark, G. M. (**1984**). ''An acoustic model of a multiple-channel cochlear implant,'' J. Acoust. Soc. Am. **76**, 97–103.

Carlyon, R. P. (**1996**). ''Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker,'' J. Acoust. Soc. Am. **99**, 517–524.

Carlyon, R. P. (**1997**). ''The effects of two temporal cues on pitch judgments,'' J. Acoust. Soc. Am. **102**, 1097–1105.

Carlyon, R. P. (**1998**). ''The effects of resolvability on the encoding of fundamental frequency by the auditory system,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 246–254.

Carlyon, R. P., Demany, L., and Deeks, J. (**2001**). ''Temporal pitch perception and the binaural system,'' J. Acoust. Soc. Am. **109**, 686–700.

Carlyon, R. P., Moore, B. C. J., and Micheyl, C. (**2000**). ''The effect of modulation rate on the detection of frequency modulation and mistuning of complex tones,'' J. Acoust. Soc. Am. **108**, 304–315.

Carlyon, R. P., van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). ''Temporal pitch mechanisms in acoustic and electric hearing,'' J. Acoust. Soc. Am. (in press).

Cohen, M. A., Grossberg, S., and Wyse, L. L. (**1995**). ''A spectral network model of pitch perception,'' J. Acoust. Soc. Am. **98**, 862–879.

Cullen, J. K., and Long, G. R. (**1986**). ''Rate discrimination of high-pass filtered pulse trains,'' J. Acoust. Soc. Am. **79**, 114–119.

Demany, L., Semal, C., and Carlyon, R. P. (**1991**). ''Perceptual limits of octave harmony and their origin,'' J. Acoust. Soc. Am. **90**, 3019–3027.

Formby, C., and Muir, K. (**1988**). ''Modulation and gap detection for broadband and filtered signals,'' J. Acoust. Soc. Am. **84**, 545–550.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (2000). ''Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency,'' J. Acoust. Soc. Am. **108**, 263–271.

Hafter, E. F., and Dye, R. H. (**1983**). ''Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number,'' J. Acoust. Soc. Am. **73**, 644–651.

Hanna, T. E. (**1992**). ''Discrimination and identification of modulation rate using a noise carrier,'' J. Acoust. Soc. Am. **91**, 2122–2128.

Henning, G. B. (**1966**). ''Frequency discrimination of random-amplitude tones,'' J. Acoust. Soc. Am. **39**, 336–339.

Hoekstra, A. (**1979**). ''Frequency discrimination and frequency analysis in hearing,'' PhD thesis, Institute of Audiology, University Hospital, Groningen, The Netherlands.

Houtgast, T. (**1989**). ''Frequency selectivity in amplitude-modulation detection,'' J. Acoust. Soc. Am. **85**, 1676–1680.

Irino, T., and Patterson, R. D. (**1997**). ''A time-domain, level-dependent auditory filter: The gammachirp,'' J. Acoust. Soc. Am. **101**, 412–419.

Irvine, D. R. F. (**1992**). ''Physiology of the auditory brainstem,'' in *Springer Series in Auditory Research Vol. II. The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 153–231.

Kaernbach, C., and Bering, C. (**2001**). ''Exploring the temporal mechanisms involved in the pitch of unresolved harmonics,'' J. Acoust. Soc. Am. **110**, 1039–1047.

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

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). ''The lower limit of pitch as determined by rate discrimination,'' J. Acoust. Soc. Am. **108**, 1170–1180.
Lawson, D., Zerbi, M., and Wilson, B. (**1999**). ''Speech Processors for Auditory Prostheses'' (First Quarterly Progress Report), NIH Project No N01-DC-8-2105.
Levitt, H. (**1971**). ''Transformed up–down methods in psychophysics,'' J. Acoust. Soc. Am. **49**, 467–477.
Loeb, G. E., White, M. W., and Merzenich, M. M. (**1983**). ''Spatial cross-correlation,'' Biol. Cybern. **47**, 149–163.
Long, C. J., Eddington, D. K., Colburn, H. S., Rabinowitz, W. M., Whearty, M. E., and Kadel-Garcia, N. (**1998**). ''Speech Processors for Auditory Prostheses'' (11th Quarterly Progress Report), NIH Project No N01-DC-6-2100.
McKay, C. M., McDermott, H. J., and Carlyon, R. P. (**2000**). ''Place and temporal cues in pitch perception: Are they truly independent?,'' ARLO (http://ojps.aip.org/ARLO/top.html) **1**, 25–30.
Micheyl, C., and Carlyon, R. P. (**1998**). ''Effects of temporal fringes on fundamental-frequency discrimination,'' J. Acoust. Soc. Am. **104**, 3006–3018.
Micheyl, C., Moore, B. C. J., and Carlyon, R. P. (**1998**). ''The role of excitation-pattern cues and temporal cues in the frequency and modulation-rate discrimination of amplitude-modulated tones,'' J. Acoust. Soc. Am. **104**, 1039–1050.
Moore, B. C. J. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.
Moore, B. C. J., and Peters, R. W. (**1992**). ''Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity,'' J. Acoust. Soc. Am. **91**, 2881–2893.
Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). ''Spiral Vos Final Report, Part A: The Auditory Filterbank,'' Applied Psychology Unit, Cambridge, England. Contract Report: APU 2341.
Plack, C. J., and Carlyon, R. P. (**1995**). ''Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics,'' J. Acoust. Soc. Am. **98**, 1355–1364.
Pressnitzer, D., and Patterson, R. D. (**2001**). ''Distortion products and the pitch of harmonic complex tones,'' in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht), pp. 97–104.
Sachs, M. B., Voigt, H. F., and Young, E. D. (**1983**). ''Auditory nerve representation of vowels in background noise,'' J. Neurophysiol. **50**, 27–45.
Sek, A., and Moore, B. C. J. (**1995**). ''Frequency discrimination as a function of frequency, measured in several ways,'' J. Acoust. Soc. Am. **97**, 2479–2486.
Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.
Shannon, R. V. (**1983a**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.
Shannon, R. V. (**1983b**). ''Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction,'' Hear. Res. **12**, 1–16.
Shannon, R. V. (**1992**). ''Temporal modulation transfer functions in patients with cochlear implants,'' J. Acoust. Soc. Am. **91**, 2156–2164.
Shepherd, R. K., and Javel, E. (**1997**). ''Electric stimulation of the auditory nerve. I. Correlation of physiological responses with cochlear status,'' Hear. Res. **108**, 112–144.
Tong, Y. C., and Clark, G. M. (**1985**). ''Absolute identification of electric pulse rates and electrode positions by cochlear implant listeners,'' J. Acoust. Soc. Am. **74**, 73–80.
Townshend, B., Cotter, N., Compernolle, D. v., and White, R. L. (**1987**). ''Pitch perception by cochlear implant subjects,'' J. Acoust. Soc. Am. **82**, 106–115.
Tyler, R. S., Wood, E. J., and Fernandes, M. A. (**1983**). ''Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **74**, 1190–1199.
van den Honert, C., and Stypulkowski, P. H. (**1987**). ''Temporal response patterns of single auditory-nerve fibers elicited by periodic electrical stimuli,'' Hear. Res. **29**, 207–222.
van Hoesel, R. J. M., and Clark, G. M. (**1997**). ''Psychophysical studies with two binaural cochlear implant subjects,'' J. Acoust. Soc. Am. **102**, 495–507.
Wiegrebe, L., and Patterson, R. D. (**1999**). ''Quantifying the distortion products generated by SAM noise,'' J. Acoust. Soc. Am. **106**, 2709–2718.

Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli"[a)]

Leslie R. Bernstein
Department of Neuroscience, University of Connecticut Health Center, Farmington, Connecticut 06030

Constantine Trahiotis
Department of Neuroscience and Department of Surgery (Otolaryngology), University of Connecticut Health Center, Farmington, Connecticut 06030

(Received 28 February 2002; revised 24 May 2002; accepted 10 June 2002)

It is well-known that thresholds for ongoing interaural temporal disparities (ITDs) at high frequencies are larger than threshold ITDs obtained at low frequencies. These differences could reflect true differences in the binaural mechanisms that mediate performance. Alternatively, as suggested by Colburn and Esquissaud [J. Acoust. Soc. Am. Suppl. 1 **59**, S23 (1976)], they could reflect differences in the peripheral processing of the stimuli. In order to investigate this issue, threshold ITDs were measured using three types of stimuli: (1) low-frequency pure tones; (2) 100% sinusoidally amplitude-modulated (SAM) high-frequency tones, and (3) special, ''transposed'' high-frequency stimuli whose envelopes were designed to provide the high-frequency channels with information similar to that available in low-frequency channels. The data and their interpretation can be characterized by two general statements. First, threshold ITDs obtained with the transposed stimuli were generally smaller than those obtained with SAM tones and, at modulation frequencies of 128 and 64 Hz, were equal to or smaller than threshold ITDs obtained with their low-frequency pure-tone counterparts. Second, quantitative analyses revealed that the data could be well accounted for via a model based on normalized interaural correlations computed subsequent to known stages of peripheral auditory processing augmented by low-pass filtering of the envelopes within the high-frequency channels of each ear. The data and the results of the quantitative analyses appear to be consistent with the general ideas comprising Colburn and Esquissaud's hypothesis. [DOI: 10.1121/1.1497620]

PACS numbers: 43.66.Pn, 43.66.Ba [MRL]

I. INTRODUCTION

The ability to discriminate changes in ongoing interaural temporal disparities (ITDs) can be much poorer when the information is conveyed by high-frequency stimuli, as compared to when it is conveyed by low-frequency stimuli (e.g., Klumpp and Eady, 1956; Zwislocki and Feldman, 1956; McFadden and Pasanen, 1976; Nuetzel and Hafter, 1976; Henning, 1980; Bernstein and Trahiotis, 1982, 1994; Blauert, 1983). In addition, and logically consistent with those results, it has been found that functions relating extent of laterality to ITD measured with high-frequency stimuli are typically more shallow than those measured with low-frequency stimuli. That is, for a given ITD, intracranial images produced by high-frequency stimuli are perceived to be much closer to the midline than are intracranial images produced by low-frequency stimuli. (e.g., Blauert, 1982; Bernstein and Trahiotis, 1985).

One logical possibility is that these differences in the relative potency of ITDs result primarily from differences between the (central) *binaural* mechanisms that mediate interaural interactions in low- and high-frequency regions, respectively. Another possibility is that the observations reflect inherent frequency-related differences in the neural information that serves as input to the binaural portion of the auditory system. This latter possibility was favored by Colburn and Esquissaud (1976). They suggested that frequency-related differences in sensitivity to ongoing ITDs could result from the rectification and low-pass filtering that occurs as a natural part of monaural, peripheral processing. For low-frequency stimuli, such processing would result in neural impulses synchronized to the whole *waveform* (i.e., both the fine-structure *and* the envelope). For high-frequency stimuli, such processing would result in neural impulses synchronized to only the *envelope* of the waveform. An important assumption made by Colburn and Esquissaud was that the binaural (cross-correlation) mechanism that receives the two types of synchronized neural impulses operates uniformly across frequency.

We recently published data and analyses that we believe strongly support Colburn and Esquissaud's (1976) contention that the binaural comparator functions uniformly across frequency (Bernstein and Trahiotis, 1996b). Utilizing a NoSo vs NoSπ discrimination task, we found that binaural detection measured as a function of the center frequency of the stimuli could be accounted for by utilizing normalized interaural correlations computed subsequent to rectification and low-pass filtering. This type of model provides, as a function of frequency, the types of inputs Colburn and Esquissaud postulated would naturally occur for binaural comparison.

The purpose of this paper is to report the results of new

[a)]A portion of this work was presented at the 141st meeting of the Acoustical Society of America, 2001, Chicago, Illinois.

experiments that lend additional general support to Colburn and Esquissaud's (1976) thesis. The experiments were designed with the goal of providing the high-frequency channels of the binaural processor with envelope-based inputs that, other things being equal, would essentially mimic waveform-based inputs normally available in the low-frequency channels. Such stimuli were generated by capitalizing on the ''transposition'' technique described by van de Par and Kohlrausch (1997).

We measured sensitivity to changes in ITDs for high-frequency ''transposed'' stimuli and compared those thresholds to thresholds measured with low-frequency tones and to thresholds measured with high-frequency tones that were sinusoidally amplitude-modulated (SAM). It will be seen that the high-frequency ''transposed'' stimuli yielded threshold ITDs that were substantially smaller than those obtained with high-frequency SAM tones and which, for low rates of modulation, were as small or smaller than threshold ITDs measured with low-frequency pure tones. In addition, it will be seen that the data can be accounted for via normalized interaural correlations computed subsequent to transformations that reflect known stages of peripheral auditory processing with the proviso that the envelopes within the high-frequency channels are subjected to low-pass filtering at 150 Hz. The data and their analysis appear to be consistent with Colburn and Esquissaud's (1976) general idea that differences in the inputs to the binaural processor between the low-frequency and high-frequency portions of the auditory system are primary determiners of sensitivity to ITD.

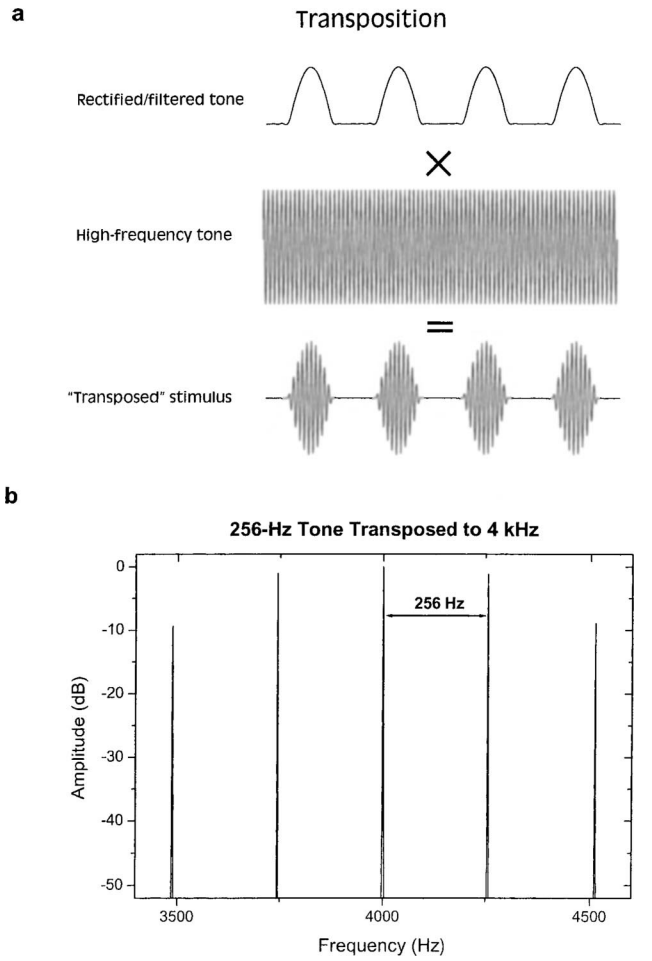


FIG. 1. Panel (a) Schematic representation of the method used to generate transposed stimuli. Panel (b) Power spectrum of a 256-Hz tone transposed to 4 kHz (see the text).

II. EXPERIMENT 1

A. Stimulus generation

Low-frequency sinusoids and high-frequency SAM tones were generated digitally in the frequency domain. The high-frequency transposed stimuli were generated employing a technique similar to that described by van der Par and Kohlrausch (1997). The general technique is illustrated in Fig. 1(a). First, the time-domain representation of a low-frequency waveform was (linearly) half-wave rectified by setting all negative values to zero. The rectified waveform was then transformed to the frequency domain and the magnitudes of components above 2 kHz were filtered out by setting them to zero. Then, the signal that resulted was transformed back to the time domain (top row) and multiplied by a high-frequency sinusoidal carrier having the desired center frequency of the transposed stimulus (middle row). The final product (bottom row) was the transposed stimulus having an envelope whose time signature mimicked that of the rectified and filtered pure tone.

Figure 1(b) displays the power spectrum of one of the transposed stimuli used in the experiment. In this case, a pure tone having a frequency of 256 Hz was transposed to 4 kHz. Like all of the transposed stimuli employed, the spectral components are symmetric and limited to ± 2 kHz around the center frequency. For this example, the technique results in the presence of sidebands at 4000 ± 256, ± 512, ± 1024, and ± 1536 Hz. Were no rectification applied to the 256-Hz tone, only the two sidebands at 4000 ± 256 Hz would be present.

It is important to understand why linear half-wave rectification followed by spectral limiting at 2 kHz was employed. A high-frequency transposed stimulus would, like any other signal, be subjected to *internal* rectification and low-pass filtering by the listener's auditory system. As a result, the internal representation of the transposed stimulus would be expected to reflect the sequential effects of the external and (perhaps, nonlinear) internal rectification and the sequential effects of external and internal low-pass filtering. Linear rectification and low-pass filtering at 2 kHz was employed because their effects would be essentially transparent when followed by internal (linear or nonlinear) rectification and the internal low-pass filtering that characterizes the neural synchrony to stimulus waveforms. This argument implicitly assumes (1) that the ''rectification'' that occurs in the peripheral auditory system removes all, or essentially all, of the negative portions of the external waveform and (2) that the cutoff of the internal low-pass filtering is substantially below 2 kHz. Numerous physiological data and analyses appear to support both assumptions (e.g., Rose *et al.*, 1967; Brugge *et al.*, 1969; Johnson, 1980; Palmer and Russell, 1986).

There are two other lines of evidence that attest to the

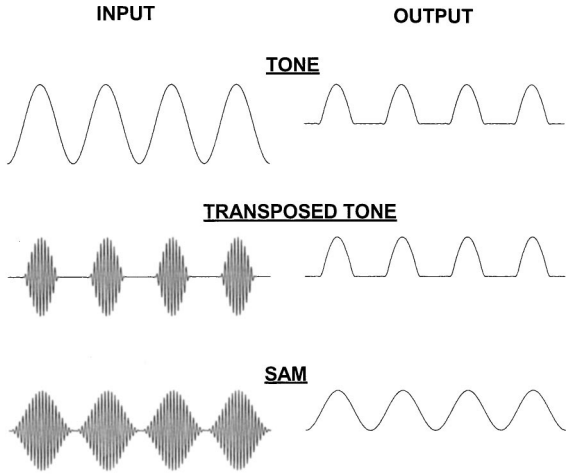


FIG. 2. Left side: A 250-Hz tone (upper), a 250-Hz tone transposed to 4 kHz (middle), and a 4-kHz tone sinusoidally amplitude modulated at 250 Hz (lower). Right side: The same three stimuli subsequent to bandpass filtering, rectification, and low-pass filtering.

suitability of the overall procedure. First, considering the stimuli themselves, we verified via computer simulations that employing a low-pass cutoff of 2 kHz had negligible effects on the envelopes of the transposed stimuli in that those envelopes differed minimally from the half-wave rectified tones used to generate them. Thus, it appears that the procedure yields physical stimuli, *per se*, that fulfill our requirements. Second, van de Par and Kohlrausch (1997) have recently shown that restricting the spectra of transposed stimuli in a similar manner such that only three or five central components remain did not adversely affect improvements in binaural detection thresholds.

B. Procedure

Detection of ongoing ITD was measured using three types of stimuli: (1) low-frequency pure tones; (2) low-frequency tones transposed to 4 kHz; (3) 100% sinusoidally amplitude-modulated (SAM) tones centered at 4 kHz. The frequencies of the pure tones and the rates of modulation of the SAM and transposed stimuli were either 32, 64, 128, 256, or 512 Hz.

Figure 2 provides an illustration of the idealized case for three types of stimulus waveforms when the frequency of the pure tone and the frequency of modulation were each 250 Hz. The waveforms are shown both at the input and at the output of putative peripheral processing (rectification and low-pass filtering that results in extraction of the envelope at high frequencies).

The figure illustrates that, for the low-frequency 250-Hz tone (top row), the effect of peripheral processing is to pass only the positive values of the waveform. That is, the waveform has been half-wave rectified. For the transposed stimulus (middle row) and for the SAM tone (bottom row), the effect of peripheral processing is to extract the envelope of the waveform. The fine-structure at 4000 Hz is removed because low-pass filtering smooths over oscillations at this frequency. Note that the pure tone (top row) and the transposed stimulus (middle row) result in output waveforms that are essentially identical half-wave rectified sinusoids. For these two stimuli, peripheral processing results in outputs characterized by distinct ''off'' regions between the ''peaks'' during which the waveform remains at or close to a value of zero. In contrast, note that the corresponding output for the SAM tone (bottom row) is an unrectified sinusoid and has no such distinct ''off'' regions. It seems reasonable to assume that period histograms of neural discharges created by the transduction of low-frequency tones and transposed tones would be relatively less dispersed in time (have a smaller variance) than period histograms for SAM tones.

To the degree that such greater neural synchrony results in smaller threshold ITDs, one would expect that, for a given pure tone or modulation frequency, threshold ITDs obtained with a high-frequency transposed stimulus would be smaller than those obtained with its high-frequency SAM tone counterpart and, ideally, be equivalent to the threshold ITDs obtained with the low-frequency pure tone. The word ''ideally'' is used because close correspondence between the outputs in low-frequency and high-frequency regions and between their respective neural inputs to the binaural processor may not always be expected to occur within the auditory system. Exceptions could occur that stem from the effects of peripheral bandpass filtering and from a ''rate limitation'' that degrades the processing of high rates of fluctuation of the envelopes of high-frequency stimuli. Both of these factors will be discussed in context when the data are presented. To the degree that these two factors play a role, one would not expect threshold ITDs obtained with transposed stimuli to be as small as those obtained with their low-frequency pure-tone counterparts.

All three types of stimuli were generated digitally with a sampling rate of 20 kHz (TDT AP2), were low-pass filtered at 8.5 kHz (TDT FLT2), and were presented via Etymotic ER-2 insert earphones at a level matching 75 dB SPL as produced by TDH-39 earphones in a 6-cc coupler.[1] The duration of each stimulus was 300 ms including 20-ms $\cos^2$ rise–decay ramps. For the high-frequency stimuli, a continuous diotic noise low-pass filtered at 1300 Hz (No equivalent to 30 dB SPL) was presented to preclude the listeners' use of any information at low spectral frequencies (e.g., Nuetzel and Hafter, 1976, 1981; Bernstein and Trahiotis, 1994).

Threshold ITDs were determined using a two-cue, two-alternative, forced choice, adaptive task. Each trial consisted of a warning interval (500 ms) and four 300-ms observation intervals separated by 400 ms. Each interval was marked visually by a computer monitor. Feedback was provided for approximately 400 ms after the listener responded. The stimuli in the first and fourth intervals were diotic. The listener's task was to detect the presence of an ITD (left-ear leading) that was presented with equal *a priori* probability in either the second or the third interval. The remaining interval, like the first and fourth intervals, contained diotic stimuli.

For the low-frequency tones and the high-frequency SAM stimuli, the starting phase of the components comprising each stimulus (prior to the imposition of an ITD) was chosen randomly for each observation interval within and across trials. All of the waveforms required for a given trial were computed immediately prior to that trial. Because of

the time required to generate the high-frequency transposed stimuli, it was necessary to calculate the transposed waveforms prior to each adaptive run. Twenty independently calculated tokens of the desired type of transposed stimulus were stored and one of them was chosen, with replacement, for each observation interval within each trial. Twenty tokens were used to ensure that the results were not dependent upon any particular stimulus. This number of tokens was considered to be sufficiently large based on Siegel and Colburn's (1989) findings that only ten independently generated tokens of noise yielded essentially equivalent performance to that measured with "running" noise in a binaural discrimination task.

For all three types of stimuli, ongoing ITDs were imposed by applying linear phase shifts to the representation of the signals in the frequency domain and then gating the signals destined for the left and right ears coincidentally, after transformation to the time domain. The ITD for a particular trial was determined adaptively in order to estimate 70.7% correct (Levitt, 1971). The initial step size for the adaptive track corresponded to a factor of 1.584 (equivalent to a 2-dB change of ITD) and was reduced to a factor of 1.122 (equivalent to a 0.5-dB change of ITD) after two reversals. A run was terminated after 12 reversals and threshold was defined as the geometric mean of the ITD across the last ten reversals.

Four normal-hearing adults served as listeners and three consecutive thresholds were first obtained from each listener for each of the particular stimulus conditions (type of stimulus$\times$frequency), which were chosen in random order. Then, three more thresholds were obtained by revisiting the same stimulus conditions in reverse order. The same ordering of conditions was used for all listeners and all listeners received substantial practice before formal collection of data began. For each listener and stimulus condition, final estimates of thresholds were calculated by averaging the individual thresholds obtained from six adaptive runs.

C. Results and discussion

Figure 3 displays the mean threshold ITDs computed across the four listeners for the three types of stimuli. The thresholds are plotted as a function of either the frequency of the pure tone or the frequency of modulation of the high-frequency SAM and transposed stimuli. When we refer to the frequency of modulation of a transposed stimulus, we refer to the frequency of the pure tone that was used to generate it. The parameter within the plot is the type of stimulus that conveyed the ITD, and the error bars represent ± 1 standard error of the mean. Note that, as signified by the "broken" ordinate and "broken" lines through the data, no values of threshold ITD are plotted for SAM and transposed stimuli having rates of modulation of 512 Hz. This is so because, for two of the listeners, thresholds ITDs could not be determined even with ITDs of up to 1 ms. Therefore, no valid mean threshold ITD could be calculated.

Beginning with the SAM tones (squares), Fig. 3 indicates that threshold ITDs are in the range of 130 to 260 μs and are smallest for the intermediate frequencies of modulation. These values of threshold ITD are consistent with those

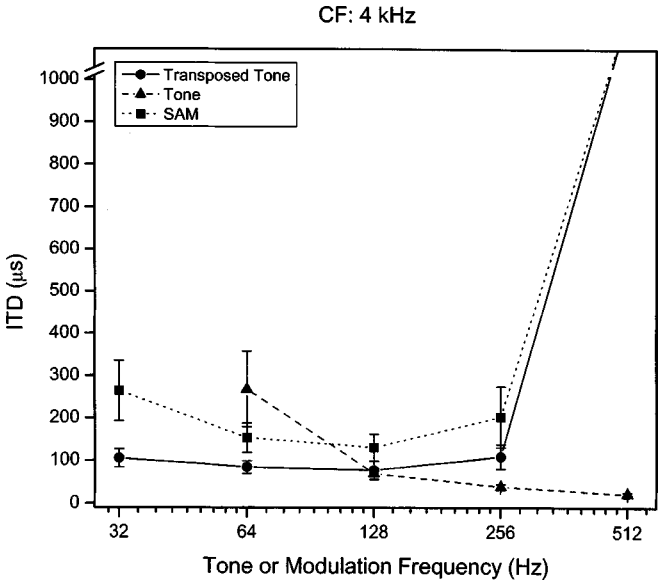


FIG. 3. Threshold ITDs averaged across the four listeners as a function of the modulation or pure-tone frequency. The center frequency of the high-frequency SAM and transposed stimuli was 4 kHz. The parameter of the plot is the type of stimulus employed. The error bars represent $\pm$ standard error of the mean. The "broken" ordinate and "broken" lines through the data indicate conditions for which average threshold ITDs could not be computed because, for a subset of the listeners, thresholds could not be determined even for ITDs of up to 1 ms.

obtained in earlier investigations with similar stimuli (e.g., Henning, 1974; Nuetzel and Hafter, 1981; Bernstein and Trahiotis, 1994)[2] and, therefore, provide a valid basis for comparison. Threshold ITDs obtained with the transposed stimuli (circles) are consistently and substantially smaller than those obtained with the SAM tones (squares). This outcome is in line with our arguments concerning the peripheral processing of the stimuli. Specifically, the threshold ITDs measured with the transposed stimuli are roughly half those measured with the SAM stimuli.

The threshold ITDs obtained with the pure tones at 128, 256, and 512 Hz (triangles) are very similar to those obtained in previous studies (e.g., Klumpp and Eady, 1956; Zwislocki and Feldman, 1956). In addition, as observed in those studies, threshold ITDs declined as frequency was increased toward 512 Hz. The relatively large mean threshold ITD of 268 μs and relatively large standard error of 89 μs at 64 Hz occurred because the threshold obtained from one of the listeners (JB) was much larger (533 μs) than those obtained from the other three listeners. Calculating the mean threshold after excluding JB's data reduced the mean threshold ITD at 64 Hz to 180 μs and the standard error to 17μs. Those values are in line with those obtained at the higher tonal frequencies.

In an effort to determine whether JB's relatively high threshold resulted from the relatively lower sensation level of the 64-Hz tone, as compared to the higher frequency tones that were presented at the same sound-pressure level, additional measures of threshold were obtained after increasing the level of the 64-Hz tone by 10 dB. This reduced JB's threshold ITD to 163 μs while having very little, if any, effect on the threshold ITDs obtained from the other three

listeners. On the basis of these findings, we consider the recalculated threshold of 180μs as being more indicative of the average listener's ability to resolve ITDs at 64 Hz.

Comparisons among threshold ITDs obtained with low-frequency pure tones (triangles) and their transposed counterparts (circles) indicate that sensitivity to ITD in high-frequency channels of the auditory system can, for stimuli having low rates of modulation, be as good as or even better than that measured in low-frequency channels. Specifically, at 128 Hz, threshold ITDs for transposed and tonal stimuli are essentially equivalent, being 76 and 69 μs, respectively. At 64 Hz, the threshold ITD obtained with the transposed stimulus (95 μs) is actually *smaller* than that obtained with the pure tone, independent of whether one uses the plotted mean threshold ITD (268 μs) or the recalculated mean (180 μs) to represent threshold for the pure-tone condition.

The data obtained at higher rates of modulation, 256 and 512 Hz, however, indicate that threshold ITDs obtained with transposed stimuli are larger than their pure-tone counterparts. The mean threshold ITD obtained with the pure tone of 256 Hz is smaller than that obtained with the transposed stimulus and, while listeners were quite sensitive to ITDs conveyed by a pure tone of 512 Hz, they were quite insensitive to ITDs conveyed by its transposed counterpart. For two different reasons, this outcome was not surprising. First, both SAM tones and transposed stimuli contain ''sidebands'' that would be subjected to increasing amounts of attenuation via peripheral filtering as the rate of modulation (and thus the separation in frequency between the sidebands) is increased such that the sidebands fall within the ''skirt'' of the filter. Nuetzel and Hafter (1981) specifically discussed how peripheral filtering would lead to attenuation of the sidebands of SAM tones and how that attenuation would result in reductions in depth of modulation which, in turn, could lead to degradations in sensitivity to ITD. Bernstein and Trahiotis (1996a) showed how reductions in depth of modulation result in poorer ITD thresholds by considering how changes in depth of modulation affect the normalized interaural correlation. More recently, van der Par and Kohlrausch (1997) also considered how peripheral attenuation of the sidebands of high-frequency transposed stimuli could degrade binaural detection in an MLD task.

The second reason this outcome was expected is that there appears to exist a limitation in the ability of the auditory system to follow rates of fluctuation of the envelope that are greater than about 150 Hz. Data supporting the existence of such a limitation have been reported in several binaural investigations (e.g., McFadden and Pasanen, 1976; Bernstein and Trahiotis, 1992a, 1992b, 1994) and, as discussed in the latter three of those studies, the process limiting the ability to follow rapidly changing envelopes appears to operate independently of peripheral bandpass filtering. It is interesting, historically, that Nuetzel and Hafter (1981), who favored an explanation based solely on peripheral filtering, acknowledged the logical possibility that such a rate limitation could have affected the ITD thresholds they measured using high-frequency SAM tones.

Additional empirical evidence that an envelope rate limitation is manifest at high spectral frequencies has been provided by Kohlrausch *et al.* (2000) and Ewert and Dau (2000). In both of those studies, temporal modulation transfer functions (TMTFs) were measured at various center frequencies. The patterning of the data and their quantitative analyses led them to include a low-pass filter in their model that serves to attenuate, independent of the center frequency of the stimulus, fluctuations of the envelope that are more rapid than 150 Hz. The common inferences from these studies and from a more recent study by Moore and Glasberg (2001) are that (1) there appears to be a monaural process that functionally acts as a low-pass filter on the envelope-based information that serves as input to more central stages of processing and (2) the low-pass filtering of the envelope appears to be functionally independent of the center frequency of the stimuli and, by necessity, independent of the width of initial peripheral bandpass filtering.

In summary, threshold ITDs obtained with high-frequency transposed stimuli: (1) are consistently smaller than those obtained with high-frequency SAM tones and (2) at frequencies of modulation of 128 and 64 Hz, are as small or smaller than threshold ITDs obtained with low-frequency pure tones. In our view, these findings are consistent with Colburn and Esquissaud's (1976) general hypothesis that transformations affecting the inputs to the binaural processor are responsible for the finding that threshold ITDs obtained at high frequencies are typically larger than those obtained at low frequencies.

The reader is reminded that two of the four listeners were essentially unable to perform the task with SAM and transposed stimuli having a rate of modulation of 512 Hz. This outcome motivated us to determine whether there were consistent inter-individual differences in relative sensitivity to ITD across the frequencies of modulation tested. In order to do so, the data from each listener were normalized by dividing the threshold ITD in each condition by the listener's threshold ITD measured with the SAM tone having a rate of modulation of 128 Hz. That stimulus was chosen as the ''standard'' for comparison because rates of modulation close to that frequency have been shown in several studies to yield relatively small threshold ITDs (e.g., Henning, 1974; Nuetzel and Hafter, 1981; Bernstein and Trahiotis, 1994). Normalizing the data in this manner permits one to make useful comparisons of relative performance within and across individual listeners, even when there are differences in the types of stimuli employed and in absolute sensitivity to ITD.

Figure 4 contains the normalized thresholds for data obtained with the SAM tones and the transposed stimuli plotted as bar graphs. The data are grouped by modulation frequency so that within- and across-listener trends in the data can be easily discerned. The horizontal dotted line at a value of 1.0 represents, for each listener, the threshold ITD obtained with the 128-Hz SAM reference stimulus. For all listeners, for rates of modulation below 512 Hz, threshold ITDs obtained with the transposed stimuli (filled bars) are smaller than those obtained with the SAM tone (unfilled bars). The only exception occurred for our most sensitive listener, AC, at a frequency of modulation of 256 Hz. Her un-normalized threshold ITDs for the SAM and transposed stimuli were 70 and 77 μs, respectively, indicating excellent sensitivity to

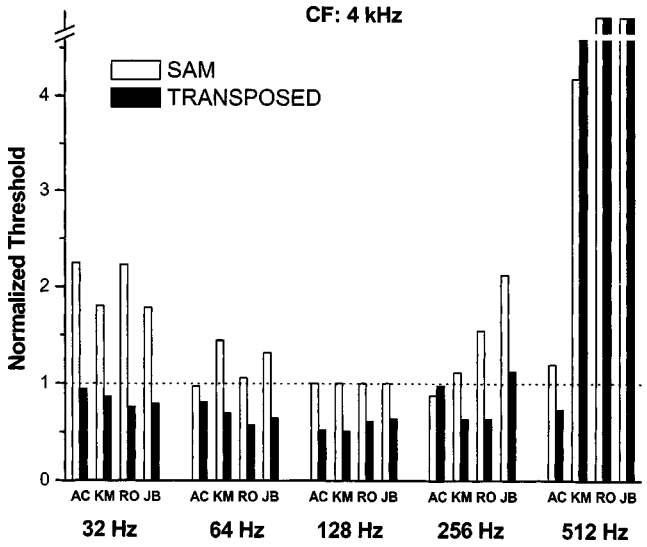


FIG. 4. Normalized threshold ITDs for the SAM (open bars) and transposed stimuli (filled bars) centered at 4 kHz. The data from each listener were normalized by dividing the threshold ITD in each condition by the listener's threshold ITD measured with the SAM tone having a rate of modulation of 128 Hz. The data for the four individual listeners are grouped by modulation frequency. The broken ordinate and broken bars indicate conditions in which thresholds could not be determined even for ITDs as large as 1 ms.

ITDs conveyed by both types of stimuli. The patterning of the normalized data clearly indicates that for both SAM tones and transposed stimuli there were essentially no inter-individual differences in relative sensitivity to ITD for rates of modulation of 32, 64, and to 128 Hz. In contrast, at 256 Hz, the heights of the bars reflect moderate inter-individual differences and at 512 Hz, there are large inter-individual differences. At the latter frequency, one of the listeners (AC) performed as well as for the lower rates, one of the listeners (KM) required approximately four to five times the ITD required at 128 Hz, and two of the listeners (RO and JB) could not perform the task given repeated attempts with ITDs of up to 1 ms.

Our interpretation of the relations among the data in Fig. 4 is that the data obtained from each individual confirm the representative nature of the averaged threshold ITDs depicted in Fig. 3. In addition, we believe that the inter-individual differences that did occur most likely did not stem from inter-individual differences in the ability to process ITDs, *per se*.

III. EXPERIMENT 2

Following the collection of data with stimuli centered at 4 kHz, data were obtained with SAM tones and transposed stimuli centered at either 6 or 10 kHz in order to assess the generalizability of the findings. At the higher frequencies, the widths of the peripheral filters are greater than at 4 kHz, and any reduction in the depth of modulation that occurred at 4 kHz would be expected to be less severe or absent at 6 and 10 kHz. Therefore, to the degree that reductions of the depth of modulation were responsible for loss of sensitivity to ITD for high rates of modulation at 4 kHz, one would expect performance to be *improved* by increasing the center frequency to 6 or 10 kHz. On the other hand, recent experiments have indicated that sensitivity to ITDs conveyed by the envelopes of conventional high-frequency stimuli (e.g., SAM tones) having center frequencies higher than 4 kHz can be much *poorer* than that observed at 4 kHz. (e.g., Henning, 1974; Bernstein and Trahiotis, 1994). A similar finding with transposed stimuli could indicate that sensitivity to ITD, *per se*, decreases at the higher frequencies as a result of unknown factors probably not associated with the peripheral processing of the stimuli.

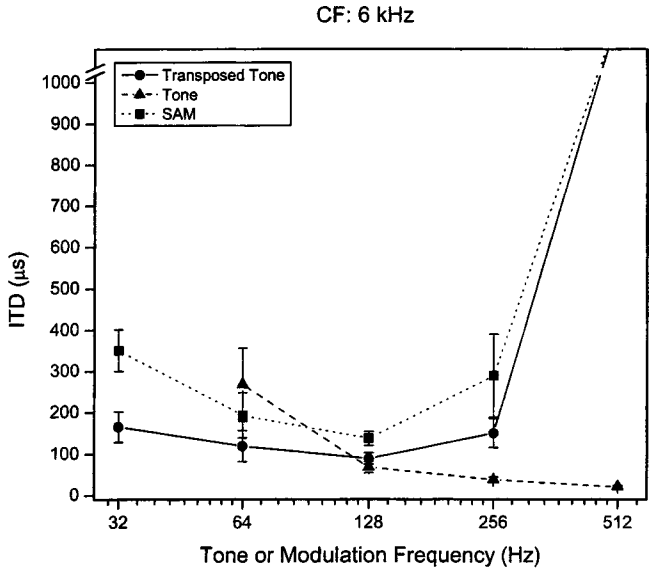
The procedures used to obtain threshold ITDs at 6 and 10 kHz were the same as those described for experiment 1, save for the fact that the stimuli were generated with a suitably higher sampling rate (27.056 kHz) and an increased cutoff frequency of the low-pass, anti-imaging filter (12.75 kHz). The listeners were the four who participated in experiment 1. All of the data with stimuli centered at 6 kHz were collected prior to collecting the data with stimuli centered at 10 kHz. It was judged that ''blocking'' the conditions in this manner would give the listeners the greatest opportunity to achieve their best performance in what was expected to be a difficult task.

A. Results and discussion

The top and bottom panels of Fig. 5 display the mean threshold ITDs for the stimuli centered at 6 and 10 kHz, respectively. The threshold ITDs obtained with pure tones are replotted from Fig. 1. The thresholds obtained at 6 kHz are slightly, but consistently, larger than those obtained at 4 kHz and the overall patterning of the data at the two center frequencies is virtually identical. Once again, the threshold ITDs obtained with the transposed stimuli are smaller than those obtained with the SAM tones. Note that, as was the case at 4 kHz, the threshold ITD obtained with the transposed stimulus having a rate of modulation of 128 Hz is, for practical purposes, equivalent to that obtained with a 128-Hz pure tone, and the one obtained with a rate of modulation of 64 Hz is smaller than that obtained with its pure-tone counterpart.

The data obtained at 10 kHz are somewhat different in that threshold ITDs are generally larger, being, when measurable, two to three times those obtained at 4 kHz. As indicated in the figure, mean threshold ITDs could not be computed at 256 and 512 Hz. This occurred because some listeners could not perform the task at these rates of modulation even with ITDs as large as 1 ms. A comparable loss of sensitivity to ITD at very high center frequencies (8 and 12 kHz) was reported by Bernstein and Trahiotis (1994).

The differences in threshold ITDs obtained with transposed stimuli and the SAM tones having rates of modulation of 32, 64, and 128 Hz are even larger than those found at 4 and 6 kHz. This stems largely from the fact that the threshold ITDs obtained with both SAM tones and the transposed stimuli, in general, doubled when center frequency was increased to 10 kHz. Thus, when the differences in threshold ITDs between SAM and transposed stimuli are considered in terms of ratios, it appears that about the same relative improvement occurs with transposed stimuli, independent of center frequency. Perhaps the biggest departure in the patterning of the data at 10 kHz is that mean threshold ITDs



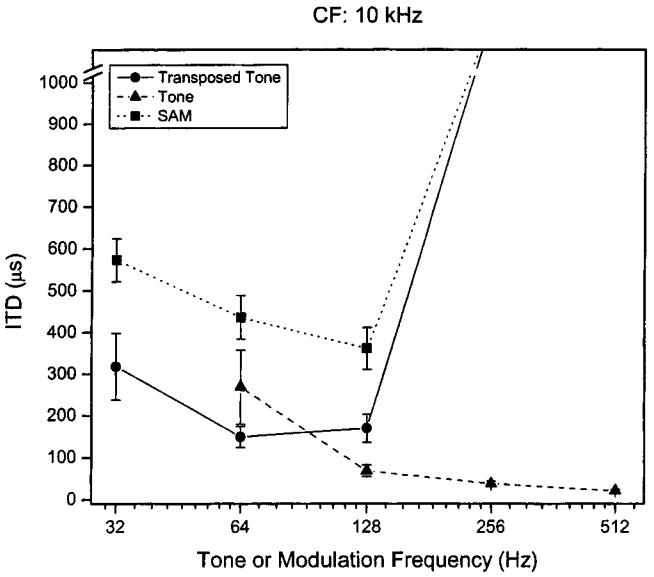
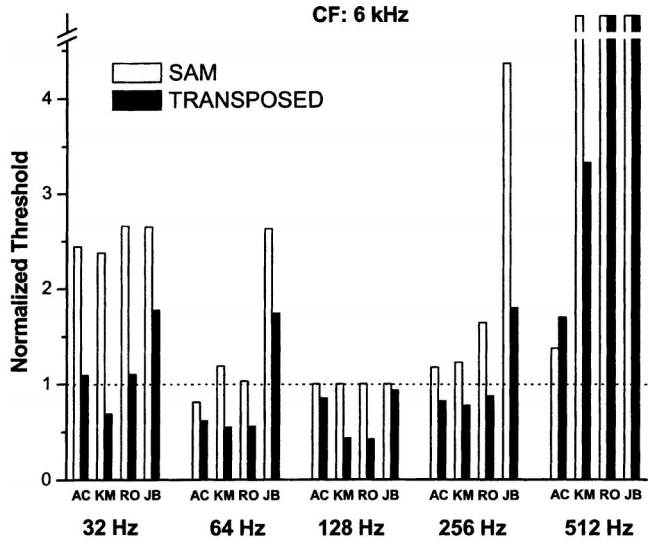


FIG. 5. Same as Fig. 3, but for SAM and transposed stimuli centered at 6 kHz (upper panel) and 10 kHz (lower panel).



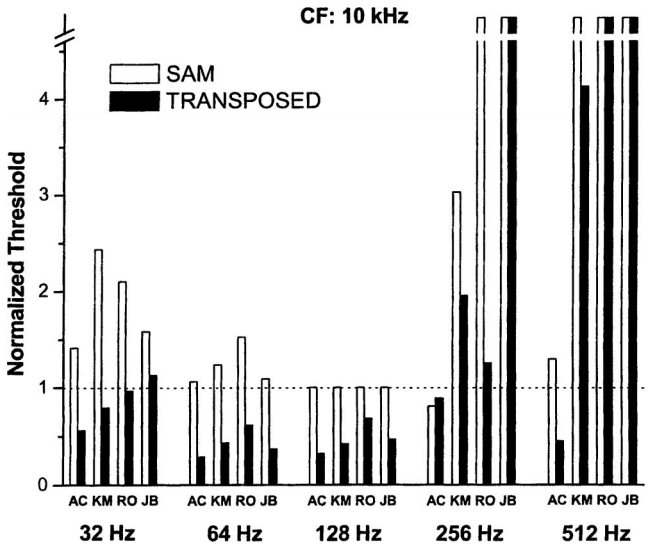


FIG. 6. Same as Fig. 4, but for SAM and transposed stimuli centered at 6 kHz (upper panel) and 10 kHz (lower panel).

could not be computed for data obtained at a rate of modulation of 256 Hz. This was so because two of the listeners, RO and JB, were unable to perform the task with either SAM or transposed stimuli.

The top and bottom panels of Fig. 6 contain individual listener's normalized thresholds at 6 and 10 kHz, respectively. At 6 kHz, the threshold ITDs obtained from all four listeners with the transposed stimuli (filled bars) were, once again, smaller than those obtained with the SAM tones (unfilled bars). The patterning of the normalized data at 6 kHz is very much like that found at 4 kHz with the exception that listener JB's normalized thresholds for rates of modulation of 64 and 256 Hz were larger than they were at 4 kHz. Still, listener JB's data clearly indicate smaller threshold ITDs with transposed stimuli than with SAM tones. The picture is much the same at 10 kHz, save for the fact that threshold ITDs were unmeasurable for listeners RO and JB in conditions in which the rate of modulation was 256 Hz. It is interesting, and to us important, that Fig. 6 reveals three stimulus conditions (a rate of modulation of 512 Hz for stimuli centered at 6 kHz, and rates of modulation of 256 and 512 Hz for stimuli centered at 10 kHz) in which listener KM was unable to perform the task with a SAM tone but was able to perform the task with a transposed stimulus.

The data obtained at center frequencies of 6 and 10 kHz, like those obtained at 4 kHz, indicate that threshold ITDs obtained with transposed stimuli are smaller than those obtained with SAM tones, and can sometimes lead to threshold ITDs that are essentially equivalent to or smaller than those obtained with low-frequency pure tones. Therefore, it appears to be generally true that the relative insensitivity to ITD typically observed with conventional high-frequency stimuli primarily stems from the nature of the information at the input to the binaural processor. That is, the high-frequency channels can support excellent sensitivity to ITD when the ''internal'' envelopes of the stimuli provide suffi-

cient information. These aspects of the data appear to be consistent with Colburn and Esquissaud's (1976) notion that frequency-related differences in sensitivity to ITD stem from frequency-related differences in the neural information that serves as input to the binaural portion of the auditory system.

On the other hand, the overall elevation in threshold ITDs observed at 10 kHz for both SAM and transposed stimuli may reflect true across-frequency differences within central binaural mechanisms that process ITDs, at least in terms of how they affect absolute sensitivity to ITDs. In our view, this outcome should not detract from the useful insights provided by Colburn and Esquissaud concerning the fundamental explanation for differences in sensitivity to ITD at low vs high frequencies.

The data obtained at all three center frequencies indicate that, in general, threshold ITDs increased as the rate of modulation was increased beyond 128 Hz. Furthermore, they increased more rapidly with rate of modulation (and more often were unmeasurable) as the center frequency of the stimuli was increased to 10 kHz. These effects cannot be explained by simply assuming that peripheral bandpass filtering causes reductions in depth of modulation of the stimuli as rate of modulation is increased and that this, in turn, degrades the binaural processing of ITDs. According to that line of argument, increasing the center frequency of the stimuli would lead to *improved* performance at the higher rates of modulation because the attendant increases in the bandwidths of the auditory filters would produce *relatively less reduction* in the depth of modulation of the stimuli. The data are not in accord with such an expectation. Instead, they appear to be consistent with there being some mechanism that serves to limit the ability to ''follow'' or to encode high rates of fluctuation of the envelope of high-frequency, complex waveforms.

IV. QUANTITATIVE ACCOUNTS AND INTERPRETATIONS OF THE DATA

We attempted to account for the data quantitatively by assuming that the listener's threshold ITDs reflect a constant change of the normalized interaural correlation. The mathematical model used to make the predictions was one we employed in previous studies (Bernstein and Trahiotis, 1996b; Bernstein *et al.*, 1999). It included ''envelope compression'' (exponent=0.23), square-law rectification, and low-pass filtering at 425 Hz to capture the loss of neural synchrony to the fine structure of the stimuli that occurs as the center frequency is increased (Weiss and Rose, 1988). For this study, the model was supplemented by an initial stage of bandpass filtering via Gammatone filters (see Patterson *et al.*, 1995) which, like the stimuli, were centered at either 4, 6, or 10 kHz.

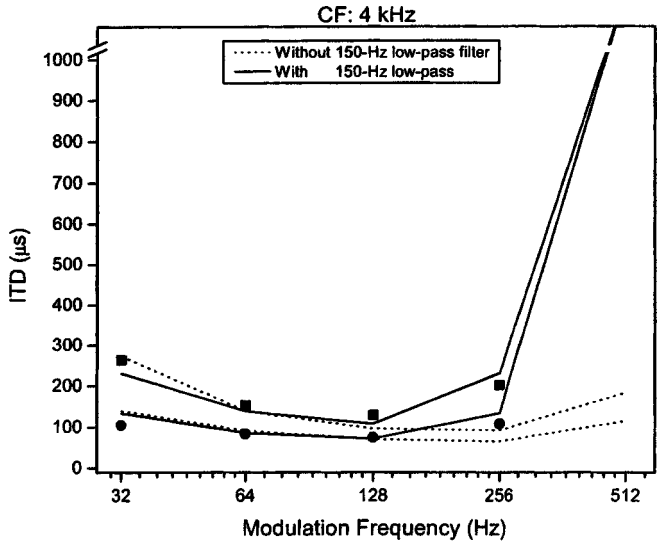
In order to make the predictions, it was necessary to determine functions relating ITD to normalized interaural correlation. This was done separately for SAM and transposed stimuli at each of the three center frequencies and at each of a large set of rates of modulation that included those actually used in the experiment. Numerical measures were obtained by implementing the peripheral stages of the model with MATLAB and then computing the normalized interaural correlation between the model's ''left'' and ''right'' outputs for a wide range of ITDs. Then, using a least-squares criterion, polynomials were fit to the paired values of normalized correlation and ITD.

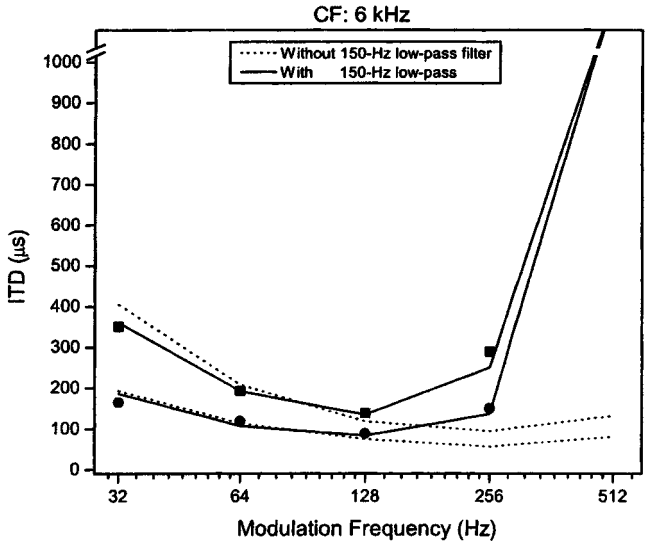
In order to arrive at predicted threshold ITDs, we sought the criterion value of normalized interaural correlation that maximized the amount of variance accounted for between predicted and obtained values of threshold ITD for data obtained with both SAM and transposed stimuli. A separate fitting procedure was carried out for the mean data at 4, 6, and 10 kHz in order to determine whether the criterion values of interaural correlation depended on center frequency. Stimulus conditions for which a mean threshold could not be computed (see Figs. 3 and 5) were not included in the computations of the amount of variance accounted for by the model. Nevertheless, predictions for such stimulus conditions were computed in order to determine what the model would predict.

The three panels of Fig. 7 contain the mean threshold ITDs for the SAM (squares) and transposed (circles) stimulus conditions along with the predictions from the model shown as dotted lines. The solid lines will be discussed below. Qualitatively and in general, the model appears to predict successfully the threshold ITDs for both SAM and transposed stimuli having rates of modulation of 32, 64, or 128 Hz. Quantitatively, the amount of variance in the data accounted[3] for by the model for those three frequencies of modulation was only 43% at 4 kHz, 10% at 6 kHz, and 64% at 10 kHz. At higher rates of modulation, the model fails to capture the dramatic increase in thresholds as the rate of modulation was increased to and beyond 256 Hz.

In an attempt to provide a satisfactory account of the loss of sensitivity to ITD at the higher rates of modulation, we further augmented the model by adding a final stage of monaural, 150-Hz low-pass filtering. The cutoff frequency was the same as that used by Kohlrausch *et al.* (2000) and Ewert and Dau (2000). The new predictions are indicated by the solid lines within each panel of Fig. 7 and appear to provide an improved fit to the data, especially for threshold ITDs obtained at center frequencies of 4 and 6 kHz. At those two center frequencies, the augmented model appears to account both for the elevated thresholds obtained at a rate of modulation of 256 Hz and for the fact that the average listener was essentially unable to perform the task at a rate of modulation of 512 Hz. For stimulus conditions for which a mean threshold could be defined, the amount of variance in the data that was accounted for by the model was 86% at 4 kHz, 96% at 6 kHz, and 77% at 10 kHz. It should be noted that a second-order low-pass filter was required to fit our binaural data, while a first-order filter appeared to fit the data of Kohlrausch *et al.* (2000) and Ewert and Dau (2000). The reasons for this difference are not understood at this time.

The changes of normalized interaural correlation ($\Delta\rho$) computed with the model (after bandpass filtering, compression, rectification, and low-pass filtering) required to fit the data were 0.000 23 at 4 kHz, 0.000 51 at 6 kHz, and 0.001 70 at 10 kHz. At face value, these values of $\Delta\rho$ suggest that sensitivity to envelope-based $\Delta\rho$ declines with increasing center frequency. This type of finding is consistent with our





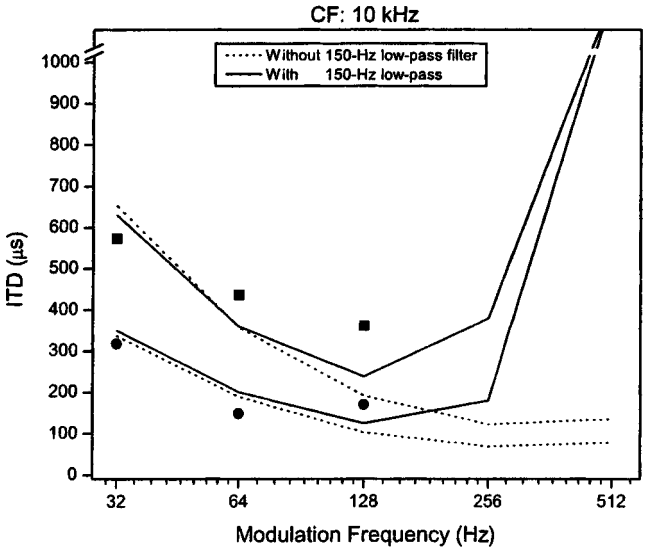


FIG. 7. Threshold ITDs for the SAM (squares) and transposed (circles) stimuli replotted from Figs. 3 and 5. The dotted lines represent predictions based on a constant criterion change in the normalized correlation computed subsequent to compression, rectification, and low-pass filtering at 425 Hz (see the text). The solid lines represent predictions obtained when the peripheral processing was supplemented by an additional 150-Hz low-pass filter.

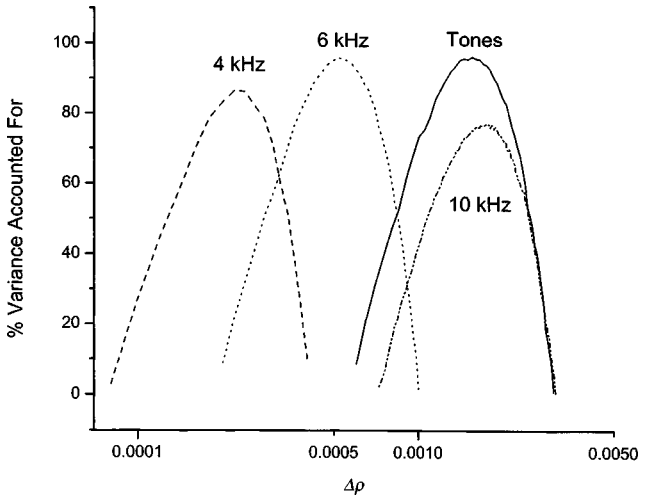


FIG. 8. Variance accounted for by the predictions as a function of the criterion $\Delta\rho$ for the pure-tones stimuli and for the stimuli centered at 4, 6, and 10 kHz, respectively.

previous research (Bernstein and Trahiotis, 1994). In order to evaluate the relative precision of the fits to the data obtained across center frequency in this study, we recomputed the percentages of variance accounted for at each center frequency while varying the criterion value of $\Delta\rho$. Figure 8 shows the results of the computations. The peaked nature of the plots indicates quite clearly that the fits are relatively precise and robust because relatively small changes in the criterion value of $\Delta\rho$ lead to relatively large changes in the amounts of variance accounted for. Furthermore, there is little overlap among the individual plots corresponding to fits obtained at the different center frequencies. Based on these findings it appears that our data and analyses are sufficiently precise to support the conclusion that sensitivity to envelope-based $\Delta\rho$ declines as center frequency is increased from 4 to 10 kHz.

Figure 8 also contains a plot of variance accounted for as a function of criterion $\Delta\rho$ for the tonal stimuli having frequencies of 128, 256, and 512 Hz. The mean threshold obtained at 64 Hz was excluded from the analysis because, as discussed much earlier in this presentation, it was not representative of performance measured across the four listeners. The 150-Hz low-pass filter was not included in the model because its function is to attenuate *modulations of amplitude* which are not present in tonal stimuli. In fact, including such a filter would appear to be folly because it would severely attenuate the internal, rectified representation of the signal and lead to the absurd prediction that sensitivity to ITD declines dramatically as the frequency of the signal increases beyond 150 Hz.

For the low-frequency tones, the criterion value of $\Delta\rho$ that best fit the data was 0.001 54. The plot representing the fits peaks in the region of 96% of variance accounted for, indicating that a correlation-based model that incorporates stages of peripheral auditory processing provides an excellent account of how threshold ITDs vary with frequency for pure tones. Note also that the plot representing the fits with the tonal stimuli overlaps greatly with the plot representing the fits for threshold ITDs obtained at a center frequency of

10 kHz. Therefore, based on the values of $\Delta\rho$ shown in Fig. 8, it appears that listeners are *least* sensitive to changes in interaural correlation for low-frequency pure tones and SAM and transposed stimuli centered at 10 kHz.

This outcome is somewhat counterintuitive because, as mentioned at the very beginning of this paper, threshold ITDs measured with low-frequency pure tones are typically smaller than those measured when the ITDs are conveyed by the envelopes of high-frequency stimuli. In fact, the data in Figs. 3 and 5, indicate that this is so. *A priori*, within a correlation-based approach, one might expect that values of threshold ITD and their $\Delta\rho$ counterparts would vary in a one-to-one fashion. For any particular stimulus, that is certainly the case. Considered across types of stimuli, however, such a relation does not occur. Specifically, the small threshold ITDs found with low-frequency tones correspond, through our model, to larger values of $\Delta\rho$ than do, for example, the larger threshold ITDs obtained at 4 and 6 kHz. This comes about because the empirically measured functions relating $\Delta\rho$ to ITD for the low-frequency pure tones are steeper than those measured for the high-frequency stimuli.

In an attempt to understand what aspect or aspects of the model lead to this outcome, we performed several computer-based analyses while omitting one or more of the peripheral stages of the model. It appears that the factor that is responsible is the differential effect that compression has on tonal stimuli and on the high-frequency complex waveforms. The type of ''envelope compression'' employed simply scaled the amplitudes of the pure tones but imposed a true compressive nonlinearity on the envelopes of the high-frequency SAM and transposed waveforms.

We investigated the effects of removing compression from the model. Doing so drastically reduced the amounts of variance accounted for. This was not completely unexpected because in a prior investigation (Bernstein *et al.*, 1999) we had demonstrated that the inclusion of envelope-based compression was necessary in the sense that it allowed us to account for binaural detection with maskers of divergent temporal features. In that study, we demonstrated that the data could not be accounted for if compression were not included. In addition, the form of compression used in the model is not arbitrary. Rather, it conforms to physiologically-based measures of basilar-membrane motion (e.g., Ruggero *et al.*, 1997) and is also constrained by successful fits to behavioral detection data obtained in independent investigations. For all these reasons, we believe that the form of compression used in the model is appropriate and that the differences in $\Delta\rho$ that account for the data are valid.

At this time, we can offer no satisfactory explanation for why listeners are relatively more sensitive to changes in $\Delta\rho$ conveyed by high-frequency stimuli centered at 4 and 6 kHz than they are to changes in $\Delta\rho$ conveyed by low-frequency pure tones of 128, 256, and 512 Hz. Our only speculation is that binaural detection is known to be constrained by additive ''internal noise'' which appears to decline dramatically for frequencies above 100 Hz (e.g., Shaw and Piercy, 1962). Yost (1988) has demonstrated that such internal noise also limits the magnitude of the MLD at low frequencies even when insert earphones, like the ones used in this study and which tend to attenuate this type of internal noise, are employed.

V. SUMMARY

Following van der Par and Kohlrausch (1997), we employed a procedure termed ''transposition'' in an attempt to provide the high-frequency channels of the auditory system with information like that normally available only at low frequencies. In these experiments, transposition entailed multiplication (modulation) of a high-frequency sinusoid by a rectified, low-pass filtered, low-frequency tone. Our findings indicate that threshold ITDs obtained with the transposed stimuli were generally smaller than those obtained with SAM tones and, at modulation frequencies of 128 and 64 Hz, were equal to or smaller than threshold ITDs obtained with their low-frequency pure-tone counterparts. Our quantitative analyses revealed that the data could be well accounted for via a model based on normalized interaural correlations computed subsequent to known stages of peripheral auditory processing augmented by low-pass filtering of the envelopes within the high-frequency channels of each ear. The data and analyses appear to be consistent with the general ideas advanced by Colburn and Esquissaud at a meeting of the Acoustical Society of America in 1976. They conjectured that the greater potency of ITDs typically observed for low frequencies as compared to high frequencies results from differences in the specific aspects of the waveform that are coded peripherally rather than from differences in the more central binaural mechanisms that process information from the different frequency regions. It should be understood that, in principle, any of a variety of high-frequency stimuli other than transposed ones, may, because of the temporal characteristics of their envelopes, foster enhanced sensitivity to ITDs. This would not detract from the general validity of the suggestions made by Colburn and Esquissaud so long as a model that assumes that a common *binaural* mechanism operates across frequency (such as the model employed here) is able to account for the data.

ACKNOWLEDGMENTS

This research was supported by research grants Nos. NIH DC-04147, DC-04073, and DC-00234 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. The authors thank Dr. Wes Grantham and one anonymous reviewer for their helpful comments that served to strengthen the logic and clarity of our presentation.

[1]We were surprised to find that stimulus levels produced according to the calibration supplied with the Etymotic ER-2 earphones sounded less loud than stimuli presented at nominally the same level via TDH-39 earphones, according to their calibration. Dr. Mead Killion, of Etymotic Research, validated our listening experience and agreed with us that the two respective methods of calibration would be expected to produce levels of stimulation differing by about 10 dB. We chose to ''calibrate'' the outputs of the Etymotic earphones to the nominal levels produced by the TDH-39s so that listeners in this study would receive levels of stimulation directly comparable to those utilized by us and others in prior psychophysical experiments employing TDH-39s. We verified that the levels from the Etymotic earphones were appropriate by presenting a high-frequency, stimulus to one ear via an Etymotic ER-2 earphone and simultaneously to the other ear via

a TDH-39 earphone. We then adjusted the relative levels between the two ears to produce a ''centered'' intracranial image, as is produced by diotic stimuli in normal-hearing listeners. In order to produce a centered image, it was necessary to impose a 10-dB larger voltage on the Etymotic ER-2 than would be expected on the basis of its calibration. Incidentally, the same type of ear-to-ear comparison allows one to compare and to cross calibrate any earphone to any other one, local variations in the frequency response of the earphones notwithstanding.

[2]In order to make this comparison, the values of ITD reported by Henning (1974) must be doubled. Henning introduced an ITD once per trial to the left or right ear. This provided twice as much information as compared to introducing the ITD to the same ear, once per trial, as in the current study and the study by Bernstein and Trahiotis (1994). Nueztel and Hafter (1981) employed a procedure similar to Henning's but the thresholds they report are twice the value of the ITD presented in each interval.

[3]The formula used to compute the percentage of the variance for which our predicted values of threshold accounted was $100\times(1-[\Sigma(O_i-P_i)^2]/[\Sigma(O_i-\overline{O})^2])$, where O_i and P_i represent individual observed and predicted values of threshold, respectively, and $\overline{O}$ represents the mean of the observed values of threshold.

Bernstein, L. R., and Trahiotis, C. (**1982**). ''Detection of interaural delay in high-frequency noise,'' J. Acoust. Soc. Am. **71**, 147–152.

Bernstein, L. R., and Trahiotis, C. (**1985**). ''Lateralization of sinusoidally amplitude-modulated tones: Effects of spectral locus and temporal variation,'' J. Acoust. Soc. Am. **78**, 514–523.

Bernstein, L. R., and Trahiotis, C. (**1992a**). ''Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies,'' J. Acoust. Soc. Am. **91**, 306–316.

Bernstein, L. R., and Trahiotis, C. (**1992b**). ''Detection of antiphasic sinusoids added to the envelopes of high-frequency bands of noise,'' Hear. Res. **62**, 157–165.

Bernstein, L. R., and Trahiotis, C. (**1994**). ''Detection of interaural delay in high-frequency SAM tones, two-tone complexes, and bands of noise,'' J. Acoust. Soc. Am. **95**, 3561–3567.

Bernstein, L. R., and Trahiotis, C. (**1996a**). ''On the use of the normalized correlation as an index of interaural envelope correlation,'' J. Acoust. Soc. Am. **100**, 1754–1763.

Bernstein, L. R., and Trahiotis, C. (**1996b**). ''The normalized correlation: Accounting for binaural detection across center frequency,'' J. Acoust. Soc. Am. **100**, 3774–3784.

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). ''The normalized correlation: Accounting for NoSπ thresholds obtained with Gaussian and low-noise masking noise,'' J. Acoust. Soc. Am. **106**, 870–876.

Blauert, J. (**1982**). ''Binaural localization: Multiple images and applications in room- and electroacoustics,'' in *Localization of Sound: Theory and Application*, edited by R. W. Gatehouse (Amphora, Groton).

Blauert, J. (**1983**). *Spatial Hearing* (MIT Press, Cambridge, MA).

Brugge, J. F., Anderson, D. J., Hind, J. E., and Rose, J. E. (**1969**). ''Time structure of discharges in single auditory nerve fibers in the squirrel monkey in response to complex periodic sounds,'' J. Neurophysiol. **32**, 386–401.

Colburn, H. S., and Esquissaud, P. (**1976**). ''An auditory-nerve model for interaural time discrimination of high-frequency complex stimuli,'' J. Acoust. Soc. Am. Suppl. 1 **59**, S23.

Ewert, S. D., and Dau, T. (**2000**). ''Characterizing frequency selectivity for envelope fluctuations,'' J. Acoust. Soc. Am. **108**, 1181–1196.

Henning, G. B. (**1974**). ''Detectability of interaural delay in high-frequency complex waveforms,'' J. Acoust. Soc. Am. **55**, 84–90.

Henning, G. B. (**1980**). ''Some observations on the lateralization of complex waveforms,'' J. Acoust. Soc. Am. **68**, 446–453.

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

Klumpp, R. G., and Eady, H. R. (**1956**). ''Some measurements of interaural time difference thresholds,'' J. Acoust. Soc. Am. **28**, 859–860.

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). ''The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers,'' J. Acoust. Soc. Am. **108**, 723–734.

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

McFadden, D., and Pasanen, E. G. (**1976**). ''Lateralization at high frequencies based on interaural time differences,'' J. Acoust. Soc. Am. **59**, 634–639.

Moore, B. C. J., and Glasberg, B. R. (**2001**). ''Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **110**, 1067–1073.

Nuetzel, J. M., and Hafter, E. R. (**1976**). ''Lateralization of complex waveforms: Effects of fine structure, amplitude, and duration,'' J. Acoust. Soc. Am. **60**, 1339–1346.

Nuetzel, J. M., and Hafter, E. R. (**1981**). ''Discrimination of interaural delays in complex waveforms: Spectral effects,'' J. Acoust. Soc. Am. **69**, 1112–1118.

Palmer, A. M., 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.

Par, S. van de, and Kohlrausch, A. (**1997**). ''A new approach to comparing binaural masking level differences at low and high frequencies,'' J. Acoust. Soc. Am. **101**, 1671–1680.

Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.

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.

Ruggero, M. A., Rich, N. C., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.

Shaw, E. A. G., and Piercy, J. E. (**1962**). ''Physiological noise in relation to audiometry,'' J. Acoust. Soc. Am. Suppl. 1 **34**, S45.

Siegel, R. A., and Colburn, H. S. (**1989**). ''Binaural processing of noisy stimuli: Internal/external noise ratios for diotic and dichotic stimuli,'' J. Acoust. Soc. Am. **86**, 2122–2128.

Weiss, T. F., and Rose, C. (**1988**). ''A comparison of synchronization filters in different auditory receptor organs,'' Hear. Res. **33**, 175–180.

Yost, W. A. (**1988**). ''The masking-level difference and overall masking level: Restating the internal noise hypothesis,'' J. Acoust. Soc. Am. **83**, 1517-1521.

Zwislocki, J., and Feldman, R. S. (**1956**). ''Just noticeable differences in dichotic phase,'' J. Acoust. Soc. Am. **28**, 860–864.

Interaural level differences and the level-meter model

William M. Hartmann[a] and Zachary A. Constan
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

(Received 23 May 2001; revised 19 June 2002; accepted 20 June 2002)

The interaural level difference (ILD) plays a significant role in sound localization. However, the definition of ILD for noise is open to some interpretation because it is not obvious how to deal with the inevitable level fluctuations. In this article, the ILD is interpreted as an energylike (time-integrated) measure of stimulus level, independent of other stimulus details—particularly interaural correlation. This concept is called the ''level-meter model.'' The model was tested by measuring human ILD thresholds for noise stimuli that were interaurally correlated, or anticorrelated, or uncorrelated. An additional test (not involving lateralization) measured the threshold for level discrimination based on loudness. According to the level-meter model, all four thresholds should be the same. The experimental results showed that the predictions of the level-meter model held good to within about half a dB, although thresholds for level discrimination were systematically higher than ILDs. Among the ILDs themselves, thresholds were slightly higher for uncorrelated noise. The latter result could be explained by replacing the level-meter model with a loudness-meter model, incorporating temporal integration. The same model accounted for the bandwidth dependence of the threshold. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1500759]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Fe [LRB]

I. INTRODUCTION

An important contribution to the human ability to localize a sound is the difference in level between the two ears, or interaural level difference (ILD). The object of the work described in this article was to gain insight into the way the auditory system uses ILD information in a broadband noise signal. Specifically, we wanted to know whether it matters if the noises in the two ears are mutually coherent or not. Although coherence is necessary for the use of interaural time differences, it is not obvious that coherence plays any role in the use of ILD information.

To bring the question into clear focus, we formed a null hypothesis about the way that the binaural system measures the ILD, namely the ''level-meter model.'' According to the level-meter model the binaural system measures the power in the left ear with one meter, measures the power in the right ear with a second meter, compares those two power measurements, and uses the difference to localize or lateralize. Binaural coherence is unimportant to the operation of this system because the system only measures the average levels from the two ears. This model would predict that the ability to detect small differences in interaural level should be independent of whether the noise is interaurally correlated or uncorrelated. Implicit in the level-meter model is an idealized integration time, long enough to average over stimulus fluctuations so that the precision of the level measurement is limited only by the noise duration.

The alternative to the level-meter is a model in which the binaural system does not reduce levels in left and right ears to single individual magnitudes prior to comparison. For instance, a system might tend to track the difference between left and right signals as a function of time. Such an ILD-tracking system is sensitive to the difference between correlated and uncorrelated noise because of inherent fluctuations in noise power. For perfectly correlated noise (cross correlation$=1$), the signals in the two ears are identical except for a fixed level difference. The instantaneous ILD is constant even though the noise fluctuates. By contrast, if the noise signals in the two ears are perfectly uncorrelated (cross correlation$=0$), then the fluctuations in left and right ears are independent, and the instantaneous ILD varies as well. The ILD fluctuation leads to additional variance, causing this ILD-tracking model to predict that the ability to detect small differences in interaural level should be better for correlated noise than for uncorrelated.

In their study of monaural versus binaural discrimination, Jesteadt and Wier (1977) made a similar distinction between models. Their *independent threshold* model, like the level meter model, is insensitive to interaural coherence. Their *information integrating* model, like the tracking model, is sensitive, possibly giving different results for interaurally correlated or uncorrelated noise.

Apart from models of measurement, there is the matter of perception. A perfectly coherent (diotic) noise is normally perceived as a compact image near the center of the head. Introducing a small ILD moves the image to the left or right but retains the compact character if the ILD is less than about 8 dB (Blauert, 1983). By contrast, a binaurally uncorrelated noise forms a fuzzy image that fills the head. *A priori*, it seems likely that listeners should be able to lateralize the compact image more successfully than the fuzzy image. Thus, both the ILD-tracking model and common experience suggest that ILD sensitivity should be greatest for correlated noise, contrary to the prediction of the level-meter model.

The role of binaural coherence in connection with ILD

[a]Electronic mail: hartmann@pa.msu.edu

sensitivity has been studied before. In unpublished work, Grantham and Ahlstrom (1982) measured ILD sensitivity for broadband noise and narrow-band noise centered on several frequencies. For two out of three listeners, ILD thresholds tended to be smaller for diotic-except-for-level (correlated) noise than for completely uncorrelated noise. The advantage observed for correlated noise was greater when noise bursts were brief, 30 ms, than when noise bursts were 500 ms in duration. Similarly, an abstract by Neutzel (1982) reported slightly smaller ILD thresholds for broadband correlated noise compared to uncorrelated noise. The experiments of the present article expand on these works, including additional conditions.

II. EXPERIMENT 1—BROAD BAND

Experiment 1 measured the human ability to detect small interaural differences in the level of broadband noise under three conditions of noise coherence:

(1) Diotic-except-for-level: Noises were identical in the two ears except for a difference in level. Diotic noise is called *No* in the traditional notation of masking level difference (MLD) studies; the cross-correlation function for zero lag is 1.
(2) Anticorrelated: Noises were identical in the two ears except for a difference in level and an overall factor of -1 for one ear. Thus each spectral component in the left ear was 180 degrees out of phase with the corresponding component in the right ear. Anticorrelated noise is called $N\pi$ in traditional MLD notation; the cross correlation for zero lag is -1.
(3) Uncorrelated: Noise sources for left and right ears were independent. Uncorrelated noise is called *Nu* in traditional MLD notation; the ensemble-averaged cross correlation is 0 for all values of the lag.

The ability of listeners to detect ILDs was probed in a lateralization experiment using headphones. If lateralization, based on ILD information, is only a matter of comparing levels in the left and right ears, as conjectured in the level-meter model, then lateralization thresholds should be the same for all three noise conditions. Furthermore, the model suggests that lateralization cannot be more accurate than the initial independent level measurements in the two ears. As will be shown in Sec. IV, the model predicts that threshold ILDs should be equal to the difference limen in intensity, as also suggested years ago by von Békésy (1930). It is an indication of the power of the level-meter model that it unifies such disparate percepts as lateralization and intensity discrimination. To test this idea, we added a fourth condition,

(4) DLI (difference limen in intensity): Noises in left and right ears were always identical.

A. Method

1. Lateralization experiment

In the lateralization experiment, noise was presented on two successive intervals. The two intervals had ILDs of equal magnitude but opposite sign. The sign reversal was expected to cause the image on the second interval to be either to the right or to the left of the image on the first, corresponding to ''left–right'' or ''right–left'' trials, respectively. The listener was required to indicate the direction of the change in a two-alternative forced-choice task.

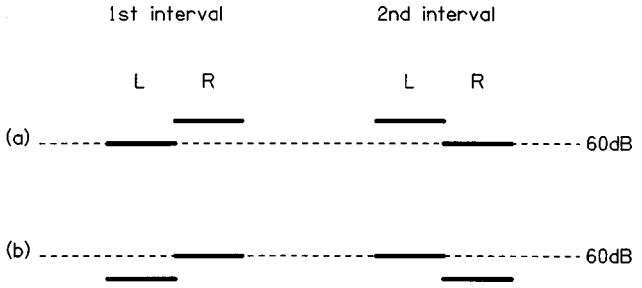


FIG. 1. There were two options (a) and (b) for the level variation on a *right–left* trial as shown. The two options were used randomly and equally often. The total power (left+right) was the same on both time intervals, but the ILDs were reversed from the first interval to the second. There were similar options for *left–right* trials.

On every interval, one ear received the standard level of 60 dB SPL. As shown in Fig. 1, a right–left trial could be made either by making the right-ear level larger on the first interval or making the left-ear level smaller on the first interval. Both options were used with equal probability according to a random schedule. Whatever the ILD on the first interval, its sign was reversed to make the ILD for the second. Similarly, two corresponding sequences were randomly used for left–right trials.

The magnitude of the ILD was varied in a staircase fashion. The staircase was one-up three-down, targeting the 79% correct point on a psychometric function. For every incorrect response the interaural level difference (ILD) was increased by an increment. After three successive correct responses, the ILD was decreased by an increment. The magnitude of the increment was caused to vary according to the staircase value, with the goal of reaching threshold quickly and providing accuracy in the vicinity of threshold. When the experimental run began, the initial ILD was 2 dB, and the ILD increment was 0.5 dB. As the staircase progressed, the ILD increment depended on the ILD. When the ILD was greater than 1.5 dB the increment was 0.5 dB. When the ILD was between 1.0 and 1.5 dB, the increment was 0.2 dB. When the ILD was less than 1.0 dB, the increment was 0.1 dB. The minimum possible ILD was 0.1 dB.

A staircase run continued until the staircase had changed direction 14 times. The first four turning points were discarded and the average and standard deviation for the remaining ten turning levels became the data for the run. The duration of runs ranged from 2 to 4 minutes.

2. DLI experiment

The stimulus for the DLI experiment was a minor variation on the lateralization experiment. To make the DLI experiment, the stimulus for the right ear from the lateralization experiment was sent to both left and right ears. Therefore, the experiment was diotic. The listener's task was to say

whether the first or second interval was the louder. Otherwise, the DLI experiment was identical to the lateralization experiments.

3. All experiments

The experiments of four types, (1) lateralization of correlated noise, (2) lateralization of anticorrelated noise, (3) lateralization of uncorrelated noise, and (4) DLI for diotic noise, were done in random order, except that no experiment type was done more than twice in succession. After a few training runs, listeners completed runs for which data were collected. Listeners received no feedback. From the experience of Grantham and Ahlstrom (1982), and from our own pilot experiments, we expected the distinctions among the four types of experiments to be subtle. Therefore, we insisted that the runs lead to rather tight staircases with small variance. Runs with a standard deviation among turning levels less than 0.3 dB were considered ''tight'' and were accepted. Runs with larger standard deviation were repeated. Final data were based on the last six tight runs for each listener on each of the experiment types, a total of 24 runs, or 144 runs for the six listeners.

B. Stimuli

All stimuli were white, Gaussian, broadband noises generated by Zener diodes biased near the breakdown knee. The spectrum was verified by a spectrum analyzer. The Gaussian character was checked by repeated sampling and plotting a histogram of the values. For the uncorrelated noise experiment independent noise generators were used. The noises were low-pass filtered at 10 kHz (−48 dB/oct). The spectrum level was 20 dB, making the stimulus level 60 dB SPL. Stimuli were turned on with matched voltage-controlled amplifiers, controlled by a common gating signal with a rise/fall time of 30 ms. Therefore, lateralization was not affected by interaural onset differences. The gate was triggered by a Tucker-Davis TG6 module to establish precise timing. After a 300-ms warning interval, marked by a green pilot lamp, there were the two stimulus intervals, 500 ms in duration, separated by a 500-ms silent gap. After the stimulus intervals, a red pilot lamp requested a response from the listener. There was no time limit for the response. The experiment was self-paced with the next trial beginning 300 ms after a response. The 500-ms stimulus duration placed a theoretical limit on our ability to test the level meter model.

Listeners were seated in a double-walled sound-attenuating room (Acoustic Systems model RE-244) and received the stimuli via Sennheiser HD 480 headphones. They made their responses by pressing one of two buttons on a response box.

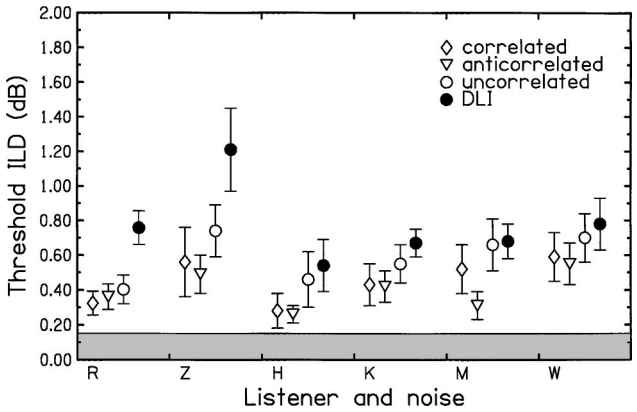


FIG. 2. Results of experiment 1: Threshold ILDs are shown for three noise types: Diamonds for binaurally correlated noise, triangles for anticorrelated noise, and open circles for uncorrelated noise. Filled circles show difference limens for the level of diotic noise (DLI). The experimental method could not measure thresholds below 0.15 dB.

C. Listeners

There were six listeners, R, Z, H, K, M, and W. Listeners R, Z, H, and M were males, ages 21, 29, 60, and 19. Listeners K and W were female, ages 21 and 19.

D. Results

For each listener and experiment type, the mean of the six runs and the standard deviation were found. These data appear in Fig. 2. Of particular interest was a comparison among the different experiment types. This comparison, averaged over listeners, is shown in the top line of Table I.

Figure 2 shows that all thresholds fell between 0.2 and 1.3 dB. Thus, they were all similar. However, some systematic differences are also visible. For every listener the largest threshold was the DLI. Also, among the ILDs (open symbols), every listener's largest threshold was for uncorrelated noise. The ILD threshold values for correlated and uncorrelated noise from Table I are 0.451 and 0.585 dB, respectively. These can be compared with values of 0.33 and 0.40 dB obtained by Neutzel (1982) using a four-interval procedure.

An analysis of variance was conducted on the four thresholds: lateralization of noise (correlated, anticorrelated, and uncorrelated) and the difference limen in intensity. The test indicated significant differences among the four, $F(3,15)=16.7$, $p<0.01$. *Posthoc* comparisons of the individual means found that the DLI was greater than all three threshold ILDs ($p_{\max}<0.03$). Further, the threshold ILD for anticorrelated noise was less than that for uncorrelated noise, $F(1,15)=10.42$, $p=0.03$. These paired comparisons, like all the others in this article, include a Bonferroni correction factor (Dunn, 1961) for multiple comparisons.

TABLE I. Threshold ILDs in dB from experiments 1–3, averaged across listeners for each stimulus type.

Experiment	Bandwidth (kHz)	Correlated	Anticorrelated	Uncorrelated	DLI
1	10	0.451	0.398	0.585	0.773
2	1	0.679	0.733	0.835	1.021
3	1 (rnd)	0.756	0.766	0.880	1.152

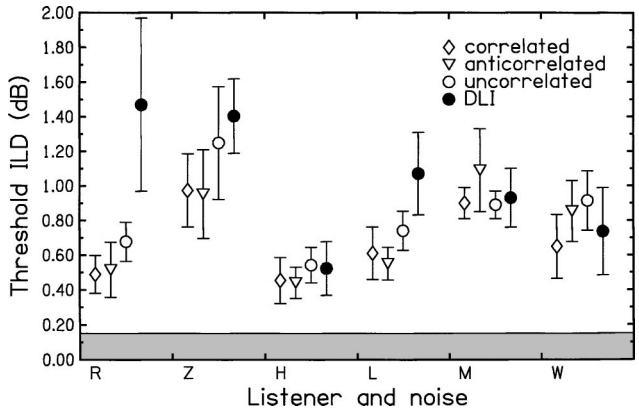


FIG. 3. Same as Fig. 2 but showing the results of experiment 2.

An analysis of variance on the three ILD thresholds alone indicated a significant difference among these noise types, overall, $F(2,10)=16.67$, $p<0.001$. *Posthoc* Bonferroni-protected comparisons of the individual means showed significant differences between the correlated and uncorrelated noises, $F(1,10)=16.07$, $p=0.007$, and between the anticorrelated and uncorrelated noises, $F(1,10)=31.30$, $p<0.001$.

III. EXPERIMENT 2

Experiment 1 found a weak effect of noise coherence on the ability to lateralize a noise. Because the noise was broadband, 0–10 000 Hz, it is not entirely clear how to interpret this result. The human binaural system is sensitive to interaural coherence in envelope fluctuations over a broad frequency range (Henning, 1974). However, the system is insensitive to coherence in the fine structure above about 1500 Hz and cannot use interaural time differences (ITD) in the fine structure to lateralize above this frequency. In Experiment 1, most of the noise power was outside this low-frequency range. By contrast, listeners' sensitivity to ILD is essentially independent of frequency (Yost, 1981).

The goal of experiment 2 was to provide a more severe test of the concept that ILD sensitivity is independent of noise type. The noise was low-pass filtered so that all the noise power was below 1 kHz. It was expected that interaural coherence would be more important in experiment 2 because both fine-structure ITD and envelope ITD contribute to lateralization and to perceived auditory source width across the entire frequency range.

A. Method

Experiment 2 was identical to experiment 1 in method except that the low-pass filter cutoff was decreased from 10 to 1 kHz, again −48 dB/oct. The noise level in the headphones remained 60 dB. In another change, listener K was replaced by listener L, a male age 20.

B. Results

The results of experiment 2 are shown in Fig. 3, comparable to Fig. 2 for experiment 1. Averages over listeners are given in Table I. The comparison shows that lateralization performance in experiment 2 declined compared to experiment 1. For the five listeners common to experiments 1 and 2, the average threshold across the four experiment types was 0.82 dB on experiment 2 compared to 0.55 dB on experiment 1. For every listener and every experiment type, threshold either increased or stayed about the same when the bandwidth was reduced from 10 to 1 kHz. Evidently all listeners had made good use of power above 1 kHz in experiment 1.

Experiment 2 was done with the intention of magnifying differences among the different experiment types. However, Fig. 3 shows that quite the reverse occurred. Unlike experiment 1, there is no stimulus type that leads to the greatest threshold for all listeners, and there is no universal agreement within the ILDs (open symbols) either.

An analysis of variance performed on the four experiment types showed that the difference among thresholds was marginally significant, $F(3,15)=3.35$, $p=0.045$. However, paired comparisons indicated no significant differences, $p_{\min}\geqslant 0.06$.

In an analysis of variance on the three ILD thresholds alone, the three noise types were found to be significantly different overall, $F(2,10)=4.25$, $p<0.05$. This was chiefly due to a marginally significant mean difference between the thresholds for correlated and uncorrelated noise, $F(1,10)=8.22$, $p=0.05$. The other individual comparisons did not approach significance. In the end, it was only barely possible to challenge the level-meter model on the basis of the results of experiment 2.

IV. EXPERIMENT 3

There is a remote possibility that the lateralization tasks in experiments 1 and 2 were not really binaural. Inspection of the protocol in Fig. 1 shows that it would theoretically have been possible for a listener to monitor a single ear and perform the task based on a loudness comparison among the two intervals. As will be seen in the discussion section that follows, this would have been a poor strategy because the experiments showed that level discrimination was less successful than lateralization. However, this strategy remains a theoretical possibility. The purpose of experiment 3 was to provide a check on experiment 2 by making it very unnatural to perform the task using a monaural strategy.

A. Method

Experiment 3 was identical to experiment 2 except that the standard level (always 60 dB in experiments 1 and 2) was randomized on every experimental interval. The distribution of levels was rectangular, centered on 60 dB, and $R=5$ dB in width. According to Green (1988), a level randomization of R dB should lead to a threshold level of $R[1-\sqrt{2(1-P_C)}]$, where P_C is the percentage of correct responses on a two-alternative task. For $P_C=0.79$, the expected threshold for decisions based on level alone is 1.8 dB. The level randomization was applied to the lateralization tasks only. It could not logically be applied to the diotic level discrimination task, and thus the DLI task remained identical to experiment 2. Listeners were the same as in experiment 2.

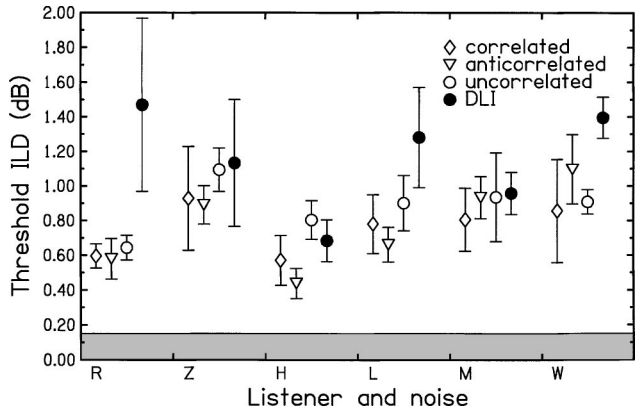


FIG. 4. Same as Fig. 2 but showing the results of experiment 3.

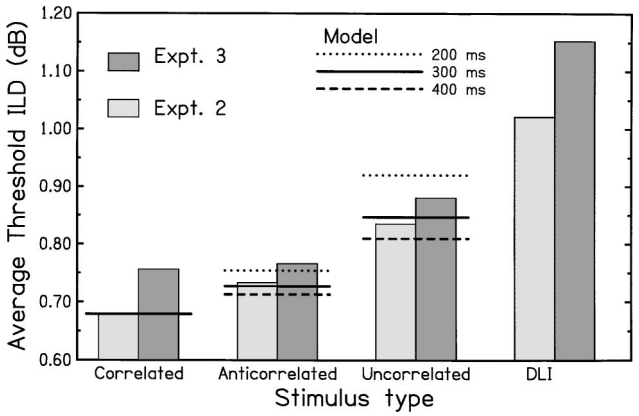


FIG. 5. Thresholds, averaged over the six listeners, for experiments 2 and 3. The vertical scale has been greatly expanded compared to Figs. 3 and 4 to show details better. Dotted lines show the predictions of the loudness-level meter model for integration times of 200 and 400 ms. The dashed line shows the prediction for 300 ms.

B. Results

The results of experiment 3 are shown in Fig. 4. They can be compared with Fig. 3 for experiment 2 (fixed standard). Averages over listeners appear in Table I. The largest value ever seen for any listener was 1.1 dB, notably less than 1.8 dB expected for judgments based on level in a single ear. We conclude that listeners did not make their decisions based on level alone, but used interaural level differences to do the task.

It was expected that experiment 3 would lead to the same results as experiment 2. The data suggest that the results were indeed similar. Of the 18 ILD thresholds (6 listeners$\times$3 noise conditions), 11 increased and 5 decreased when the standard was randomized. No change was outside the error bars. The only difficulty encountered in experiment 3 was that listener *R*, who was clearly having problems with the DLI task in experiment 2 (Fig. 3), consistently failed to satisfy our criterion for tight DLI staircases on the identical task in experiment 3. Therefore, his DLI data point for experiment 2 was used also for experiment 3.

Table I shows that the order of thresholds among experiment types in experiment 3 was the same as in experiment 2. An analysis of variance performed on the four experiment types showed that the difference among thresholds was significant, $F(3,15)=6.53$, $p<0.01$. Comparisons of paired differences indicated only two significant differences, namely DLI vs ILD for correlated noise, $p=0.01$, and DLI vs ILD for anticorrelated noise, $p=0.01$.

An analysis of variance performed on the three ILD thresholds showed that trends among the different noise types for experiment 3 were similar to those for experiment 2. Subjects were, however, somewhat more variable in exhibiting those trends. Accordingly, the overall analysis of variance on the noise results approached, but did not reach significance, $F(2,10)=2.66$, $p=0.12$.

To further compare experiments 2 and 3, the averages over listeners are given in Fig. 5. The plots are nearly parallel. The most likely explanation for the small consistent difference between the thresholds for experiments 2 and 3 is that the level randomization on experiment 3 slightly disrupted listener concentration.

V. DISCUSSION

A. The level-meter model

The level-meter model predicts that the thresholds for the four tasks in the experiments should all be the same. According to this model the signals in left and right ears are not compared on a moment-to-moment basis; only an average representation of the signal levels is compared—average neural driven firing rate, for example. Because the level measurements by the two ears are independent in this model, it is immaterial whether the signals are correlated or not. Therefore, one expects the threshold ILDs to be the same for all three noise types. The level-meter model also predicts that the difference limen in intensity (DLI) should be the same as the threshold ILDs.

The argument for this equivalence is as follows: Suppose that a level measurement in a single ear has a standard deviation σ. The distribution of interaural level differences then has a standard deviation of $\sigma\sqrt{2}$. Therefore, if a power increment, Δ, is added to one ear, the ratio of internal ILD to standard deviation for a single interval is $\Delta/(\sigma\sqrt{2})$. Because the ILD, Δ, is added to one interval and subtracted from the other in the lateralization task, the means of the interval distributions differ by 2Δ, and the standard deviation of the difference distribution becomes 2σ. Therefore, a comparison between first and second intervals leads to a d' of $2\Delta/(2\sigma)$, or Δ/σ.

In the DLI experiment, the level difference between the two intervals is Δ, and the ratio of internal level difference to the standard deviation for a differential measurement is $\Delta/(\sigma\sqrt{2})$. However, this is the result for a single ear. In this experiment, both ears receive the same stimulus. As shown by Jesteadt and Wier (1977), d' for diotic DLIs is better than for monaural by the square root of 2. Therefore, the ratio of internal difference to standard deviation is Δ/σ, the same result as obtained for lateralization.

A brief argument that leads to the same conclusion is that for both ILD and DLI tasks, both ears are subjected to a level difference of Δ between the first and second intervals.

Therefore, a level meter should behave the same on both types of experiments.

B. ILD vs DLI

Although the thresholds in Figs. 2–5 are all approximately the same, there are systematic differences for different experiment types. These differences show that the level-meter model cannot be exactly right. The largest systematic difference is between the DLI and the ILD threshold for correlated noise. The ILD is notably smaller.

The smaller ILD threshold suggests the involvement of a specific binaural comparison process. For example, neurons in the lateral superior olive that are sensitive to finite ILDs are not involved in the DLI experiment because ILDs are zero in that experiment. Such neurons are activated by the ILD experiments, and information from them could be responsible for the lower threshold for ILD compared with DLI.

Alternatively, our theoretical comparison of the DLI and ILD tasks may be incomplete because the two tasks impose different memory requirements. The DLI task requires the comparison of two sequential loudnesses; an ILD task requires the comparison of two sequential lateral positions.

C. The loudness-meter model

According to the level-meter model, binaural coherence should play no role in ILD thresholds. The experiments of this article have shown that this concept is approximately true. However, it does not appear to be entirely true. Our best estimate for the deviations from the simple level-meter model appear as the averaged data for experiment 2 in Fig. 5. There, ILD thresholds are shown to be lowest for correlated noise, highest for uncorrelated noise, and intermediate for anticorrelated noise, though the difference between correlated and anticorrelated noise was not found to be statistically significant.

We discovered that a straightforward modification of the level-meter model can account for the small deviations observed experimentally. A clue actually appeared in the work by Grantham and Ahlstrom (1982). Their measurements showed that ILD thresholds were similar for correlated and uncorrelated noise when the noise bursts were 500 ms in duration, but thresholds were usually larger for uncorrelated noise when the noise bursts were only 30 ms in duration. This result indicates that listeners are able to average over the additional variability in uncorrelated noise given the opportunity to do so, but that the additional variability leads to increased thresholds when listeners are prevented from integrating. This insight further suggests that for long stimulus durations, as used in the experiments of this article, the thresholds for uncorrelated noise are limited by the finite human auditory integration time. The auditory integration time is an element in the perception of loudness.

The remainder of this section introduces the loudness-meter model as an improved alternative to the level-meter model. The loudness-meter model includes two elements of loudness, namely temporal integration and compression. It also includes half-wave rectification of the signal, a physiologically based feature that is common in current auditory models. This section shows that with reasonable values of the integration time, the additional variance for uncorrelated noise is just large enough to account for the threshold differences seen experimentally.

The loudness-meter model, from signal to decision criterion, is as follows:

Let the signal in ear e ($e=L$ or R) be written as

$$x_e(t)=\sum_{n=1}^{N} A_n^e \cos(2\pi f_n t+\phi_n^e), \tag{1}$$

where A_n^e and ϕ_n^e are the amplitude and phase of component n in ear e. The amplitude is given by a Rayleigh distribution (e.g., Hartmann, 1997) and the phase is rectangularly distributed over 360 degrees. On different experimental intervals, the amplitudes and phases of the components will be different, and these differences lead to the stimulus variability.

The excitation rate is a half-wave rectified and compressed version of the signal,

$$r_e(t)=\{\mathcal{H}[x_e(t)]\}^{\eta}, \tag{2}$$

where $\mathcal{H}$ is the half-wave operator, and η is the compression exponent. The half-wave rectification will be directly responsible for the fact that anticorrelated noise leads to greater variability than correlated noise. The compression exponent is taken to be $\eta=0.6$, consistent with the observed rule for loudness (Fletcher, 1953; Stevens, 1955).

The excitation attributable to ear e, obtained by the listener on an experimental observation interval, is the integral of the excitation rate, $r(t)$,

$$E_e=\int_0^T dt\, r_e(t), \tag{3}$$

where T is the integration time. In the limit that the integration time is infinite, the system functions like the ideal level meter and the variability in E_e depends only on the duration of the excitation.

Finally, the left-or-right decision criterion, based on ILD, is the excitation difference relative to the mean,

$$\Delta=2\frac{E_R-E_L}{E_R+E_L}. \tag{4}$$

The advantage in normalizing Δ by the mean in this way is that it makes the decision criterion independent of the overall level of the experiment. The sign of criterion Δ determines whether the image is perceived to be on the left or the right. The magnitude of Δ determines whether the image is close to the midline or off to the side.

Because of the stimulus variability, the values of E_R and E_L are different on different experimental intervals, and the value of Δ varies from trial to trial. For correlated noise, however, any variation in E_R is perfectly mirrored in E_L. Because of the normalization, Δ becomes a function of the stimulus ILD only and there is no stimulus variability. If noises are not perfectly interaurally correlated, there is a trial-to-trial variation in Δ, characterized by a stimulus variance, σ_S^2.

D. The loudness-meter model and coherence

1. The sensitivity index

Average $\overline{\Delta}$ is the expected value of the decision variable, Δ. It is the ensemble average, i.e., the average over all possible signals, as specified by the sets of amplitudes and phases $\{A^L\}\otimes\{A^R\}\otimes\{\phi^L\}\otimes\{\phi^R\}$.

In an experiment with stimuli drawn from the ensemble, the variance of Δ is $\mathrm{var}(\Delta)=\sigma^2$. Together with $\overline{\Delta}$, the variance determines the sensitivity index d' (Green and Swets, 1966),

$$d'=\overline{\Delta}/\sigma. \tag{5}$$

The variance is composed of two parts, an internal noise σ_N^2 and a stimulus noise σ_S^2. If these two forms of noise are assumed to be additive and independent, the total variance is given by

$$\sigma^2=\sigma_N^2+\sigma_S^2. \tag{6}$$

2. Detailed calculation

A key to the calculation is that when the stimulus is coherent (diotic-except-for-level), there is no stimulus variance, i.e., $\sigma_S=0$. Because the stimulus variance is zero, all the total variance is internal, and the measured ILD threshold, L_R-L_L, gives an estimate for σ_N. That calculation is simplified by using an analytic form for the ensemble average $\overline{\Delta}$ in terms of the interaural level difference (L_R-L_L) in dB,

$$\overline{\Delta}=2\,\frac{10^{\eta(L_R-L_L)/20}-1}{10^{\eta(L_R-L_L)/20}+1} \tag{7}$$

or

$$\overline{\Delta}=2\tanh[0.057\,56\eta(L_R-L_L)].$$

Equation (7) is easily proved for noise that is diotic-except-for-level. To prove it for other conditions requires that the ensemble average of several nonlinear functions be equal to the functions of their ensemble-averaged arguments, not generally a valid step. However, numerical experiments using ensembles of 2000 waveforms showed that Eq. (7) holds good to much better than 1% accuracy.

The loudness-meter model was tested on the average results for experiment 2. The calculation was a three-step process. The first step required finding the internal noise, σ_N, from the threshold ILD for correlated noise using Eqs. (7) and (5) with $d'=1.16$. The ILD threshold of 0.68 dB led to $\sigma_N=0.0405$. Because of the normalization in Eq. (4) both Δ and σ are dimensionless.

The second step required the computation of the stimulus variance σ_S^2 as a function of integration time. Calculations of the variance of Δ, based on 2000 waveforms for a given integration time, led to the values of σ_S in Table II. As it turned out, the variance for uncorrelated noise was larger than the variance for anticorrelated noise by about a factor of 2. It was not *a priori* evident that this would be the case, but it always was, for any value of integration time we studied (25 to 400 ms).

TABLE II. Values of internal and stimulus noise computed for experiment 2.

T (ms)	σ_N	$\sigma_{S=\pi}$	$\sigma_{S=u}$
200	0.0405	0.019 38	0.036 90
300	0.0405	0.015 34	0.030 04
400	0.0405	0.012 85	0.026 24

The third step added the internal and stimulus variances according to Eq. (6) and used the total to predict threshold ILDs (L_R-L_L) from Eqs. (5) and (7). The predicted thresholds for integration times of 200 and 400 ms are given by the dashed and dotted lines in Fig. 5. All these useful integration times are less than the duration of our experimental stimuli, 500 ms, indicating that the experimental method did not impose an important duration limitation. The best agreement between model and experiment is obtained with an integration time of 300 ms, shown in Fig. 5 by the solid line. An integration time of 300 ms agrees with the value obtained by Plomp and Bouman (1959) for low-frequency stimuli such as ours, although it should be noted that their integration window was exponential whereas ours was rectangular. With an integration time of 300 ms, the agreement between the model and experiment is excellent.

E. The loudness-meter model and bandwidth

As shown by Table II, the noise that limits performance in experiment 2 is mostly internal noise. In the spirit of the loudness-meter model, it ought to be possible to improve the signal-to-noise ratio by increasing the bandwidth of the stimulus so that more channels of the binaural system are used. Because experiments 1 and 2 showed that ILD thresholds always decreased when the bandwidth was increased from 1 to 10 kHz, there is preliminary evidence that such an effect might be operating. For instance, the ILD threshold for correlated noise decreased from 0.68 to 0.45 dB with the increased bandwidth. It is interesting to ask whether this decrease can be predicted by the loudness-meter model.

The number of auditory channels involved in processing can be taken to be the number of critical bands within the range of the stimulus. A top frequency of 1000 Hz corresponds to 15.5 Cam, and a top frequency of 10 000 Hz corresponds to 35.2 Cam.

The ratio of the number of channels is 35.2/15.5 or 2.27 and the square root is 1.51. From Table II, the internal noise in experiment 2 is $\sigma_N=0.0405$, and dividing it by 1.51 leads to an expected value of $\sigma_N=0.0269$ for experiment 1. According to Eqs. (5) and (7) this leads to an expected threshold ILD of 0.45 dB, in exact agreement with experiment 1.

Several details of the loudness meter model for lateralization invite comparison with details from the study of loudness *per se*. Table II shows that the ILD experiments require that the internal noise needs to be several times the stimulus noise. This result agrees with the conclusions of loudness discrimination experiments (e.g., Raab and Goldberg, 1975). As noted above, the ILD experiments with correlated noise show improved performance with increasing bandwidth that agrees with the statistical ideal. Ideal bandwidth dependence

was found by Green (1960). By contrast, Raab and Goldberg (1975) found only a weak dependence on bandwidth, and deBoer's (1966) model for discrimination actually considers the internal noise to be independent of bandwidth. Buus (1990) concluded that discrimination would improve with increasing bandwidth, as expected, for low-level noise but not for high. Applied to the bandwidth results in the present article, the Buus conclusions suggest that discrimination would improve when bandwidth is increased from 1 to 10 kHz, but not by as much as the ideal energy detector because of the upward spread of excitation in the 1-kHz case. What prevents us from reaching a firmer conclusion on this matter is that previous work has employed briefer noise bursts than ours and band-pass noises.

F. Alternative calculations and caveats

The compression exponent η was chosen to be 0.6 in the loudness-meter model because this is the value generally found in loudness experiments. This value of the exponent, combined with an integration time of 300 ms, led to good agreement with experimental data. However, alternative exponents are possible. We made a parametric study employing exponents of 0.2, 0.4, 0.6, 0.8, 1.0, 1.4, and 2.0. Although the values of Δ from Eq. (7) varied considerably, those variations were tracked by corresponding changes in σ_S from the numerical study of the variance of Δ. In the end, the calculations proved remarkably insensitive to the compression exponent. For instance, an exponent of 0.4, as suggested by the recent masking-level-difference study by Bernstein *et al.* (1999), led to good agreement with the measured thresholds when the integration time was increased to about 400 ms. Further, the predictions using an exponent of 0.4 and an integration time of 300 ms were not unreasonable in view of the uncertainty in the experimental thresholds.

A second, and more radical, reformulation of the model eliminated both the stages of half-wave rectification and compression. Instead, the calculation simply squared the input signals, $x_e(t)$, to obtain a power, and integrated the power to make a level-meter model with finite integration time, i.e., an *integrating level-meter model*. The purpose of this calculation was to allow us to separately assess the roles of temporal integration and of the other elements modeling the auditory system. The results of the calculation are easily described. Most important, the integrating level-meter model predicts no difference between correlated and anticorrelated noise—both have zero stimulus variance. Thus, this model does not fit the data in Fig. 5 as well as the level meter model, but it cannot be ruled out experimentally because the difference in ILD thresholds for correlated and anticorrelated noise did not reach statistical significance in any of our three experiments. Otherwise, the integrating level meter model fits the data for uncorrelated noise with an integration time of 240 ms, a not unreasonable value. Thus, in the end, the only aspect of the loudness meter model that is unequivocally indicated by our measured ILD thresholds is the temporal integration.

VI. CONCLUSIONS

The guiding hypothesis for this article was the level-meter model, which says that the human binaural system deals with interaural level differences (ILDs) for noise stimuli by making separate measurements of excitation from left and right ears and comparing those measurements at a higher level. According to the model, temporal fine structure of the noise is unimportant. Only the average excitation is important. Accordingly, the model predicts that ILD thresholds are the same whether the noise in the two ears is diotic-except-for-level (correlated), inverted (anticorrelated), or completely independent (uncorrelated). The model further predicts that ILD thresholds should be the same as intensity difference limens for diotic noise.

Experiments with noise having bandwidths of 10 and 1 kHz were performed to test the level-meter model. The results of the experiments can be briefly summarized by saying that the level-meter model works rather well. Although individuals differed from one another, thresholds for the four different experiment types agreed to within about half a dB. Therefore, for most practical purposes, one can rely on the level-meter concept.

At a finer degree of detail, the experiments showed some departure from the level-meter model. The largest difference was between the difference limen in intensity (DLI) and the ILD threshold for correlated noise. This difference might be due to specific binaural processes that contribute to ILD sensitivity but are excluded from diotic tasks like the difference limen. It might result from the different memory requirements of the two tasks.

Of greater interest were the differences among the ILD experiments themselves, with different noise coherence. Here, the differences were smaller. These differences could be explained by replacing the level-meter model by a loudness-meter model, incorporating temporal integration. The model further assumed that the total variance was the sum of internal noise and stimulus noise. Predictions of the model were in good agreement with the experimental results. The model also successfully accounted for the observed bandwidth dependence of ILD thresholds. An experimental observation that was not addressed by the model was that roving the level of the noise tended to increase thresholds (experiment 3 versus experiment 2). Averaged over listeners, the change was roughly the same for all experiment types, and it was never more than 0.2 dB. We tentatively attributed the change to a slight distraction caused by the rove.

To guard against misinterpretation, it is important to note that all the sounds considered in this article had similar amplitude spectra in left and right ears. A naive level-meter imposes no such requirement. It would lateralize equally well given a 100-Hz tone in one ear and a 5000-Hz tone in the other. The auditory system would not behave like this level meter because it would not fuse these tones into a single image. Although questions of spectral similarity have been deliberately avoided in this article, one imagines that one coherence-independent level meter per critical band might serve as an adequate model when different ears receive different spectral shapes.

There is recent animal physiological work that also sug-

gests that the processing of ILDs in broadband noise is independent of the binaural coherence of the noise. Egnor (2001) examined the posterior part of the ventral nucleus of the lateral lemniscus—the first stage of ILD processing in the barn owl. She found no difference in neural activity when coherent noise was replaced by incoherent noise. Further, the owls retained their vertical plane behavioral response when the coherence of the noise was changed. By contrast, replacing coherent noise by incoherent noise eliminated the ability of barn owls to localize in the horizontal plane (Saberi *et al.*, 1998). Because barn owls appear to use ILDs and ITDs separately to localize in separate planes, the insensitivity to coherence in ILD may not be surprising.

Tollin and Yin (2002) studied the spatial receptive field (SRF—neural firing rate as a function of azimuth angle) of LSO neurons in cat. For the single neuron investigated, they showed that the shape of the SRF was unchanged when broadband coherent noise was replaced by incoherent noise. The peak was reduced by 15% to 25%, however. To summarize, both these studies of other species are broadly consistent with the coherence insensitivity demonstrated in the present article.

The results of this work have significance for the localization of steady-state sounds in rooms. Because of reflections from room surfaces, the signals to the two ears are always incoherent to some degree. In a large reverberant environment they are uncorrelated above 500 Hz (Lindevald and Benade, 1986). It is not possible to use interaural time differences on binaurally uncorrelated signals because there is no common feature to time. By contrast, the present experiments show that binaural coherence has an almost negligible effect on the use of interaural level differences. It may be that standing waves in the room cause the ILDs to be misleading about the true location of the source, but the nervous system is entirely capable of making use of this information or disinformation. This result tends to focus attention on the ILD as an important element in sound localization in a room.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institute of Deafness and other Communicative Disorders. We are grateful to Dr. Brad Rakerd and Mary Jo Hidecker for help with the statistical evaluation of the results and to Dr. Leslie Bernstein, Dr. Steve Colburn, and an anonymous reviewer for useful comments.

[1]In the search for a replacement for listener K in experiment 2, three other listeners (male students C and X, and middle-aged female J) were tested and repeatedly failed to meet our criterion for tight staircases when the bandwidth was 1 kHz. Especially curious was the case of listener C. Listener C had among the lowest thresholds for a bandwidth of 10 kHz. He also had low thresholds for a bandwidth of 1 kHz for all noise conditions except correlated noise. However, C had great difficulty in lateralizing correlated noise with a bandwidth of 1 kHz. According to C's own description, he perceived a ''hole,'' which is an absence of noise. The hole moved opposite to the ILD-induced image. The effect was time dependent; C could perform well for most of a run before the hole would capture his attention, causing him to answer incorrectly 100 percent of the time, even at large ILDs. The authors introduced stimuli of shorter duration in order to reduce the salience of an ITD cue, but to no effect. Louder stimuli and the introduction of an interaural delay also failed to improve performance. The hole appeared less often when the low-pass cutoff was increased to 2 kHz, and it disappeared completely for a cutoff above 3 kHz. The authors have no explanation for any of this.

[2]Critical band numbers on the *Cam* scale refer to critical bands measured in Cambridge using notched-noise measurements (Glasberg and Moore, 1990; Moore, 1995). The Cam units are identical to the units regrettably called ERBs in much of the *fin de siecle* psychoacoustical literature.

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). ''The normalized interaural correlation: Accounting for $NoS\pi$ thresholds obtained with Gaussian and 'low-noise' noise,'' J. Acoust. Soc. Am. **106**, 870–876.

Blauert, J. (**1983**). *Spatial Hearing*, translated by J. S. Allen (MIT, Cambridge, MA).

Buus, S. (**1990**). ''Level discrimination of frozen and random noise,'' J. Acoust. Soc. Am. **87**, 2643–2654.

deBoer, E. (**1966**). ''Intensity discrimination of fluctuating signals,'' J. Acoust. Soc. Am. **40**, 552–560.

Dunn, O. J. (**1961**). ''Multiple comparisons among means,'' J. Am. Stat. Assoc. **56**, 52–64.

Egnor, R. (**2001**). ''Effects of binaural decorrelation on neural and behavioral processing of interaural level differences in the barn owl (Tyto alba),'' J. Comp. Physiol., A 589–595.

Fletcher, H. (**1953**). *Speech and Hearing in Communication*, edited by J. B. Allen (Acoustical Society of America, New York, 1995) (reprint).

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched noise data,'' Hear. Res. **47**, 103–138.

Grantham, D. W., and Ahlstrom, J. B. (**1982**). ''Interaural intensity discrimination of noise as a function of center frequency, duration, and interaural correlation,'' J. Acoust. Soc. Am. Suppl. 1 **71**, S86.

Green, D. M. (**1960**). ''Auditory detection of a noise signal,'' J. Acoust. Soc. Am. **32**, 121–131.

Green, D. M. (**1988**). *Profile Analysis Auditory Intensity Discrimination*, Oxford Psychology Series No. 13 (Oxford, New York), p. 20.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (Springer-Verlag, New York), Appendix I.

Henning, G. B. (**1974**). ''Detectability of interaural delay in high-frequency complex waveforms,'' J. Acoust. Soc. Am. **55**, 84–90.

Jesteadt, W., and Wier, C. C. (**1977**). ''Comparison of monaural and binaural discrimination of intensity and frequency,'' J. Acoust. Soc. Am. **61**, 1599–1603.

Lindevald, I. M., and Benade, A. H. (**1986**). ''Two-ear correlations in the statistical sound fields of rooms,'' J. Acoust. Soc. Am. **80**, 661–664.

Moore, B. C. J. (**1995**). ''Frequency analysis and masking,'' in *Hearing, Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, San Diego).

Neutzel, J. M. (**1982**). ''Sensitivity to interaural intensity differences in tones and noise as measured with a roving level procedure,'' J. Acoust. Soc. Am. Suppl. 1 **71**, S47.

Plomp, R., and Bouman, M. A. (**1959**). ''Relation between hearing threshold and duration for tone pulses,'' J. Acoust. Soc. Am. **31**, 749–758.

Raab, D. H., and Goldberg, I. A. (**1975**). ''Auditory intensity discrimination with bursts of reproducible noise,'' J. Acoust. Soc. Am. **57**, 437–447.

Saberi, K., Takahashi, Y., Konishi, M., Albeck, Y., Arthur, B. J., and Farahbod, H. (**1998**). ''Effects of decorrelation on neural and behavior detection of spatial cues,'' Neuron **21**, 789–798.

Stevens, S. S. (**1955**). ''The measurement of loudness,'' J. Acoust. Soc. Am. **27**, 815–829.

Tollin, D. J., and Yin, T. C. T. (**2002**). ''The coding of spatial location by single units in the lateral superior olive of the cat, I. Spatial receptive fields in azimuth,'' J. Neurosci. **22**, 1454–1467.

von Békésy, G. (**1930**). ''Zur Theorie des Hörens: Über das Richtungshören bie einer Zeitdifferenze oder Lautstärkeungleichheit der beidseitigen Schalleinwirkungen,'' Phys. Z. **31**, 824–838, 857–868; cited by Blauert (1983), p. 162.

Yost, W. A. (**1981**). ''Lateral position of sinusoids presented with interaural intensive and temporal differences,'' J. Acoust. Soc. Am. **70**, 397–409.

Temporal weighting in sound localization[a)]

G. Christopher Stecker[b)] and Ervin R. Hafter
Department of Psychology, University of California, Berkeley, California 94720

(Received 18 October 2001; revised 29 May 2002; accepted 6 June 2002)

The dynamics of sound localization were studied using a free-field direct localization task (pointing to sound sources) and an observer-weighting analysis that assessed the relative influence of each click in a click-train stimulus. In agreement with previous studies of the precedence effect and binaural adaptation, weighting functions showed increased influence of the onset click when the interclick interval (ICI) was short (<5 ms). For longer ICIs, all clicks in a train contributed roughly the same amount to listeners' localization responses. Finally, when a short gap was introduced in the middle of a train, the influence of the click immediately following the gap increased, in agreement with the ''restarting'' results obtained by Hafter and Buell [J. Acoust. Soc. Am. **88**, 806–812 (1990)]. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1497366]

PACS numbers: 43.66.Qp, 43.66.Mk [LRB]

I. INTRODUCTION

Sound localization in the natural world is based on a variety of cues including interaural-time (ITD) and -level (ILD) differences, as well as spectral cues produced by the direction-dependent filtering of sound by the head, shoulders, and pinnae. The interaural differences provide cues to azimuth (direction in the horizontal plane), whereas spectral cues, characterized by the directional transfer function (DTF), provide monaural information that is especially useful in vertical localization. Each of these cues is susceptible to distortion by environmental factors, such as the presence of echoes, reverberation, and competing sources. However, even in some highly reverberant spaces, listeners are relatively unaffected by echoes in their ability to localize sound sources. In some cases, this ability may be partly attributable to the availability of redundant spatial cues, but an additional factor is the perceptual dominance of spatial cues contained in a stimulus onset—which are unaffected by the presence of echoes—over those contained in later portions (Zurek, 1980). This dominance is exhibited in a relatively large and well-studied class of phenomena shown using a variety of different approaches, and known variously as the ''precedence effect'' (Wallach *et al.*, 1949), ''Haas effect'' (after Haas, 1972), ''law of the first wavefront'' (Blauert, 1983), ''binaural adaptation'' (Hafter, 1997), ''echo suppression'' (Clifton, 1987), and occasionally ''onset dominance'' (Freyman *et al.*, 1997). A recent review can be found in Litovsky *et al.* (1999). Throughout the remainder of this paper, we use the descriptive term, ''onset dominance,'' to refer to the general phenomenon of increased influence of onsets in spatial hearing.[1] The goal of the current study is to investigate onset dominance by developing a temporal weighting function for the localization of click-train stimuli presented in the free field.

Three types of stimuli have been used to estimate the temporal extent of onset dominance in past studies. In the first, paired stimuli (for example, paired clicks or noise bursts) are presented with a delay between the first (lead) and second (lag) stimulus. Lead and lag are presented with different spatial or intracranial positions, and listeners are asked to make spatial judgments regarding the lagging stimulus or the combined (fused) image of both lead and lag. Commonly employed in studies of the precedence effect (Litovsky *et al.*, 1999), this method reveals a temporary reduction in spatial sensitivity from approximately 1 to 10 ms following the leading stimulus (Zurek, 1980). A second approach, employed by Hafter and colleagues (Hafter and Buell, 1990; Hafter *et al.*, 1988b; Hafter and Dye, 1983) compares spatial discrimination performance for stimuli of different durations, where the different portions of each stimulus (clicks in a train) present redundant spatial information to the listener. While lateralization performance generally improves with stimulus duration—as expected if listeners respond based on pooled information from all clicks—improvement for high-rate stimuli [interclick interval (ICI) shorter than approximately 12 ms] is suboptimal, as if later clicks are less effective than earlier clicks. In modeling this effect, termed ''binaural adaptation,'' Hafter and Dye (1983) showed that the number of informative events (i.e., the effective number of clicks) available to a listener, N, is a compressive power function of the number of acoustic clicks, n

$$N \propto n^k, \quad \text{where} \quad 0 \leqslant k \leqslant 1, \quad \text{and} \quad k = f(\text{ICI}). \tag{1}$$

If one assumes that the relative effectiveness (or ''weight'') of individual clicks declines monotonically following the stimulus onset, then the weight (w_j) on each click j can be estimated by calculating the finite difference of Eq. (1)

$$w_j = \Delta N_j \propto (j)^k - (j-1)^k, \quad \text{for} \quad j = \{1,2,...,n\} \tag{2}$$

(Hafter and Buell, 1990; Hafter *et al.*, 1983). The exponent k of Eq. (1) is ICI dependent, with $k \approx 1$ (optimal use of all

[a)]Portions of this work have appeared in the first author's doctoral dissertation and in a poster presentation given at the 24th Annual Midwinter Research Meeting of the Association for Research in Otolaryngology in February 2001.

[b)]Current address: Kresge Hearing Research Institute, University of Michigan Medical Center, Ann Arbor, MI 48109. Electronic mail: cstecker@umich.edu

clicks) at ICIs longer than approximately 12 ms; k (and consequently, N) grows smaller and smaller as ICI approaches 2 ms. Below this value, $k \approx 0$ and performance for trains of up to 32 clicks is hardly better than for single clicks (Hafter *et al.*, 1988b). Thus, weighting functions estimated by Eq. (2) show a monotonic decline in the effectiveness of clicks over the course of a stimulus, with the slope of decline related to ICI.

Both of the above methods estimate temporal weighting functions by comparing the performance levels achieved with stimuli of different overall durations. In the precedence method, effects of stimulus rate and duration are confounded. The subtractive method described by Hafter *et al.* (1988b) has the advantage of estimating the influence, or ''perceptual weight,'' of each click in an extended stimulus of a given rate. However, there are indications that the assumptions underlying Eq. (2) may not hold for all stimulus arrangements (see, e.g., Saberi, 1996), possibly because localization judgments reflect the retroactive evaluation of spatial information carried by all parts of the stimulus.

A more direct approach is to estimate weights for each portion of an extended stimulus independently; this can be accomplished using observer-weighting analyses (Ahumada and Lovell, 1971; Berg, 1989; Saberi, 1996; Shinn-Cunningham *et al.*, 1993; Stellmack *et al.*, 1999). These techniques were developed to help ascertain the relative influence of multiple stimulus components on a subject's perception or psychophysical performance. In short, observer-weighting analyses (see, e.g., Berg, 1989) relate random variation of a number of independent stimulus components to variation in subject responses (e.g., detection, scaling, etc.). Stimulus components that, when varied, induce systematic changes in the response are assigned high weights by the analysis, while those which do not affect the response receive low weights.

In the context of onset dominance in sound localization, observer weighting was used by Shinn-Cunningham *et al.* (1993, 1995) to estimate the relative influence of leading and lagging noise bursts, presented over headphones, on the perceived lateral position of their combined image. The analysis revealed high weights for the lead and correspondingly low weights on the lag—indicating onset dominance—over lead–lag delays of 1 to 10 ms, as expected based on earlier studies of the precedence effect. Similarly, Stellmack *et al.* (1999) used observer-weighting analysis in a task where listeners were asked to discriminate the lateral position of lead or lag clicks (i.e., not the combined image), again presented over headphones. They found high lead weights at short delays (1–4 ms), regardless of whether subjects were asked to judge the position of the lead or lag click.

Two previous studies using observer-weighting techniques to estimate temporal weighting functions for the lateralization of extended stimuli were completed by Saberi (1996) and Dizon *et al.* (1998). Saberi (1996) estimated weighting functions for trains of 2–16 clicks, similar to stimuli employed by Hafter and Dye (1983). Saberi's study employed a lateralization task in which listeners were asked to identify each stimulus as having clicks drawn from one of two normal distributions of ITD. The two distributions were centered at left-leading and right-leading ITD values corresponding to each listener's discrimination threshold (defined as 75% correct performance), with standard deviations of 100 μs. Each click in a train possessed an ITD drawn at random from that trial's distribution, and relative weights were computed for individual clicks. For ICIs of 1.8–12 ms, the first click received consistently higher weight than did later clicks; this effect was somewhat diminished at long ICIs and for long trains. Except at an ICI of 12 ms, weights did not decrease monotonically over the duration of the stimulus, as would be expected based on the model of Hafter and Buell (1990). Rather, weights were reduced to a constant level immediately following the onset (click 1).

Dizon *et al.* (1998) estimated similar weighting functions for broadband noises with varying ITD. Each stimulus was divided into 4–6 temporal ''slices,'' each with an ITD chosen at random from a discrete distribution of five values spanning the range from -400 μs (left-leading) to $+400$ μs (right-leading). As in the studies of Shinn-Cunningham *et al.* (1993, 1995), subjects estimated the perceived lateral position of the fused image, and linear regression was used to calculate weights for the different slices. Slice duration was varied from 2 to 10 ms as an experimental parameter. The results revealed high weight for the first slice, regardless of slice length, and approximately equal (and low) weights for the remaining slices. This pattern of weights matches closely the functions obtained by Saberi (1996), despite differences in procedure (clicks vs continuous noise, identification vs adjustment).

As described earlier, real-world sound localization is based on a combination of acoustic cues, including ITD, ILD, and the DTF. The observer-weighting studies described above employed only headphone listening, manipulating only ITD as a spatial cue. Manipulation of a single parameter in this manner helps to simplify experimental designs and the identification of potential mechanisms. However, we are ultimately interested in extrapolating these findings to real-world listening, where all three cue types are present. In this respect, pure-ITD stimuli are not satisfactory, because they are not mere simplifications of natural stimuli; rather, they present ITD- and ILD cues which are in conflict with each other and with spectral cues derived from the DTF.

Previous work has shown that binaural adaptation affects the processing of ILD cues similarly to that of ITD cues (Hafter *et al.*, 1983), even in situations where the two cue types are presented together (Hafter *et al.*, 1990). Similar results were found by Hafter *et al.* (1988a) to hold for free-field stimuli. However, there are some indications of important differences between pure-ITD and free-field listening with regard to precedence-like effects. Blauert *et al.* (1989), for instance, found no evidence of an active ''restarting'' phenomenon—as observed by Hafter and Buell (1990) under headphone-listening conditions—in the free-field precedence effect. In addition, the findings of Rakerd and Hartmann (1985) suggest that the presence of echoes alters the way in which listeners make combined use of ITD and ILD cues for localization.

Here, we present sounds in the free field via loudspeakers. In that manner, the ITDs, ILDs, and spectral cues are

related in a natural fashion and are consistent with the individual subjects' everyday listening experiences. To obtain localization judgments, we employed a direct localization (pointing) task. This task was similar to the adjustment procedures used by Shinn-Cunningham *et al.* (1993, 1995) and Dizon *et al.* (1998), but utilized a visual pointer in the free field rather than an acoustic one presented over headphones.

II. EXPERIMENTAL METHODS

A. Subjects

Subjects included the first author (CS) and four paid subjects (HW, LL, LS, and TL) naive to the purpose of the experiments. All subjects had normal audiograms from 125–8000 Hz. Not all subjects participated in all experiments or conditions. Later sections indicate the particular subjects involved in each experiment.

B. Stimuli

Following previous work (Hafter and Dye, 1983; Saberi, 1996), stimuli throughout this study consisted of trains of high-frequency narrow-band clicks (Gaussian-windowed tone bursts), sampled at 50 kHz. Carrier frequency was fixed at 4 kHz and the Gaussian envelope, centered on a peak of the carrier waveform, had a total duration of 2 ms (measured at the points where the Gaussian window function falls below the limits of 16-bit truncation). Duration measured at $\pm 1\sigma$ was 0.6 ms. The measured bandwidth (at -3 dB) of the click was approximately 900 Hz. Trains of 2 or 16 clicks were synthesized with interclick intervals (ICI), defined as time between click peaks, of 1, 3, 5, 8, or 14 ms. At 1-ms ICI, the 2-ms Gaussian envelopes overlap for half their total duration, but cross over at a point 59.9 dB below the peak of either envelope, resulting in negligible overlap in energy. Different stimuli were presented at equal absolute levels, so that SPL at the listener's position varied systematically as a function of ICI, from 32 dB at 14 ms ICI to 39 dB at 1 ms ICI. All stimuli were clearly audible for all subjects.

C. Setup and arrangement

As depicted in Fig. 1, listeners were seated in an anechoic chamber (Eckels Corp., $8.3\times5.4\times4.0$ m), facing an array of 12 loudspeakers (Audax model MHD12P25 FSM-SQ) placed at ear level along the left, right, and front walls. Loudspeakers were spaced 5.5° apart in listener-centered azimuth, with the center-most loudspeakers placed 2.75° to the left and right of the listener's midline. Stimuli were delayed and attenuated such that all loudspeakers produced sounds at the listener's position that were equalized in level and aligned in time. The delays and attenuations simulated a circular array of loudspeakers located 6.1 m from the listener's location. During the experiment, the loudspeakers were obscured visually by an acoustically transparent white curtain hung 1.07 m in front of the listener. A laser pointer was mounted near the listener's right hand—approximately 56 cm to the right, 30 cm below, and 15 cm in front of the listener's intracranial midpoint—on a pair of high-precision potentiometers allowing free rotation in both azimuth and elevation. It projected a bright spot of red light upon the curtain, which listeners used to make localization responses (see below). The laser position was recorded digitally by sampling the potentiometer settings with a pair of 8-bit analog-to-digital converters. Note that, since the curtain was hung in a straight line across the room, the accuracy of laser pointer readings was not equal across the entire field; away from the midline, a given angular rotation of the pointer produced a larger displacement of the point in head-centered coordinates. However, because all responses were transformed to head-centered coordinates for analysis, this distortion did not act to systematically bias responses; rather, the reduced pointer accuracy results in somewhat increased variance in responses away from the midline. This should probably not be a concern, especially considering that auditory spatial acuity is reduced in those regions as well (Mills, 1958). The room was lit by two 25-watt soft-white bulbs mounted near the ceiling to either side of the listener. The space beyond the curtain, including the loudspeakers, was darkened.

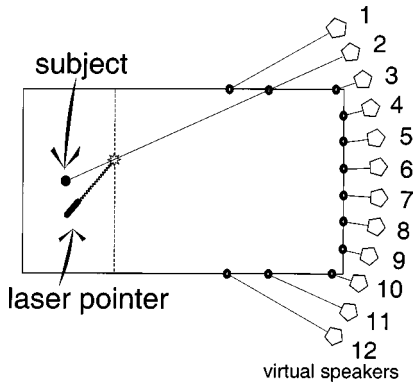


FIG. 1. Arrangement of the anechoic chamber used in the current study, drawn to scale. As discussed in the text, experiments took place in an $8.3\times5.4\times4.0$-m anechoic chamber. The subject was seated in the center of one end of the room; an array of 12 loudspeakers was positioned along the walls of the opposite end. Loudspeakers were spaced 5.5° apart, relative to the listener's position. Stimuli were delayed and attenuated so as to simulate a circular array of loudspeakers with a radius of 6.10 m, centered on the listener. Actual distances ranged from 3.43 (speaker 12) to 5.61 m (speakers 4 and 9). Loudspeakers were visually obscured by an acoustically transparent curtain placed 1.07 m in front of the listener. This curtain also acted as a target for the hand-guided laser pointer, mounted next to the subject. Here, the subject is shown pointing to speaker 2 (-24.75°).

The spectral characteristics of individual loudspeakers were equalized through digital inverse filtering. Each day, impulse responses from each loudspeaker were recorded digitally and used to construct time-domain inverse filters that produced effectively ''flat'' spectral responses in phase and level (± 1 dB in the range 2–6 kHz).

D. Stimulus presentation and listeners' task

A stimulus location, θ_L, was chosen at random on each trial. This location defined the center of a group of three or five loudspeakers that presented the individual click stimuli. The stimulus itself was a train of 2 or 16 clicks, depending on the experiment. Each click in a train was presented from a single loudspeaker, selected at random from within the

TABLE I. Loudspeaker conditions.

Condition	Stimulus locations	Stimulus azimuth (θ_L)	Speaker group (relative to θ_L)
N3	2,...,11	−24.75°,..., +24.75°	−5.5°, 0°, +5.5°
W3	3,...,10	−19.25°,..., +19.25°	−11°, 0°, +11°
W5	3,...,10	−19.25°,..., +19.25°	−11°, −5.5°, 0°, +5.5°, +11°

group. This random variation of location was necessary for the computation of observer weights for each click (see Sec. III).

Three conditions defining the placement of loudspeakers in a group were employed. The conditions, summarized in Table I, are denoted N3, W3, and W5. In each case, the letter indicates the range of azimuths around θ_L, either ±5° (''narrow,'' for N3) or ±11° (''wide,'' for W3 and W5). The number (3 or 5) indicates the number of loudspeakers in the group. For example, in condition W3, clicks were presented from one of three loudspeakers on a given trial; these were located at θ_L, $\theta_L-11°$, or $\theta_L+11°$ azimuth. Because we used a fixed array of loudspeakers, some values of θ_L at the far ends of the array were not achievable in each condition; ranges of achievable stimulus azimuths are also shown in Table I.

Figure 2 shows a timeline of two hypothetical trials that might appear in condition W3. In the first trial, location 4 ($\theta_L=-13.75°$) has been selected as the stimulus location for presentation of a 16-click train. Individual clicks are presented from speakers 2, 4, and 6 (−24.75°, −13.75°, and −2.75°). The stimulus on trial 2 is presented from location 7 ($\theta_L=+2.75°$), and individual clicks are delivered to speakers 5, 7, and 9 (−8.25°, +2.75°, and +13.75°).

The listener's task on each trial was to point to the location of the stimulus with the laser pointer. The subject guided the pointer using the right hand and held a small response box in the left hand. At the beginning of the trial, subjects were instructed to face forward and maneuver the laser pointer so that its spot was located in the ''home'' position, above 28° elevation and between −16° (to the right) and +17° (to the left) azimuth. In the home position, the spot was at or above the top of the curtain, and just beyond the subject's gaze. A stimulus was presented from one of eight or ten potential locations, following which subjects were instructed to direct their gaze, without moving their heads, to foveate the perceived location. They were not told to expect stimuli with multiple apparent locations but, in the event that more than one acoustic image was perceived, they were to respond to the leftmost image. This instruction ensured that when subjects perceived multiple images (as expected, e.g., at long ICIs), the weights were not biased by listeners' strategies to favor early or late clicks. Rather, since all clicks appeared in the leftmost position with equal probability, the appearance of multiple images should have produced equal weights on all clicks. Next, they were to maneuver the laser pointer to project its spot directly at the point of their gaze and to record the location by pressing a button on the response box. The positions of both potentiometers were recorded and transformed to coordinates of the listener's gaze for analysis. Following the response, subjects returned the laser spot to the home position and, following a 1-s delay, the next trial began. Each experimental run consisted of 100 uninterrupted trials. Subjects were allowed to take breaks between runs, and completed between 4 and 16 runs in a given condition.

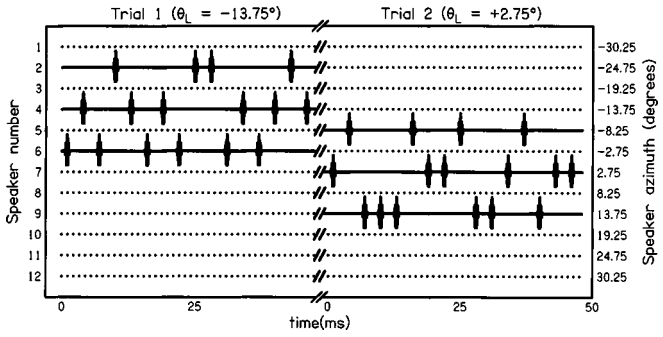


FIG. 2. Timelines of two hypothetical trials in condition W3. On the left, a 16-click train is presented in trial 1 with $\theta_L=-13.75°$; this corresponds to speaker 4. Individual clicks are presented from the group of speakers 2, 4, and 6 ($\theta_i\in\{-24.75°,-13.75°,-2.75°\}$). On the right, trial 2 is presented with $\theta_L=+2.75°$ (speaker 7); individual clicks are presented from the group of speakers 5, 7, and 9 ($\theta_i\in\{-8.25°,+2.75°,+13.75°\}$). Trials in condition N3 present individual clicks from three speakers 5.5° apart, while trials in condition W5 present individual clicks from five speakers separated by 5.5°.

III. ANALYTICAL METHODS

A. Regression model

The listener's task was to indicate a single location belonging to a group of stimuli with potentially disparate locations. Regardless of whether listeners *actually perceived* a single location for this type of stimulus, they had to respond in a way that combined the locations of the individual clicks. The analysis used here assumes the following linear combination:

$$\hat{\theta}_R=\sum_{i=1}^{n} w_i\theta_i+C, \tag{3}$$

where $\hat{\theta}_R$ is the (predicted) response location in azimuth, θ_i is the azimuth of the ith click (of n total clicks), w_i is the perceptual *weight* applied to the ith click, and C is a constant that reflects overall bias in response locations. Because each click provides potentially usable information for the task, a reasonable listening strategy would place equal weight on each click (i.e., all w_i are equal and responses indicate the mean location of all clicks). For a discrimination task (e.g., if subjects were asked to discriminate θ_L on two presentations) this strategy is not merely reasonable, but ''optimal,'' in information-theoretic terms (assuming equality and independence of the performance-limiting noise associated with each click) (Saberi, 1996). In contrast, when not all w_i are equal, a suboptimal strategy, favoring some clicks over others, is indicated. Since the localization task used here has no ''correct'' answer, no strategy can be considered optimal in quite this way; however, equal weighting serves as a useful null hypothesis to which obtained weighting patterns can be compared.

B. Weight calculation

Using response azimuth (θ_R) as the dependent variable and the azimuths of the n individual clicks ($\theta_i, i=1,...,n$) as independent predictor variables for multiple linear regression, a least-squares fit to Eq. (3), minimizing $(\theta_R-\hat{\theta}_R)^2$, was computed for each combination of subject, stimulus condition, and ICI. Regression coefficients, also known as "beta weights," obtained from these analyses provided estimates for w_i. Normalized weights

$$\dot{w}_i=\frac{w_i}{\Sigma_{j=1}^{n} w_j}, \tag{4}$$

were then computed to provide a relative weighting function for each combination of subject, condition, and ICI. While the raw weights provide a meaningful interpretation (degrees shift in response per degree shift in click location), normalized weights stress the *relative* influence of each click and also allow weights to be averaged across subjects (since $\Sigma_{\dot{w}_i}=1$ for each subject, by definition). Normalized weights are plotted throughout this paper, along with 95% confidence intervals (see the Appendix).

IV. EXPERIMENT 1: EFFECTS OF ICI AND SPATIAL SEPARATION ON WEIGHTING FUNCTIONS FOR 2-CLICK TRAINS

Experiment 1 used pairs of clicks to study the localization-dominance aspect of the precedence effect (Litovsky *et al.*, 1999) with a direct-localization task. While the spatial separation of stimuli was somewhat smaller (5.5°–22° azimuth) than in many studies of localization dominance, the stimuli were designed to be similar to those used in previous work and spanned a comparable range of ICI (1–14 ms). The results of this experiment are compared to those obtained using observer-weighting paradigms for lateralization (Dizon *et al.*, 1998; Saberi, 1996; Shinn-Cunningham *et al.*, 1993; Stellmack *et al.*, 1999).

A. Methods

Four subjects participated in this experiment. Two (CS, LL) listened to stimuli presented in condition N3 and three (CS, HW, TL) listened to stimuli presented in condition W3.

The ICI for trains of 2 clicks varied from 1 to 14 ms. Each subject completed four runs of 100 trials at each ICI, except subject TL, who listened only to ICIs of 3 and 8 ms.

B. Results

Figure 3 presents the mean weighting functions (across subjects) obtained for click pairs in both conditions. For click pairs, $\dot{w}_2=1-\dot{w}_1$, so the normalized click weight $\dot{w}_1$ is plotted on its own, as a function of ICI. At the shorter ICIs (1–5 ms), weight assigned to click 1 was significantly greater than that assigned to click 2 (i.e., $\dot{w}_1>0.5$). There were no significant differences between the weights obtained in conditions N3 (filled circles) and W3 (open circles), except at the shortest ICI (1 ms), where the difference can be explained by differences in the degree of onset dominance exhibited by different subjects at this value of ICI (see Fig. 4). Smaller symbols plot comparable weights obtained in four other observer-weighting studies (Saberi, 1996; Shinn-Cunningham *et al.*, 1993, 1995; Stellmack *et al.*, 1999).

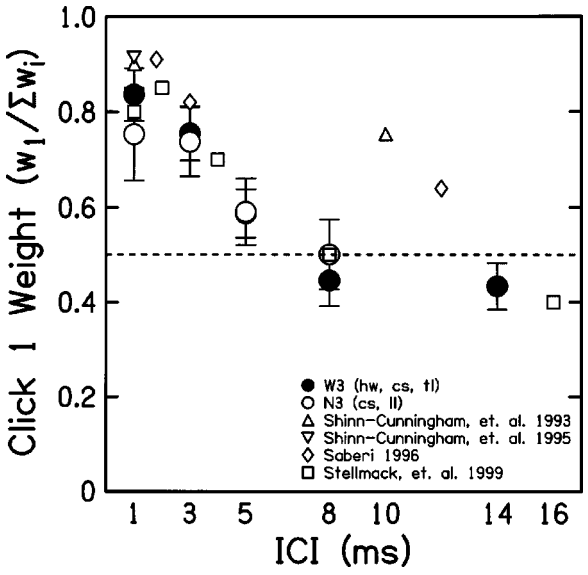


FIG. 3. Relative weights obtained for click pair stimuli. Normalized click 1 weights ($\dot{w}_1$) are plotted as a function of ICI (weights for click 2 are given by $1-\dot{w}_1$), for loudspeaker conditions W3 (filled circles, three speakers spanning 22°) and N3 (open circles, three speakers spanning 11°). Error bars indicate mean 95% confidence intervals on the weight estimates (see the Appendix). Dashed line indicates the value (0.5) of weights expected if all clicks had equal influence on the response (an "optimal" strategy). Diamonds and squares plot equivalent weights measured by Saberi (1996) and Stellmack *et al.* (1999), respectively, while triangles present those measured by Shinn-Cunningham *et al.* (1993) and Shinn-Cunningham *et al.* (1995) using paired bursts of white noise (upward-facing triangles) and narrowband noise centered at 1450 Hz (downward-facing triangles).

Figure 4 summarizes the degree of onset dominance, or precedence, observed for individual subjects. Plotted is the precedence ratio (*pr*), defined by Saberi (1996)

$$pr=\frac{w_1}{\Sigma_{i=2}^{n} w_i}. \tag{5}$$

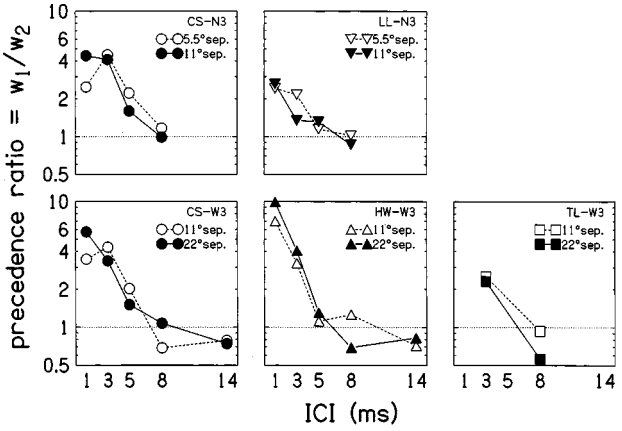


FIG. 4. Precedence ratios for each subject, comparing trials separated by 5.5/11 and 11/22 deg. The precedence ratio of Eq. (5) is plotted as a function of ICI, separately for trials with different angular separations between click 1 and 2 (open versus filled symbols). Separate panels plot data obtained from individual subjects. Upper panels show ratios calculated under condition N3, lower panels plot ratios for condition W3. In all panels, dotted lines indicate the ratios that would be obtained if clicks were weighted equally ($w_1=w_2, pr=1$).

For 2-click trains, $pr = w_1/w_2$. This ratio indicates the relative influence of the onset and later-arriving portions of the stimulus on the localization responses. Larger values indicate a stronger influence of the onset, and a value of 1 indicates equal weight between the onset click and the remainder of the stimulus. Precedence ratios are plotted against ICI in Fig. 4. A logarithmic scale is used to avoid overemphasizing large values of *pr* and to more clearly visualize the trend of *pr* values declining with increasing ICI. Separate lines indicate trials with different spatial separations between loudspeakers. Since the positions of clicks 1 and 2 were selected randomly from three possibilities (relative to θ_L) on each trial, there were three possible separations on any given trial: 0°, 5.5°, or 11° in condition N3 and 0°, 11°, or 22° in condition W3. By comparing performance on W3 trials separated by 11° with those trials separated by 22° (for instance), we determined whether precedence differed between trials with wide or narrow separations completed by individual subjects. As can be seen from the figure, there were no systematic differences between the separations. Especially at short ICI, however, large differences between the precedence ratios were calculated for different subjects, with HW (condition W3) showing very large *pr* values compared to subject LL (condition N3). It seems likely that the difference between conditions N3 and W3 apparent in Fig. 3 reflects this intersubject difference rather than an effect of spatial separation. In comparison, subject CS—who listened in both conditions—showed similar *pr* values in both conditions.

C. Discussion

The primary finding of this experiment is that localization was dominated by the first click at short ICIs but that weights for the two clicks were approximately equal at longer ICIs (8–14 ms). Recall that subjects were instructed to point to the leftmost auditory image when more than one image was apparent. Consistent pointing to the leftmost click would cause subjects to point to the lead on 50% of trials; equal weights reflect the fact that both clicks are equally likely to be in the leftmost position. If instructed differently, subjects would likely have been able to accurately localize both clicks at long ICIs. These results are in fair quantitative agreement with those obtained in other studies using headphone stimulation. As shown in Fig. 3, weights obtained by Shinn-Cunningham *et al.* (1993, 1995), Saberi (1996), and Stellmack *et al.* (1999) were similar to those measured in this experiment, except at longer delays, where the results of Saberi (1996) and Shinn-Cunningham *et al.* (1993) indicate stronger precedence effects than those of the current study or Stellmack *et al.* (1999). Shinn-Cunningham *et al.* (1993) also conducted a meta-analysis of lead weights estimated from previous studies of the precedence effect using discrimination measures (Gaskell, 1983; Saberi and Perrott, 1990; Zurek, 1980). These tended to decline more gradually (i.e., $\dot{w}_1$ remained above 0.7 for ICIs around 8–10 ms) than weights measured in matching tasks. The results of this experiment are in agreement with that observation, suggesting that precedence at intermediate delays (~5–10 ms) may be weaker for localization dominance than for lag discrimination.

Spatial separation was found to have no significant effect on the weighting functions. Though consistent with the results of Yang and Grantham (1997), this result contrasts with that of Shinn-Cunningham *et al.* (1993), in which some subjects showed separation-dependent weighting functions at long lead–lag delays and high stimulus levels. Considering the magnitude of interaural separations used in that study (ΔITD up to ± 1 ms) relative to the equivalent angular separations used here—a maximum angular separation of 22° corresponds roughly to ΔITD of 150–200 μs (Kuhn, 1987)—it may be that very large separations act to reduce precedence in a way that narrower separations do not.

A final feature of the weighting functions plotted in Fig. 3 bears discussion: there is a consistent tendency at the longer ICIs (8–14 ms) for a larger weight on click 2 than on click 1. The difference is small, but it is statistically significant in condition W3. This result may reflect a general bias of subjects to point toward the later stimulus, perhaps through the perception (occasionally reported by subjects) of apparent motion. It may also relate to the increase of weights observed toward the end of the stimulus in experiments 2–4, discussed below. Similar effects were seen by Stellmack *et al.* (1999), who measured elevated echo weights at ICIs from 16–32 ms even when subjects were asked to judge the location of the lead click. Also, ITD-threshold data collected by Tollin and Henning (1998) show some indication that a diotic lag click interferes with discrimination of the lead's ITD as ICI increases from 0.8 to 12.8 ms, suggesting increased influence of the lag at these delays.

V. EXPERIMENT 2: EFFECTS OF ICI ON WEIGHTING FUNCTIONS FOR 16-CLICK TRAINS

While studies of the precedence effect have generally used pairs of stimuli with a variable delay (''lead'' and ''lag''), studies of onset dominance and binaural adaptation have used single stimuli with extended durations. Experiment 2 examined the form of localization weighting for trains of 16 bandlimited clicks, similar to stimuli used by Saberi (1996) and Hafter and Dye (1983) to measure binaural adaption.

A. Methods

All five subjects (CS, LL, LS, TL, and HW) participated in this experiment. Not all subjects were tested at all ICIs; specifically, TL was not tested at 8 ms ICI and LL was not tested at 5 ms ICI.

Stimuli were trains of 16 clicks, as described in Sec. II B. Each subject completed 4 runs of 100 trials at each of four ICIs (3, 5, 8, and 14 ms). Loudspeaker condition W3 defined the spatial layout of stimuli. All other aspects of the experimental procedure, stimulus presentation, and analysis were as described in the section on general methods. Aside from differences in stimuli used and listeners participating, this experiment was identical to experiment 1.

B. Results

Normalized weighting functions, averaged across subjects, appear in Fig. 5. Weights for click 1 were significantly

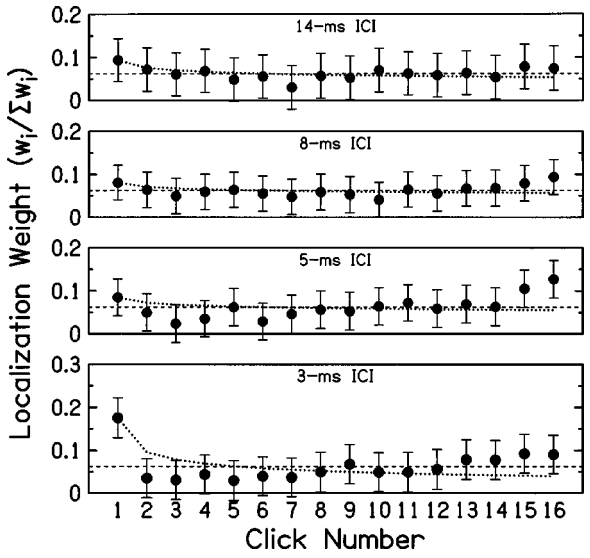


FIG. 5. Normalized weights for 16-click trains. Plotted weights are the mean of four subjects per condition, with each panel showing data for a single value of ICI. Error bars indicate means of normalized 95% confidence intervals (see the Appendix), as in Fig. 3. Dashed lines indicate equal weights for all clicks. At short ICI, the results show a significant increase in the weight for click 1 relative to the remaining weights, while weights are approximately even for longer ICIs. The increase is not accompanied by monotonically decreasing weights for clicks following the first, as in the relative effectiveness of individual clicks estimated by Hafter and Buell (1990) [dotted lines, from Eq. (2) fit with k equal to 0.85, 0.91, 0.89, and 0.63 for ICIs of 14, 8, 5, and 3 ms, respectively]. Rather, suppression of the weights for later clicks seems to be rather abrupt. Additionally, there is a tendency for weights to recover over the duration of the stimulus, resulting in somewhat elevated weights toward the stimulus offset.

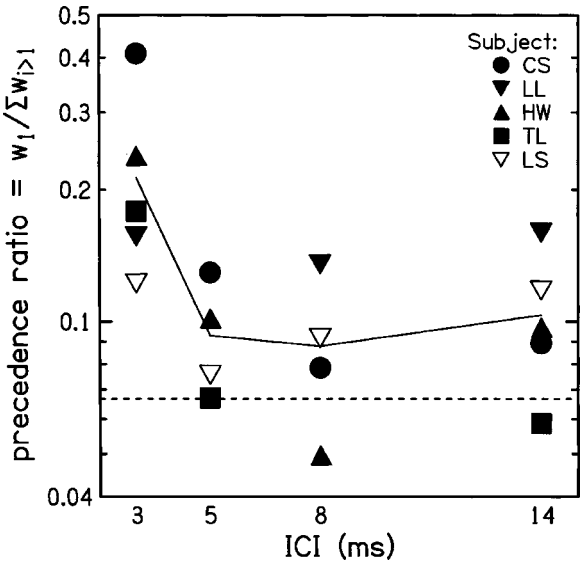


FIG. 6. Precedence ratio for 16-click trains, plotted as a function of ICI. Ratios for individual subjects are indicated by symbol type; the solid line indicates mean ratio across subjects. The dotted line indicates the ratio that would be obtained if all clicks received equal weight. Two aspects of the figure should be noted: First, precedence ratios are largest for 3-ms ICI, and decrease as the ICI is extended. Second, the average ratio is slightly greater than 1, even for long ICIs, indicating that the onset has a somewhat stronger influence than later clicks even at slower rates.

larger than weights for clicks 2–16 at short ICI (3 ms). At longer ICIs, weights were approximately equal for all clicks. Additionally, at ICIs of 3 and 5 ms, there is a consistent tendency for weights to increase from click 2 to click 16. At 5-ms ICI, this tendency manifested in a weight for click 16 that was significantly larger than the optimal value of 1/16 $=0.0625$ (and also larger than most of the preceding clicks' weights).

Figure 6 displays the precedence ratios computed from weighting functions for each subject. It can be seen from the figure that ratios were largest at 3-ms ICI, quantifying the onset dominance apparent in Fig. 5. Interestingly, however, the mean precedence ratio remained slightly above 1/15 (the expected equal-weighting value) for all ICIs, in all conditions. This indicates that although the weighting functions for 8- and 14-ms ICI in Fig. 5 did not deviate significantly from equal weighting, there was a tendency for onset dominance even for these stimuli, with click 1 receiving somewhat more weight than later clicks.

C. Discussion

The weighting functions plotted in Fig. 5 demonstrate a large onset emphasis for the shortest ICI (3 ms), with relatively even weights for the longer ICIs (8 and 14 ms). Even at short ICIs, however, weights for clicks beyond the first remained positive, indicating that all clicks had some influence on the responses. These results are consistent with previous findings in precedence and binaural adaptation (Hafter and Dye, 1983; Litovsky *et al.*, 1999; Shinn-Cunningham *et al.*, 1993). However, they disagree with the model of binaural adaptation proposed by Hafter and Buell (1990), which predicts a monotonic decrease in the relative effectiveness of each click. Predictions of that model, calculated by fitting Eq. (2) to $\dot{w}_1$ at each ICI, are plotted as dotted lines in Fig. 5. The obtained weighting functions reveal an immediate reduction in weights following click 1. Across conditions, the weight for click 2 tended to be among the smallest, and was significantly overestimated by Eq. (2) at 3-ms ICI. This result is consistent with the findings of Saberi (1996) and Dizon *et al.* (1998), who also showed a rather abrupt reduction of postonset weights at short ICIs.

Across conditions, the largest weights other than click 1 appeared near the end of the stimulus (for ICIs of 5–8 ms, they were slightly *larger* than those of click 1). This ''recovery'' of weights is one of the more intriguing aspects of the obtained weighting functions. No such recovery was seen by Saberi (1996), and of course no such effect would be observed in precedence-type studies where only two clicks are presented—although results showing an increased influence of click 2 (e.g., experiment 1; Stellmack *et al.*, 1999; Tollin and Henning, 1998) may be related. A number of experimental issues remain to be explored with respect to this finding; however, weighting functions for stimuli varying in duration suggest an increased influence of cues near the end of a stimulus, rather than a recovery from transient suppression following the onset (Stecker, 2000). This effect is reminiscent of ''recency'' effects observed in tests of verbal memory (Glanzer and Cunitz, 1966), and may be related to the integration of spatial information in sensory memory and/or response-planning mechanisms. These issues will be addressed in future work.

Increasing weights are, of course, not compatible with the relative-effectiveness model [Eq. (2)] presented by Hafter and Buell (1990), which assumes only monotonically nonincreasing weighting functions. However, the results are not necessarily at odds with the binaural-adaptation results (Hafter and Buell, 1990; Hafter *et al.*, 1988b; Hafter and Dye, 1983) summarized by Eq. (1), showing suboptimal improvement with duration at short ICIs.

A further consideration when comparing the results of this experiment with those found in the binaural adaptation literature is the difference between discrimination measures and the localization task employed here. There are some indications that the extent of laterality produced by modulated high-frequency stimuli is not necessarily predicted by the discriminability of their interaural cues (Bernstein and Trahiotis, 1994; Buell *et al.*, 1994). Some precedence studies have also shown that lagging stimuli can occasionally be discriminated—possibly based on nonspatial cues—in stimuli that produce single fused images (Litovsky *et al.*, 1999; Saberi and Perrott, 1990). Finally, Tollin and Henning (1998) found discrepancies between listeners' ability to independently lateralize both clicks in a pair and to discriminate the ITD of one click. Localization and discrimination measures must involve different brain mechanisms at some point (e.g., response selection), possibly with time courses different from those affecting purely sensory mechanisms (*viz.*, auditory mechanisms underlying onset dominance).

VI. EXPERIMENT 3: RELEASE FROM ADAPTATION WITH ACOUSTIC TRIGGER

One surprising result of the research on binaural adaptation is the so-called ''restart'' phenomenon (Hafter and Buell, 1990), whereby a brief acoustic change (or ''trigger'') in the middle of an extended stimulus produces a release from adaptation. For example, Hafter and Buell (1990) inserted a short gap between clicks 4 and 5 in an 8-click stimulus. Without the gap, the performance improvement from four to eight clicks was suboptimal at short ICI, but with it, performance was equivalent to two optimally combined 4-click stimuli, each adapted independently of the other. A similar improvement was obtained when the gap was replaced by a different acoustic ''trigger,'' such as a shortening of the ICI or the appearance of a brief tone burst (Hafter, 1997). They termed this release from binaural inhibition ''restarting,'' based on the idea that the effect of the trigger is to return sensitivity to normal, preadapted levels, hence restarting the adaptation mechanism.

Although the results of Hafter and Buell (1990) show clearly the effects of restarting with various triggers, Saberi's (1996) study of observer weighting in click-train lateralization revealed no effects of inserting gaps (4-ms gaps in trains of 1.8-ms ICI) at various positions within the trains. Saberi suggested that the difference between his results and those of Hafter and Buell (1990) may have been the result of the randomly varying ITD used in his study (compared to the static ITD used in previous studies). In this experiment, we interrupted 16-click trains with short gaps in order to explore the effects of restarting on sound localization.

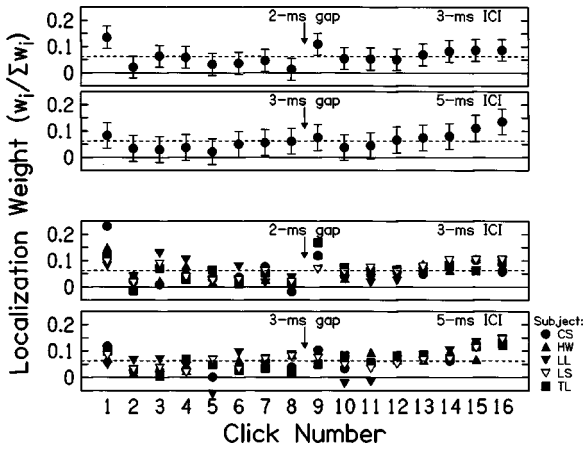


FIG. 7. Temporal weighting for interrupted click trains. Upper panels: mean normalized weighting functions obtained for 16-click trains with a short gap added to the ICI between clicks 8 and 9. For trains with 3-ms ICI (top panel), the gap was 2-ms long, while for trains with 5-ms ICI (second panel), the gap duration was 3 ms. Error bars indicate 95% confidence intervals on weight estimates, and dashed lines indicate the value expected for equal weight on all clicks. Lower panels: Normalized weighting functions for individual subjects. Error bars are omitted for clarity.

A. Methods

Five subjects (CS, HW, LL, LS, and TL) participated in this experiment. Subjects completed 4 runs of 100 trials at each ICI (3 or 5 ms).

Stimuli were trains of 16 clicks, generated as in experiment 2, with one major difference: the ICI between clicks 8 and 9 was lengthened by 2 ms (for trains with 3-ms ICI) or 3 ms (for 5-ms trains). As in experiment 2, stimuli were presented using loudspeaker condition W3. Other aspects of the experimental procedure and analytical technique were unchanged from experiment 2.

B. Results

Figure 7 plots the weighting functions for conditions incorporating a gap between clicks 8 and 9. The two upper panels show the mean normalized weighting functions, with 95% confidence intervals on the weight estimates. For a 3-ms ICI, a clear and significant elevation of the weight on click 9 ($\dot{w}_9$) was found, indicating that the gap was an effective restarting trigger. For 5-ms ICI, there was no significant increase of the weight on click 9. Lower panels plot the normalized weights for individual subjects; each subject is represented by a different symbol. Note the variation between the weighting functions for the different subjects, especially with regard to $\dot{w}_9$. At 3-ms ICI, subjects CS, HW, and TL showed rather large weights on click 9 (and also quite small weights on click 8). Subjects LL and LS, on the other hand, showed weights on click 9 that were approximately the equal-weight value of 1/16; these same subjects also had the lowest weights on click 1, and the weakest onset dominance (precedence) at short ICIs, as seen in Fig. 6. A similar, though less apparent, trend can be seen in the weights for trains with 5-ms ICI.

C. Discussion

At the shortest ICI (3 ms), introducing a gap of 2 ms between clicks 8 and 9 was sufficient to increase the weight on click 9, as expected from previous research on binaural adaptation (Hafter and Buell, 1990). However, the results also show clear intersubject variability with regard to restarting. Interestingly, the subjects that failed to show restarting showed the least onset dominance as well. In contrast to the results at 3-ms ICI, 3-ms gaps within trains of 5-ms ICI did not produce restarting. Some subjects in this condition (CS and TL) did produce larger weights on click 9 than on click 8, and those subjects also had the largest weights on click 1, however, the effects were much smaller than for 3-ms ICI, and not statistically significant.

Saberi (1996) found no effect of introducing gaps in trains of clicks presented over headphones. He reasoned that restarting may have been prevented by the random variation in ITD of his stimuli, compared to the static ITD employed by Hafter and Buell (1990). Based on the current results, we suggest that variation of interaural cues does not explain the difference, since spatial cues were also varied between clicks in this experiment, and restarting was apparent, at least for trains with 3-ms ICI. On the other hand, because the current paradigm employs free-field listening, stimuli carry both ITD and ILD cues, as well as cross-frequency cues related to the DTF. If variation of spatial cues causes the localization mechanism to alter the strategy used in combining time-and level cues, then the differences among the results of the three studies could be related to spatial variation, as Saberi suggested. However, more work comparing the influence of ITD and ILD in this paradigm is necessary before drawing strong conclusions along these lines.

The appearance of intersubject variation in the degree of restart (as seen in Fig. 7) suggests an alternate explanation of the difference between Saberi's results (Saberi, 1996) and those of the current study: namely, that the two studies sampled listeners with stronger (in our case) or weaker (in Saberi's) restarting. This explanation, however, appears unlikely for two reasons: First, the results for subjects LL and LS suggest that the degree of restarting may be related to the degree of onset dominance for a given listener; Saberi's results, in contrast, show large onset weights, but no effect of gaps. Second, Hafter and Buell (1990) observed little variability across subjects concerning the improvement in performance afforded by restarting.

VII. EXPERIMENT 4: EFFECTS OF VARYING ONLY ONE CLICK IN A STIMULUS

In experiments 1–3, every click contained in a stimulus train was subjected to random variation in its location. As pointed out by Saberi (1996), an important difference between observer-weighting approaches and earlier studies of binaural adaptation is that the latter have employed static interaural cues in assessing performance. Having already established that onset dominance, its rate dependency, and the restarting effect of Hafter and Buell (1990) are observed using the current paradigm, it is still quite possible that the obtained weighting functions differ in important ways from the time course of onset dominance for static stimuli. For example, if change in location were a sufficient trigger for restarting, then the random variations in click location could have acted to produce restarting at random times within each stimulus used in the study. Such occasional restarting would act to increase the weights of clicks following the restart (as was seen in Fig. 7), thus flattening *all* the functions obtained in this study. As a check for this kind of effect, experiment 4 examined localization weights for stimuli which did not possess random variation in the location of each click.

Ideally, we would derive a method for obtaining weighting functions directly from localization performance using static stimuli. The method of subtraction employed by Hafter and Buell (1990) utilizes static stimuli; however, the assumption underlying the estimation of temporal weighting functions—specifically that weights are monotonically nonincreasing—appear invalid for the regression paradigm, based on the results of experiments 2 and 3. The observer-weighting analysis employed in the current study avoids this assumption, but cannot be used for static stimuli, since it requires independent variation in each click's location. In experiment 4, we modified the observer-weighting technique by varying the location of only one click in the stimulus. All other clicks were emitted from a common location. In this paradigm, the relative influence of the varied click (or ''probe'') and the main ''body'' of the stimulus were assessed using multiple linear regression, as before. Because only two weights were computed (probe and body), and the body weight encompassed 15 clicks (for 16-click trains), the interpretation of the weighting functions is not as straightforward as in the other experiments. However, plotting probe weight as a function of probe position produces a function that is comparable to the weighting functions derived in experiments 1–3.

A. Methods

Two subjects (CS and HW) participated in this experiment. Each completed 8 runs of 100 trials at each tested ICI: 3, 5, and 14 ms in the probe condition and 4 runs of 100 trials at each of the same ICIs in a control condition (see below).

Stimuli were trains of 16 clicks, generated as in experiment 2. One of the 16 clicks was chosen as the ''probe'' on each trial. There were four potential probe positions corresponding to clicks 1, 2, 9, and 16. A new position was chosen at random from this set on each trial. The remaining clicks (e.g., clicks 1–8 and 10–16 when click 9 was the probe), termed the ''body,'' shared a common location (the stimulus location—θ_L—chosen on that trial). The location of the probe was selected in the same manner that individual click locations were selected in experiments 1–3. Loudspeaker condition W5 was used in this experiment; here, the probe was presented from -11, -5.5, 0, $+5.5$, or $+11$ degrees relative to the body on any given trial. We chose to use this increased spatial resolution to provide a more stable estimate of the single probe weight. A control condition employed condition W5 and the same subjects as the probe condition but was otherwise identical to experiment 2. Other aspects of the experimental procedure were identical to those used in

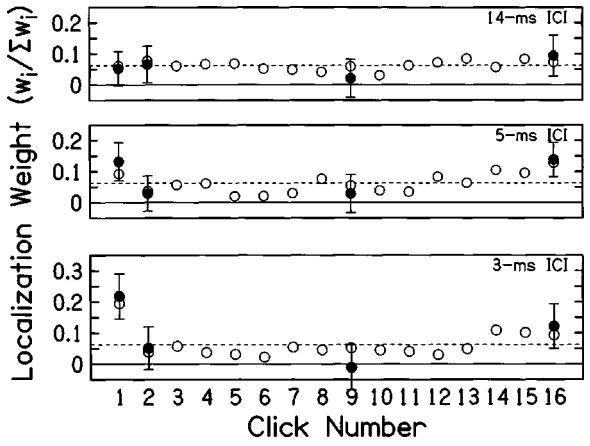


FIG. 8. Mean normalized weights obtained using the probe method (experiment 4). Filled symbols indicate the normalized probe weight (see the text), plotted as a function of the probe position within a static train of 16 clicks. Error bars indicate 95% confidence intervals on probe weight estimates. Confidence intervals appear larger here than in experiments 1–3, partly because weights here were calculated using ≈200 trials each, as opposed to 400 trials in experiments 2 and 3. Open symbols plot the weighting functions obtained for 16-click trains in a control condition similar to experiment 2 but using the probe method's loudspeaker configuration (w5). Confidence intervals for the control condition (not shown) were similar to experiment 2. Finally, dashed lines indicate values expected for equal weighting of all clicks.

experiment 2. However, the analytical technique was modified for the probe method employed by the new design.

Because all clicks comprising the body of the stimulus shared a common location, they did not vary independently and hence were not assigned weights individually. Rather, the model used here included two predictors: body and probe

$$\hat{\theta}_R = w_{\text{probe}}\theta_{\text{probe}} + w_{\text{body}}\theta_{\text{body}} + k. \tag{6}$$

Since the body contained 15 clicks, while the probe contained only 1, the body was expected to exert a stronger influence on localization (i.e., w_{body} should be somewhat larger than w_{probe}), regardless of the probe position. As before, normalized weights were computed

$$\dot{w}_{\text{probe}} = \frac{w_{\text{probe}}}{w_{\text{probe}} + w_{\text{body}}}. \tag{7}$$

The normalized probe weight expresses the relative influence of the probe on localization responses, just as before, and would be expected to vary as a function of probe position, with large values for a probe at click 1, small values for a probe at click 2, and so on.

B. Results

Figure 8 shows the mean normalized probe weights as a function of probe position. Open symbols indicate weighting functions obtained in the control condition (similar to the method used in experiment 2) for comparison. Weights obtained in the control condition did not differ significantly from those obtained in experiment 2. There were additionally no significant differences between the results for the probe and control conditions. For both 3- and 5-ms ICI there was a trend for slightly more extreme weights (higher at clicks 1 and 16, lower at click 9) in the probe condition consistent with the notion that spatial variation may have acted to flatten weights in the control condition (and in experiments 1–3); however, none of these effects was statistically significant.

C. Discussion

The lack of significant systematic differences between the weights calculated using the probe method and those calculated with random variation on all clicks lends support to the form of observer weights reported in this study. The results of this experiment show a nonsignificant tendency for less-flattened weights using the probe method compared to the control. This would seem to suggest a small homogenizing effect of click variation on obtained weights in experiments 1–3. However, conclusions should be tempered by consideration of the subjects participating in experiment 4, who were more experienced than other subjects, and tended to show larger effects of onset dominance at short ICI in the earlier experiments. Although the results of experiment 4 are expressed in relation to a control condition employing only those subjects, it is possible that other subjects would have been less sensitive to the manipulation. Nevertheless, the correspondence between weighting functions obtained in the two methods suggests that the results of the previous experiments were not largely affected by continuous variation in the locations of clicks; if anything, they may have acted to reduce differences between the weights for different clicks.

VIII. SUMMARY AND CONCLUSIONS

(1) *Onsets dominate localization at high stimulus rates*. Consistent with the findings of previous research on precedence and binaural adaptation, click 1 received larger weights than subsequent clicks when the ICI was short (1–5 ms). This was the case for click pairs (experiment 1) as well as 16-click trains (experiment 2). At slower rates (ICI=8 or 14 ms), weights for click 1 were not significantly different from subsequent clicks.

(2) *Reduction of localization weights following onset is immediate*. For 16-click trains, weighting functions at short ICI (3–5 ms) were characterized by an immediate reduction in weights for clicks following the onset (e.g., click 2), with steady or increasing weights for the remainder of the stimulus. This finding contrasts with that obtained by Hafter and Buell (1990). Using a subtractive method to estimate weighting functions, they suggested that click effectiveness decreases monotonically following the onset. The current study agrees with that of Saberi (1996), who also found immediately reduced weights for clicks 2 and later using an observer-weighting method, and Dizon *et al.* (1998), who found similar functions for ITD-varying noises.

(3) *Onset dominance is not affected by spatial spread of the stimulus over the range* 5.5–22°. There were no significant differences between weighting functions for 2-click trains presented from speakers spanning 11° or 22° azimuth; neither were there significant differences between those obtained from trials where clicks 1 and 2 in a pair were separated by 5.5°, 11°, or 22° (experiment 1). Other

studies have shown an effect of spatial separation on precedence at much wider separations (Shinn-Cunningham *et al.*, 1993). The current results suggest that this influence is not a straightforward function of separation, or that the ''spatial window'' of precedence is larger than 22° in azimuth.

(4) *Acoustic triggers can act to increase the weights on post-trigger clicks*. Consistent with previous work on the restart phenomenon in binaural adaptation (Hafter and Buell, 1990), introducing a gap between clicks 8 and 9 acted to increase the weight on click 9, at least for short ICI.

(5) *Weighting functions are not specific to randomly varying stimuli*. The influence of a single click is essentially the same regardless of whether other clicks in the train are presented from randomly varying locations (experiment 2) or from a single static location (experiment 4).

ACKNOWLEDGMENTS

Support for this work was provided by NIH/NIDCD Research Grant 00087. We thank Miriam Valenzuela, Frédéric Theunissen, and Bruce Berg for their assistance in developing the observer-weighting method employed in this study; Ephram Cohen and Erick Gallun for invaluable assistance with programming and figure preparation; Rich Ivry, Marty Banks, David Wessel, Gerald Kidd, and Chris Mason for helpful comments during the early stages of this research; and finally John Middlebrooks, Brian Mickey, Ewan Macpherson, Julie Arenberg Bierer, along with two anonymous reviewers for providing feedback on earlier versions of this paper.

APPENDIX: CONFIDENCE INTERVALS ON WEIGHT ESTIMATES

The standard error of an individual regression coefficient w_i is given as

$$s_{w_i} = \frac{s_{\theta_R \cdot \theta_i}}{s_{\theta_i}\sqrt{N-1}}, \tag{A1}$$

where $s_{\theta_R \cdot \theta_i}$ is the standard error of estimate (square root of error variance) associated with the predictor variable in question and s_{θ_i} is the standard deviation of that predictor. Here, N is the number of samples used to estimate w_i (Howell, 1997). The statistical significance of differences between coefficients can be examined using *t*-tests. Alternately, one can compute confidence intervals (CI) on the weight estimates

$$\mathrm{CI}_{w_i} = w_i \pm t_{(\alpha/2,df)} s_{w_i}, \tag{A2}$$

where $t_{(\alpha/2,df)}$ is the critical value of Student's *t* for a significance level of $\alpha/2$ (e.g., 0.025 for 95% confidence intervals) and df degrees of freedom $(N-n-1)$. Ninety-five-percent confidence intervals were calculated based on the raw weights (prior to normalization) for each subject. For plotting with normalized weights, confidence limits were normalized in the same manner as the weights themselves

$$\mathrm{CI}_{\dot{w}_i} = \dot{w}_i \pm t_{(\alpha/2,df)} s_{\dot{w}_i} = \dot{w}_i \pm \frac{t_{(\alpha/2,df)} s_{w_i}}{\sum_{j=1}^{n} w_j}. \tag{A3}$$

Plots of mean weights across subjects use error bars displaying mean confidence limits, rather than limits corresponding to the pooling of data across subjects. As such, they do not provide an estimate of intersubject variability; conversely, they overestimate the actual 95% confidence interval for the mean weight, and thus reflect a somewhat conservative estimate of significance.

[1]Some readers may object to our lumping together of potentially disparate phenomena, and we tend to agree with the general view that a number of interacting neural mechanisms at different levels may be involved in onset-dominance phenomena. Nevertheless, it is our view that at least some distinctions in the literature result primarily from differences in the experimental tasks and stimuli used in various studies. The stimuli employed here were designed to be similar to those used in previous studies of *both* binaural adaptation and the precedence effect. Labeling the target of this investigation as one or the other at this point seems mistaken, especially considering the remarkably convergent temporal extents of the two phenomena.

Ahumada, A., and Lovell, J. (**1971**). ''Stimulus features in signal detection,'' J. Acoust. Soc. Am. **49**, 1751–1756.

Berg, B. G. (**1980**). ''Analysis of weights in multiple observation tasks,'' J. Acoust. Soc. Am. **86**, 1743–1746.

Bernstein, L. R., and Trahiotis, C. (**1994**). ''Detection of interaural delay in high-frequency sinusoidally amplitude-modulated tones, two-tone complexes, and bands of noise,'' J. Acoust. Soc. Am. **95**, 3561–3567.

Blauert, J. (**1983**). *Spatial Hearing* (MIT Press, Cambridge, MA).

Blauert, J., Canévet, G., and Voinier, T. (**1989**). ''The precedence effect: No evidence for an ''active'' release process found,'' J. Acoust. Soc. Am. **85**, 2581–2586.

Buell, T. N., Trahiotis, C., and Bernstein, L. R. (**1994**). ''Lateralization of bands of noise as a function of combinations of interaural intensive differences, interaural temporal differences, and bandwidth,'' J. Acoust. Soc. Am. **95**, 1482–1489.

Clifton, R. K. (**1987**). ''Breakdown of echo suppression in the precedence effect,'' J. Acoust. Soc. Am. **82**, 1834–1835.

Dizon, R. M., Culling, J. F., Litovsky, R. Y., Shinn-Cunningham, B. G., and Colburn, H. S. (**1998**). ''On the development of a post-onset temporal weighting function,'' in *Abstracts of the Twenty-first Midwinter Research Meeting* (Association for Research in Otolaryngology, St. Petersburg Beach, FL), p. 42.

Freyman, R. L., Zurek, P. M., Balakrishnan, Y., and Chiang, Y. C. (**1997**). ''Onset dominance in lateralization,'' J. Acoust. Soc. Am. **101**, 1649–1659.

Gaskell, H. (**1983**). ''The precedence effect,'' Hear. Res. **11**, 277–303.

Glanzer, M., and Cunitz, A. R. (**1966**). ''Two storage mechanisms in free recall,'' J. Verbal Learn. Verbal Behav. **5**, 351–360.

Haas, H. (**1972**). ''The influence of a single echo on the audibility of speech,'' J. Audio Eng. Soc. **20**, 146–159.

Hafter, E. R. (**1997**). ''Binaural adaptation and the effectiveness of a stimulus beyond its onset,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), pp. 211–232.

Hafter, E. R., and Buell, T. N. (**1990**). ''Restarting the adapted binaural system,'' J. Acoust. Soc. Am. **88**, 806–812.

Hafter, E. R., Buell, T. N., Basiji, D. A., and Shriberg, E. E. (**1988a**). ''Discrimination of direction for complex sounds presented in the free-field,'' in *Basic Issues in Hearing*, edited by H. Duifhuis, J. W. Horst, and H. P. Witt (Academic, London), pp. 394–401.

Hafter, E. R., Buell, T. N., and Richards, V. M. (**1988b**). ''Onset coding in lateralization: Its form site, and function,'' in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 647–676.

Hafter, E. R., and Dye, R. H. J. (**1983**). ''Detection of interaural differences of time in trains of high-frequency clicks as a function of interclick interval and number,'' J. Acoust. Soc. Am. **72**, 644–651.

Hafter, E. R., Dye, R. H. J., and Wenzel, E. M. (**1983**). ''Detection of interaural differences of intensity in trains of high-frequency clicks as a function of interclick interval and number,'' J. Acoust. Soc. Am. **73**, 1708–1713.

Hafter, E. R., Dye, R. H. J., Wenzel, E. M., and Knecht, K. (**1990**). ''The combination of interaural time and intensity in the lateralization of high-frequency complex signals,'' J. Acoust. Soc. Am. **87**, 1702–1708.

Howell, D. C. (**1997**). *Statistical Methods for Psychology*, 4th ed. (Duxbury, Belmont, CA).

Kuhn, G. F. (**1987**). ''Physical acoustics and measurements pertaining to directional hearing,'' in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), Chap. 1, pp. 3–25.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). ''The precedence effect,'' J. Acoust. Soc. Am. **106**, 1633–1654.

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

Rakerd, B., and Hartmann, W. M. (**1985**). ''Localization of sound in rooms. II. The effects of a single reflecting surface,'' J. Acoust. Soc. Am. **78**, 524–533.

Saberi, K. (**1996**). ''Observer weighting of interaural delays in filtered impulses,'' Percept. Psychophys. **58**, 1037–1046.

Saberi, K., and Perrott, D. R. (**1990**). ''Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate,'' J. Acoust. Soc. Am. **87**, 1732–1737.

Shinn-Cunningham, B. G., Zurek, P. M., and Durlach, N. I. (**1993**). ''Adjustment and discrimination measures of the precedence effect,'' J. Acoust. Soc. Am. **93**, 2923–2932.

Shinn-Cunningham, B. G., Zurek, P. M., Durlach, N. I., and Clifton, R. K. (**1995**). ''Cross-frequency interactions in the precedence effect,'' J. Acoust. Soc. Am. **98**, 164–171.

Stecker, G. C. (**2000**). ''Observer weighting in sound localization,'' Ph.D. thesis, University of California, Berkeley.

Stellmack, M. A., Dye, R. H., and Guzman, S. J. (**1999**). ''Observer weighting of interaural delays in source and echo clicks,'' J. Acoust. Soc. Am. **105**, 377–387.

Tollin, D. J., and Henning, G. B. (**1998**). ''Some aspects of the lateralization of echoed sound in man. I. The classical interaural-delay based precedence effect,'' J. Acoust. Soc. Am. **104**, 3030–3038.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). ''The precedence effect in sound localization,'' Am. J. Psychol. **52**, 315–336.

Yang, X., and Grantham, D. W. (**1997**). ''Echo suppression and discrimination suppression aspects of the precedence effect,'' Percept. Psychophys. **59**, 1108–1117.

Zurek, P. M. (**1980**). ''The precedence effect and its possible role in the avoidance of interaural ambiguities,'' J. Acoust. Soc. Am. **67**, 952–964.

Early pitch-shift response is active in both steady and dynamic voice pitch control

Theresa A. Burnett[a),b)] and Charles R. Larson
Department of Communication Disorders, Northwestern University, Evanston, Illinois 60208

(Received 13 December 2001; revised 15 April 2002; accepted 22 April 2002)

When air conducted auditory feedback pitch is experimentally shifted upward or downward during steady phonation, voice pitch changes in response. The first pitch change is an automatic deflection opposite in direction to the feedback shift. It appears to help stabilize voice pitch by counteracting unintended changes. But what happens during an intended pitch change? If the purpose of the first pitch-shift response is to stabilize voice pitch around a fixed target, it should be suppressed during voluntary pitch changes. Alternatively, if the pitch-shift response is a general process of voice control it should be modified during intended pitch changes to bring production in line with the desired output. Auditory feedback pitch was shifted during steady pitch and upward glissando vocalizations by thirty trained singers. Contrary to the ''steady-specific'' hypothesis, pitch-shift responses occurred during dynamic pitch vocalizations. Responses were comparable in direction, peak time, and slope, but had significantly longer latency and smaller magnitude than responses elicited during steady note phonation. Results indicate that the early pitch-shift response is a general component of voice control that serves to automatically bring phonation pitch into agreement with an intended target, whether that target is constant or changing in time. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1487844]

PACS numbers: 43.70.Aj [AL]

I. INTRODUCTION

Air conducted auditory feedback is important for on-line fine pitch[1] control. Interference with auditory feedback through masking lengthens the period of time used for pitch adjustment, and ultimately reduces pitch accuracy (Elliot and Niemoeller, 1970). Moreover, the detrimental effect of auditory feedback masking increases with the distance between the target pitch and a singer's modal pitch (Mallard, Ringel, and Horii, 1978). Auditory feedback is also important for the maintanance of an internal representation of pitch. Sustained perturbation of air conducted auditory feedback pitch induces adaptive changes in phonation. For example, Elman (1981) found that voice pitch is adjusted in the presence of altered pitch feedback to allow accurate production of a target note. Similarly, when a disparity between feedback pitch and voice F_0 is created and then gradually increased over a series of vocalizations, voice F_0 changes in a manner that opposes the feedback alteration (Jones and Munhall, 2000).

Even the presentation or manipulation of a sound during phonation can trigger changes in vocal fold configuration and voice fundamental frequency. Sapir (1983) found that a loud click presented during sustained phonation induces a brief, small involuntary rise in voice F_0 (mean latency 59 ms; SD 12 ms). The short latency of this response is suggestive of midbrain-level networks linking the auditory and phonatory systems. Continuous influence of auditory feedback pitch during phonation was later described by Sapir, McClean, and Luschei (1983), who showed that voice F_0 will oscillate at the same frequency, albeit out of phase, with oscillations of a sawtooth tone loudly presented during quiet phonation. These vocal variations may have been early evidence of the pitch-shift response (PSR) later described by Larson *et al.* (1995).

Using signal processing techniques that manipulate frequency while leaving duration intact (see Bristow-Johnson, 1995 for further information), it has been shown that a single transient perturbation of auditory feedback pitch during sustained phonation results in one or more voice pitch changes (Larson *et al.*, 1995; Burnett *et al.*, 1998; Hain *et al.*, 2000). Brief pitch shifts ($\leqslant$100 ms in duration) typically evoke a single involuntary compensatory voice F_0 change with a latency of approximately 100 ms (Burnett *et al.*, 1998). The magnitude of this early pitch-shift response (VR1) usually stays below 50 cents, and does not directly correspond to the magnitude of the pitch shift (Burnett, Senner, and Larson, 1997; Burnett *et al.*, 1998; Larson *et al.*, 1995). Pitch shifts sustained longer than 100 ms are often associated with a bimodal or double pitch-shift response. Unlike VR1, the direction and magnitude of the second response (VR2) vary with task instructions, and its duration is positively correlated with the duration of the pitch-shift stimulus. Although perhaps not intentionally produced, this second PSR component certainly appears to be generated with some voluntary influence.

VR1 is influenced by voluntary pitch control. When instructed to treat a pitch-shift stimulus as a ''go'' cue to change voice pitch in a particular direction, VR1 is significantly reduced in duration and magnitude, occasionally disappearing entirely (Hain *et al.*, 2000). Differences between the early and late components of the pitch-shift response suggest that they represent two levels of auditory feedback me-

a)Currently affiliated with the National Institute of Neurological Disorders and Stroke, Bethesda, MD 20892-1416.
b)Author to whom correspondence should be addressed. Electronic mail: burnettt@ninds.nih.gov

diated online error correction, with each perhaps generated by different neural circuitry and contributing uniquely to voice pitch control. It may be that VR1 is particularly tuned to compensate for small perturbations while VR2 facilitates the correction of large production errors. Together, VR1 and VR2 reveal that the PSR involves a complex, task-dependent interaction between audition and phonation. What is as yet unclear is how the PSR behaves during dynamic voice pitch tasks.

When singing a steady note, VR1 magnitude is measured at peak time from a baseline pitch that is coincidentally the same as the one perturbed by the pitch shift stimulus. Given the response latency of approximately 100 ms, it has not been discernable whether VR1 is generated relative to the pitch that was perturbed or to the pitch that is produced at the time of the peak response. If the function of VR1 is to automatically hold phonation pitch steady by correcting any deviation in voice pitch, then it might be reduced or eliminated during dynamic vocalizations when (a) an increase or decrease in voice pitch is anticipated and desired, and (b) a response generated relative to the perturbed pitch may occur too late, when that pitch is no longer being produced. Suppression of a pitch-stabilizing response during intentional voice pitch changes could be considered analogous to suppression of the gaze stabilizing vestibulo–ocular reflex during smooth pursuit eye movements (Barnes and Greely, 1992; Waterston and Barnes, 1991).

Alternatively, VR1 may be triggered by a voice feedback error, despite intentional pitch changes. Instead of being suppressed, VR1 would theoretically be generated in accordance with an internal representation of the ongoing voice pitch pattern to correct what might be perceived as a miscalibration of the pitch control system, rather than a single error in the production of a particular note. In order to better understand the purpose of the pitch-shift response, a paradigm was designed to assess whether VR1 is a task-specific mechanism that aids steady pitch vocalization, or if it is a general mechanism of voice pitch control that is active during both steady and dynamic pitch vocalizations.

II. METHODS

Participants: Trained singers were recruited to participate in this study in order to minimize variability in task performance. There were 15 females (10 sopranos, 4 mezzo-sopranos, 1 alto) and 15 males (6 tenors, 8 baritones, 1 bass). Females averaged 21.7 years of age (SD=2.6, range=18–26) with a mean of 7.1 years of formal voice training (SD =1.6, range=5–9). Males averaged 25.3 years of age (SD =8.1, range=18–42), with a mean of 6.9 years of formal voice training (SD=1.8, range=4–10). All gave informed consent, passed a standard hearing screening, and reported no history of neurological injury or disease, or vocal pathology.

Instrumentation: Participants sat in a chair placed inside an acoustically dampened booth. They wore 75-ohm AKG K270 H/C Talkback over-the-ear headphones and vocalized into a boom-held microphone extending from the left side of the headphones. Between transduction by the microphone and feedback through the headphones, the voice signal was processed in near-real time (8–20 ms delay) through an Eventide SE Ultra-harmonizer (used to digitally shift the frequencies of acoustic signals), a Mackie 1202 VLZ micro mixer, and a Hewlett-Packard dB attenuator. The intensity of the voice signal presented over the headphones was equivalent to the loudness of the signal at the microphone. Participants were instructed to maintain voice loudness near 80 dB SPL using the display of a Dorrough loudness meter as a guide. No attempt was made to mask bone conducted auditory feedback.

A Yamaha MU10 synthesizer was used to generate a pitch pattern for vocal imitation, and its output was presented through the headphones via the mixer. Both the harmonizer and the synthesizer were controlled by musical instrument digital interface (MIDI) signals from a laboratory computer (MAX, Opcode Systems, Inc., Nashville, TN).

Procedure: Data were recorded in alternately ordered blocks of 30 vocalizations (trials) each, one called ''hold'' and the other, ''glissando.'' To ensure consistency across vocalizations, each trial was preceded by the presentation of an acoustic model generated by the synthesizer. The amplitude of the model signal was 80 dB SPL and it resembled a chorus singing the vowel /u/. The lowest pitch modeled was equal to 20% of the participant's total vocal range, and the highest required note was nine semitones above that 20% note. These limits were set to normalize the experiment across participants of different ''voices,'' and to minimize the likelihood of pitch or phonation breaks such as might occur when crossing a vocal register. In the hold condition, a model note was selected at random within the designated range, and played for a duration of 6 s. In the glissando condition, the model was a steady note (1 second) that transitioned into an upward glide (1 semitone/second for 4 seconds), and then held steady again (1 second).

Participants were instructed to mimic the model presented over the headphones, beginning when ready but only after the model had ended. They were also instructed to sing in a straight manner, without vibrato. Participants were warned they might notice a brief pitch shift at some time during their vocalizations but told to ignore it as best they could. No practice trials were provided. Water and rest breaks were granted upon request.

A single pitch-shift stimulus (PSS) was introduced during each trial. All PSS were identical, with a magnitude of −100 cents and a duration of 100 ms. Qualitatively, the PSS sounded as though the voice pitch momentarily went ''flat'' by 1 semitone. PSS were triggered at a random time between 2–3 s post onset of phonation (as marked by experimenter button press) for the holds, and 2.5–3.5 s post phonation onset for the glissandos. Differences between trials and participants in which pitch was shifted were not considered problematic because the PSR does not significantly vary with voice pitch (Burnett *et al.*, 1997).

Analysis: Four voltage signals, derived from the voice, processed voice feedback, synthesized model, and a function generator (shifted simultaneously with the voice for later use as a mark of PSS onset), were antialias filtered at 5000 and digitally recorded at a sampling frequency of 10 000 Hz using MacLab (AD Instruments). The voice signals were then

further low pass filtered; 200 Hz for males and 300 Hz for females. Voice fundamental frequency was extracted using a threshold-crossing technique with interpolation that yields a time value for each glottal pulse. Since glottal pulses differ in duration (jitter), plots of these values yield a wave in which there are sharp discontinuities between each glottal cycle. A fifth order, digital low pass filter set at 20 Hz was used to smooth these discontinuities. Finally the voice, feedback and synthesized signals were converted to cents using the equation: $\text{cents} = 100(39.86 \log_{10}(f2/f1))$ with $f1$ equal to the arbitrarily chosen note $G4$ (195.997 Hz) and $f2$ equaling the voice signal in Hz.

Individual trials by each subject within each condition were aligned in time to the onset of the PSS, and a 1500 ms segment (500 ms pre-PSS through 1000 ms post-PSS) was taken for analysis. The pre-PSS voice signal was examined, and trials featuring phonation breaks, vibrato or failure to adhere to task instructions were excluded (typically 3–8 of the 30 trials in each condition). Figure 1 shows waterfall displays of one participant's hold and glissando trials. An ensemble average of the trials was performed, resulting in one average case per experimental condition across each of the 30 participants (60 total).

Figure 2 displays the method used to measure response latency, magnitude, and peak time. A line created by a linear curve fit of the prestimulus voice signal was extended across the analysis window (hereafter called the ''baseline''). Above and below this line were parallel lines indicating the upper and lower second SD limits. Software was used to automatically mark responses, accepting only those that occurred more than 50 ms after the onset of the PSS with a duration 30 ms or greater. Once a response was delineated, the program automatically measured the crossing time, peak deviation from the baseline, and peak time. Because the second SD crossing time is a function of prestimulus voice pitch variation, response latency was defined as that time at which a straight line, extrapolated backwards from the response peak through the voice signal at the preceding second SD crossing time, intersected the baseline. Response magnitude was measured as the difference in cents between the response peak and the baseline pitch at peak time.

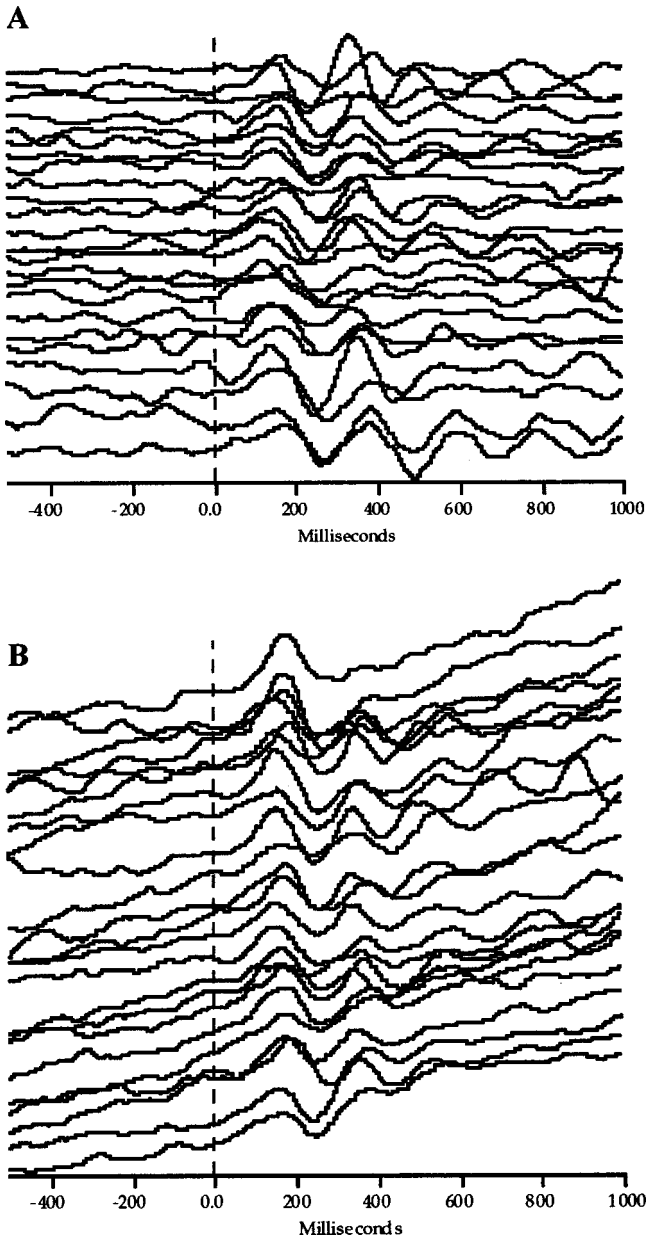


FIG. 1. Waterfall displays of all trials for one participant in (a) the hold condition, and (b) the glissando condition, prior to ensemble averaging. Pitch is represented on the vertical axis, with each trace identically scaled but offset slightly for display purposes. The dotted vertical line at time 0 represents the onset of the pitch shift stimulus.

III. RESULTS

Participants consistently suppressed vibrato and held or changed voice pitch as instructed, enabling a valid comparison of responses across the two conditions. Mean baseline slope in the hold condition was 0.64 cents/s (SD=0.99). Mean baseline slope in the glissando condition was 124.19 cents per second (SD=28.77). The second SD lines used to mark the occurrence of a response to the PSS give an indication of voice pitch stability in the 500 ms pre-PSS vocalization period. Within-participant (trial-to-trial) variability prior to PSS onset was not significantly different during glissando vocalizations from the variability during holds ($F = 1.30$; 1 df; $p = 0.26$). The second SD lines were a mean of 2.63 cents from baseline in the glissando condition (SD =1.28) and 2.34 cents from it in the hold condition (SD =1.12).

Of the 60 cases, 58 (97%) featured responses directed upward in pitch, opposite the downward direction of the PSS. The remaining two cases featured downward responses, in the same direction as the PSS. Both of these ''following'' responses occurred during glissando vocalizations. Measurement of magnitude and peak time in one averaged case was impossible because of truncation by the analysis window, and both had latency values that were markedly different from the rest of the data. Of note, it was shown in an earlier study that ''following'' responses are significantly different in latency from the more typical ''opposing'' responses (Burnett *et al.*, 1998). Therefore, these two ''following'' responses were omitted from statistical analysis.

Examination of the 58 ''opposing'' first responses (28

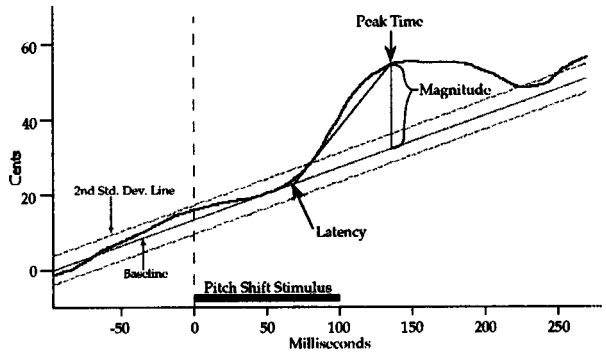


FIG. 2. Graphical representation of the techniques for measurement of response latency, magnitude and peak time. Response latency equals the time at which a straight line, extrapolated backwards from the response peak through the voice signal at the preceding second SD crossing time, intersected the baseline. Response magnitude equals the difference in cents between the response peak and the baseline pitch at peak time.

glissando; 30 hold) revealed three cases with data points that were more than 2.8 SD from the mean. All outliers were from the glissando condition, including two in the latency measure, and one in the measure of peak time. Exclusion of all cases with outlier data points left 25 participants with a complete set of data across the two vocalization tasks.

A Hotelling's T^2 test (a special case of MANOVA) was performed to determine if task had any effect on response latency, magnitude or peak time. A significant effect was found (T^2 value=18.18, $p<0.01$), which was attributed to differences across tasks for the measures of latency (95% confidence interval: 0.005 to 0.029) and magnitude (95% confidence interval: −11.22 to −0.66). In general, responses that occurred earlier tended to have a greater magnitude than those that occurred later (Pearson $r=-0.30$; Bartlett chi-square=4.57, df=1, $p<0.05$). Early VR1 responses also tended to have earlier peak times than later VR1 responses (Pearson $r=0.37$; Bartlett chi-square=7.036, df=1, $p<0.01$).

Mean response latency in the hold condition was significantly shorter (75.61 ms, SD=15.81) than in the glissando condition (92.73 ms, SD=28.40), a difference of 17.12 ms. Figure 3 shows dot plots of response latency by task for each participant.

Mean VR1 magnitude was 23.58 cents (SD=10.81) in the hold condition, significantly greater than the mean of 17.64 cents (SD=10.17) when produced on a glissando, and

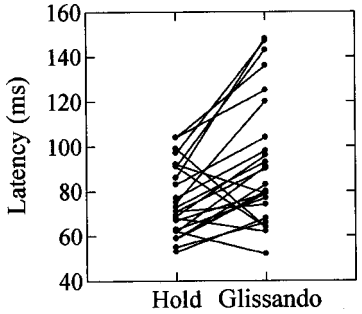


FIG. 3. Latency of the pitch-shift response elicited while holding voice pitch steady and while producing an upward glissando. Lines connect the data points for each participant.

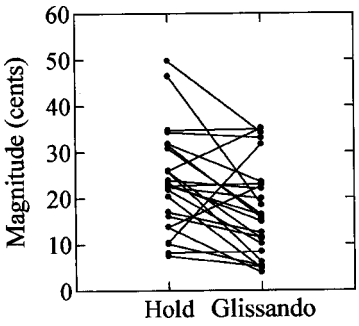


FIG. 4. Magnitude of the pitch-shift response elicited while holding voice pitch steady and while producing an upward glissando. Lines connect the data points for each participant.

a difference of approximately 5.94 cents. Figure 4 shows dot plots of response magnitude by task for each participant.

Vocalization task did not significantly affect VR1 peak time (95% confidence interval: −0.016 to 0.014). Mean peak time was 162.71 ms in the hold condition (SD=25.78) and 162.16 in the glissando condition (SD=25.27). Figure 5 shows dot plots of response peak time by task.

Given that magnitude divided by time-to-peak equals response slope, a comparable peak time but shorter latency and greater magnitude in the hold condition hint that response slope during hold might be similar to that in during the glissando condition. Indeed, they were similar. Response slope in the hold condition averaged 287.70 cents/s (SD =131.62), while average response slope during glissando production was 272.14 cents/s (SD=153.03).

IV. DISCUSSION

The purpose of this study was to examine whether the early component of the pitch-shift response is exclusive to steady pitch vocalizations, or whether it is a mechanism that could aid control of both steady and dynamic pitch vocalizations. The primary finding is that the early, automatic pitch shift response is not exclusive to steady pitch vocalizations. A downward perturbation of auditory feedback pitch induces an opposing, upward voice pitch response whether one is sustaining a steady note or continuously changing pitch in an upward glissando. The PSR system allows for intended voice pitch variation and is triggered by unintended changes in voice feedback pitch, i.e., errors. Response peak time and slope were comparable across both tasks, but latency was

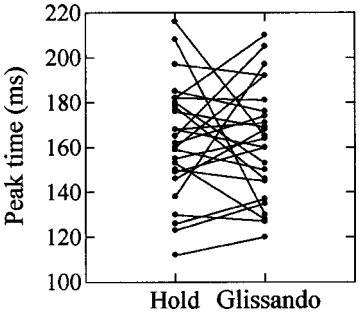


FIG. 5. Peak time of the pitch-shift response elicited while holding voice pitch steady and while producing an upward glissando. Lines connect the data points for each participant.

shorter and magnitude greater when attempting to hold voice pitch constant than when increasing pitch at a rate of approximately one semitone/second.

Pitch-shift response latencies measured in this experiment were shorter and likely more accurate than those previously reported (Burnett *et al.*, 1998; Hain *et al.*, 2000; Larson *et al.*, 1996). Differences are primarily attributed to the use of extrapolation in the measurement of response latency and a baseline pitch that is based on a linear curve-fit of the pre-PSS period rather than the mean F_0 of that period. Also, inter-trial variability was reduced and responses made more apparent by having only trained singers as participants, having participants sing /u/ rather than /a/ to ease F_0 extraction, and by including more trials in the ensemble average than were previously included. Although trained singers voluntarily change pitch more quickly than do untrained singers (Sundberg, 1979), the PSR appears to be involuntarily produced and therefore training is not believed to be a factor in response generation. However, musicians have been found to have superior preattentive auditory processing abilities (Koelsch, Schröger, and Tervaniemi, 1999) which might be linked to a more rapid automatic F_0 correction.

Shorter response latency while holding voice pitch steady may be because pitch deviations are more quickly detected from a stable background than from a changing one. However, the human auditory system is known to be remarkably adept at detecting subtle differences in complex sounds (Klatt, 1973; Tervaniemi, Schröger, and Näätänen, 1997). More likely the difference stems from response generation variables. Response generation during glissandos may require more calculation time than during holds to factor the ongoing, planned vector of voice pitch change into the calculation of response magnitude. This process allows the response to improve the accuracy of the pitch produced at the time of the peak response, rather than the pitch that was heard to be erroneous but is no longer being produced.

It may be that VR1 during all dynamic pitch tasks is smaller than when produced during steady phonation. No upward PSS was delivered and no downward glissando was produced during this experiment, so it is not possible to know if and how the PSR might differ in these conditions. Alternatively, the lower response magnitude during glissandos could reflect a gain that varies with the ongoing, intended voice pitch change. It may be that responses in the same direction as the underlying task (e.g., an upward VR1 during upward glissandos) are smaller than responses that are opposite in direction to the underlying voice pitch change (e.g., upward VR1 during a downward glissando). Steady pitch phonation would fall between these two extremes. Further investigation is needed to specify the effect of baseline slope on response magnitude.

The trigger for generation of the PSR evidently involves a comparison between air conducted auditory feedback and some other signal, the nature of which is unknown. The results of this study indicate that the comparative signal is not simply that pitch most recently heard and held in immediate auditory memory. Remaining possibilities include (a) the unshifted bone-conduction signal and (b) an internal representation of the anticipated pitch calculated by plotting an ''efference copy'' signal on an auditory–phonatory sensorimotor map. The methods of this and many earlier studies feature shifts of only the air conducted auditory feedback signal, with no attempt to mask the unshifted bone conducted signal. Although total elimination of bone conduction is not possible, Jones *et al.* (2000) demonstrated that a compensatory voice pitch change occurs in response to gradually shifted air conducted feedback in the presence of bone masking. Moreover, experience speaking under the condition of vowel formant transformed auditory feedback results in compensatory adaptations of vowel articulation patterns such that auditory feedback is altered toward the usual vowel formant pattern (Houde and Jordan, 1998). Immediate corrections and adaptations to altered air-conducted auditory feedback are practical responses, given that the air conducted signal is the only one perceived by listeners during speech communication.

Neural networks that mediate voice pitch error correction during phonation likely involve midbrain and cortical connections. Müller-Preuss (1986) recorded auditory neurons from the inferior colliculus, dorsal nucleus of the lateral lemniscus, and parts of the reticular formation in the midbrain of the squirrel monkey during self-produced and recorded vocalizations. While a majority of those cells responded to both sounds, some cells located in the marginal regions of the inferior colliculus and an area of the reticular formation distinguished between the two. These source-distinguishing neurons (fewer than 10%) showed no response or only a very weak response during self-produced calls, even when there was a clear response to the playback of prerecorded calls of comparable intensity (Müller-Preuss, 1986). Similar selectivity has been noted in the cortex. Most neurons recorded from the superior temporal gyrus that are active during the presentation of recorded vocalizations are found to be silent during ongoing vocalization, indicating a suppression of sensitivity to self-generated sounds. In conjunction with preconscious pitch processing mechanisms, source-discriminating cellular activity such as this may underlie PSR triggering.

A comparison of the effects of auditory masking on mean voice pitch and intonation contours indicates that these two aspects of speech production may be separately controlled. Temporarily masking auditory feedback during oral reading of interrogative, declarative, and imperative sentences does not significantly degrade pitch contours, although average voice F_0 is raised (Mallard *et al.*, 1978). That auditory feedback masking has a limited effect on voice intonation contours is consistent with the hypothesis that these contours are generated in an open loop manner.

Short-latency feedback control of pitch could be useful given the possibility of unplanned voice F_0 perturbations resulting from aerodynamic instability, pulsing vascular and lymphatic fluids (Orlikoff and Baken, 1989), and laryngeal configuration changes that occur during speech articulation (Honda, 1983; Ternström, Sundberg, and Colldén, 1988; Watson, 1998). The findings of this study demonstrate the availability of a closed-loop mechanism for control of pitch during dynamic variations such as may occur during speech. Had the pitch-shift response been absent during glissando vocalizations it would have been inconsistent with a hypothesis of online pitch error correction during speech. Interest-

ingly, Natke and Kalveram (2002) demonstrated that voice pitch is monitored during speech sound production, even during brief, unstressed syllables. In their study, a sustained downward shift of one semitone prior to the onset of syllable repetition resulted in a vocalization that, after hearing the shifted initial syllable, became 25 to 69 cents higher than productions with normal auditory feedback. While the responses observed by Natke *et al.* may represent an attempt to maintain habitual speaking pitch rather than an attempt to correct a particular suprasegmental contour error, it is certainly possible that online monitoring of pitch inflection and correction of errors are used to preserve speech intelligibility. It remains to be determined whether the early pitch shift response is an important aspect of pitch control during speech production.

V. CONCLUSION

Prior to this study it was known that a perturbation in auditory feedback pitch typically generates a reflex-like, small-magnitude opposing voice pitch change. The function of this reflex was hypothesized to be pitch stabilization. It was not known whether this stabilization was restricted only to steady pitch levels or whether it was also operative during dynamic pitch modulations. Results of this study indicate that the automatic, early component of the pitch-shift response is not specific to tasks during which voice pitch is to be held steady. Rather, the pitch-shift response serves to stabilize voice pitch around an intentional pitch target or vector.

ACKNOWLEDGMENTS

This research was funded by NIH Grant No. DC02764. The authors are very grateful to Timothy Hain and Kimberly Fisher for suggestions regarding experimental design, to Rokny Akhavein for programming assistance, and to Lily Trofimovich, for contributions of statistical analysis.

[1]The term ''pitch'' is used in reference to production because the controlled variable in the composition of music is believed to be relative (semitones and cents), not absolute (Hertz).

Barnes, G., and Grealy, M. (**1992**). ''Predictive mechanisms of head-eye coordination and vestibulo-ocular reflex suppression in humans,'' J. Vestib. Res. **2**, 193–212.

Bristow-Johnson, R. (**1995**). ''A detailed analysis of a time-domain formant-corrected pitch-shifting algorithm,'' J. Audio Eng. Soc. **43**, 340–352.

Burnett, T., Senner, J., and Larson, C. (**1997**). ''Voice F_0 responses to pitch-shifted auditory feedback: a preliminary study,'' J. Voice **11**, 202–211.

Burnett, T. A., Freedland, M., Larson, C. R., and Hain, T. (**1998**). ''Voice F_0 responses to manipulations in pitch feedback,'' J. Acoust. Soc. Am. **103**, 3153–3161.

Elliott, L., and Niemoeller, A. (**1970**). ''The role of hearing in controlling voice fundamental frequency,'' Int. Audiol. **9**, 47–52.

Elman, J. (**1981**). ''Effects of frequency-shifted feedback on the pitch of vocal productions,'' J. Acoust. Soc. Am. **70**, 45–50.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., and Singh, S. (**2000**). ''Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex,'' Exp. Brain Res. **130**, 133–141.

Honda, K. (**1983**). ''Relationship between pitch control and vowel articulation,'' in *Vocal Fold Physiology*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego, CA), pp. 286–297.

Houde, J. F., and Jordan, M. I. (**1998**). ''Sensorimotor adaptation in speech production,'' Science **279**, 12213–12216.

Jones, J. A., and Munhall, K. G. (**2000**). ''Perceptual calibration of F_0 production: Evidence from feedback perturbation,'' J. Acoust. Soc. Am. **108**, 1246–1251.

Klatt, D. (**1973**). ''Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception,'' J. Acoust. Soc. Am. **53**, 8–16.

Koelsch, S., Schröger, E., and Tervaniemi, M. (**1999**). ''Superior pre-attentive auditory processing in musicians,'' NeuroReport **10**, 1309–1313.

Larson, C. R., Carrell, T. D., Senner, J. E., Burnett, T. A., and Nichols, L. L. (**1995**). ''A proposal for the study of voice F_0 control using the pitch shifting technique,'' in *Vocal Fold Physiology: Voice Quality Control*, edited by M. Hirano and Osamu Fujimura (Singular Publishing Group, Inc., San Diego, CA), pp. 321–331.

Larson, C. R., White, J. P., Freedland, M. B., and Burnett, T. A. (**1996**). ''Interactions between voluntary modulations and pitch-shifted feedback signals: Implications for neural control of voice pitch,'' in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular Publishing Group, Inc., San Diego, CA), pp. 279–289.

Mallard, A. R., Ringel, R. L., and Horii, Y. (**1978**). ''Sensory contributions to control of fundamental frequency of phonation,'' Folia Phoniatr. **30**, 199–213.

Müller-Preuss, P. (**1986**). ''On the mechanisms of call coding through auditory neurons in the squirrel monkey,'' Eur. Arch. Psychiatry Neurol. Sci. **236**, 50–55.

Natke, U., and Kalveram, K. (**2002**). ''Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables,'' J. Speech Hear. Res. **44**, 577–584.

Orlikoff, R. F., and Baken, R. J. (**1989**). ''The effect of the heartbeat on vocal fundamental frequency perturbation,'' J. Speech Hear. Res. **32**, 576–582.

Sapir, S. (**1983**). ''Human laryngeal responses to auditory stimulation,'' J. Acoust. Soc. Am. **73**, 315–321.

Sapir, S., McClean, M. D., and Luschei, E. S. (**1983**). ''Effects of frequency-modulated auditory tones on the voice fundamental frequency in humans,'' J. Acoust. Soc. Am. **73**, 1070–1073.

Sundberg, J. (**1979**). ''Maximum speed of pitch changes in singers and untrained subjects,'' J. Phonetics **7**, 71–79.

Ternström, S., Sundberg, J., and Colldén, A. (**1988**). ''Articulatory F_0 perturbations and auditory feedback,'' J. Speech Hear. Res. **31**, 187–192.

Tervaniemi, M., Schroger, E., and Naatanen, R. (**1997**). ''Pre-attentive processing of spectrally complex sounds with asynchronous onsets: an event-related potential study with human participants,'' Neurosci. Lett. **227**, 197–200.

Waterston, J., and Barnes, G. (**1991**). ''Vestibulo-ocular reflex suppression during high velocity head-free pursuit in normal participants,'' Acta Oto-Laryngol., Suppl. **481**, 272–276.

Watson, B. C. (**1998**). ''Fundamental frequency during phonetically governed devoicing in normal young and aged speakers,'' J. Acoust. Soc. Am. **103**, 3642–3647.

Rules for controlling low-dimensional vocal fold models with muscle activation

Ingo R. Titze[a)]
National Center of Voice and Speech, and Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242 and the Wilbur James Gould Voice Research Center, Denver Center for the Performing Arts, Denver, Colorado 80204

Brad H. Story
Department of Speech and Hearing Sciences, University of Arizona, Tuscon, Arizona 85721

(Received 18 June 1998; revised 3 May 2002; accepted 31 May 2002)

A low-dimensional, self-oscillation model of the vocal folds is used to capture three primary modes of vibration, a shear mode and two compressional modes. The shear mode is implemented with either two vertical masses or a rotating plate, and the compressional modes are implemented with an additional bar mass between the vertically stacked masses and the lateral boundary. The combination of these elements allows for the anatomically important body-cover differentiation of vocal fold tissues. It also allows for reconciliation of lumped-element mechanics with continuum mechanics, but in this reconciliation the oscillation region is restricted to a nearly rectangular glottis (as in all low-dimensional models) and a small effective thickness of vibration (<3 mm). The model is controlled by normalized activation levels of the cricothyroid (CT), thyroarytenoid (TA), lateral cricoarytenoid (LCA), and posterior cricoarytenoid (PCA) muscles, and lung pressure. An empirically derived set of rules converts these muscle activities into physical quantities such as vocal fold strain, adduction, glottal convergence, mass, thickness, depth, and stiffness. Results show that oscillation regions in muscle activation control spaces are similar to those measured by other investigations on human subjects. [DOI: 10.1121/1.1496080]

PACS numbers: 43.70.Bk [AL]

I. INTRODUCTION

This paper addresses an ongoing search for a link between laryngeal muscle activation and mechanical properties of simple mass-spring vocal fold models. Although finite element implementation of continuum models of the vocal folds (Alipour and Titze, 1983, 1996), the tongue (Wilhelms-Tricarico, 1995), and the velum (Berry *et al.*, 1999) offer the potential for embedding muscle contractions directly into the mathematical formulations, the appeal of simple lumped-element models of tissue remains. This is partly due to the conceptual simplicity of coupled masses and springs, and partly due to the interpretive power that nonlinear dynamics has to offer to low-dimensional models (two or three degrees of freedom).

Low-dimensional models of the vocal folds have traditionally been structured to capture two or three primary modes of tissue vibration, one shear mode and one or two compressional modes (for identification and nomenclature of a complete set of modes, see Titze and Strong, 1975; Berry and Titze, 1996). In a continuum model, the compressional modes do not truly represent tissue compression, because human tissues are incompressible at audio frequencies, but there is an *apparent* compression in the horizontal direction while there is an *apparent* expansion in the vertical direction, leaving the volume conserved. In fact, all vocal fold deformations involve shear (Titze and Talkin, 1979). The one-mass model of Flanagan and Landgraf (1968) captured only one compressional mode, the two-mass model of Ishizaka and Flanagan (1972) captured both a shear mode and a compressional mode, but not in an explicit body-cover manner. The translational-rotational model of Liljencrants (1991) was designed expressly to control the shear mode independently from the compressional mode, but the body-cover distinction was still not clear. The three-mass model of Story and Titze (1995) captured three modes, one shear mode and two compressional modes, with the extra degree of freedom allowing a body-cover differentiation. Thus, if low-dimensionality, mode separation, and body-cover differentiation are all of primary interest, the translational-rotational model of Liljencrants (with minor modifications) or the three-mass model of Story and Titze are logical choices.

The problem is that vocal fold parameters such as normal mode frequencies, stiffness, effective mass in vibration, vocal fold length, vocal fold thickness, and rest position may not be under direct control by the vocalist. Rather, perceptual dimensions such as loudness, pitch, register, tightness, and roughness are likely to govern the thought processes for activation of laryngeal and respiratory muscles. Thus, to understand the oscillatory characteristics of the vocal folds in perceptual and physiologically realistic control spaces, we need one set of rules that transform perceptual variables to muscle activations, and a second set of rules that transform muscle activations to geometrical and viscoelastic parameters of the

a)Electronic mail: ingo-titze@uiowa.edu

lumped element models. We address only the second set of rules here.

It must be stated at the outset that no physiologically realistic vocal fold model is simple. If simplicity is forced onto the mechanics of the model, then the rules for control are complicated, as will be seen in this paper. On the contrary, less complicated rules are needed for models that incorporate the three-dimensional (continuum) nature of vocal fold mechanics, but at the price of greater mathematical complexity.

II. A MINIMAL PARAMETER SET

As a minimum, the following parameters are needed to capture the lowest natural modes and to relate the lumped-element mechanics of the tissue layers to the distributed surface pressures in the glottis:

(1) A compressional stiffness (K) for the body of the vocal folds and a rotational shear stiffness (κ) for the cover to control the natural frequencies of oscillation (and less directly, the fundamental frequency F_o). These two stiffnesses can be related to two independent elastic constants of an elastic continuum, a Poisson ratio and a shear modulus.
(2) The length (L), thickness (T), and depth (D) of the vocal folds to define boundary constraints and the vocal fold dimensions.
(3) The depth of the cover (D_c) to separate the body tissue layer from the cover tissue layer in terms of effective mass of vibration.
(4) The glottal half-width (ξ_{02}) at the top of the vocal fold to control the adduction of the vocal processes.
(5) The net glottal convergence (ξ_c) from the bottom of the vocal folds to the top of the vocal folds to control medial surface geometry.
(6) The mucosal upheaval point along the medial surface of the vocal folds or, correspondingly, the nodal point (z_n) of the shear mode to control upper and lower amplitudes of vibration.

With the use of these parameters, we first begin to review the mechanics of vocal fold vibration as applied to a three degrees of freedom body-cover model. Muscle activation will be brought into the picture later by rules.

III. LOW-DIMENSIONAL BODY-COVER MODELS

We begin the quantification of self-sustained oscillation with a body-cover model that has only three degrees of freedom. As illustrated in Fig. 1 in coronal cross-section, the cover is represented by either two masses m_1 and m_2 and three springs k_1, k_2, and k_c (left side) or by a rotating plate of mass m with a torsional stiffness κ and a single spring k (right side). The body is represented by a mass M in both cases, connected through a compressible spring K to the thyroid cartilage boundary. We shall refer to these two models as the *three mass* and the *bar-plate* version of the body-cover model, respectively. The models have no variation perpendicular to the plane of the paper (the anterior-posterior direction, or y direction), which means that the masses are rigid

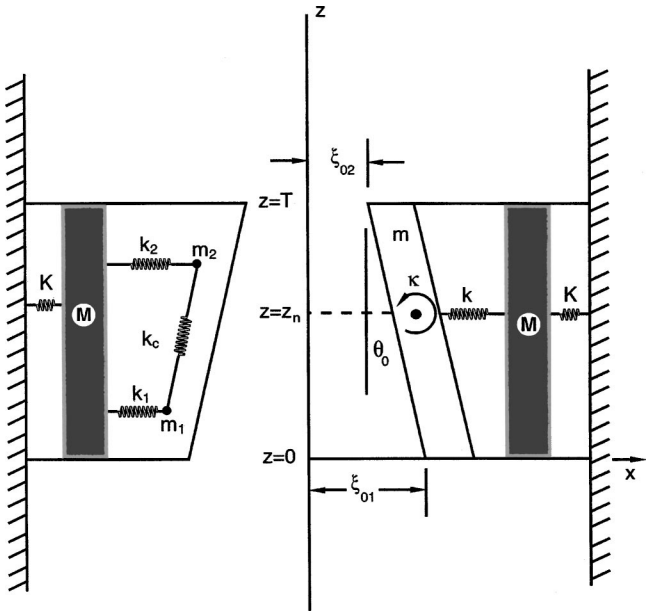


FIG. 1. Sketch of three-mass (left) and bar-plate (right) low-dimensional body-cover models in frontal cross section.

rods, plates, or bars. In the bar-plate model, the cover spring k attaches at the point of rotation (nodal point z_n) of the plate, which is not necessarily at the center of the bar. The z direction is the direction of airflow and the x direction is the direction of medial-lateral tissue displacement.

Unless otherwise stated, we make the following assumptions throughout the body-cover derivations: (1) The two vocal folds move symmetrically with respect to the glottal midplane ($x=0$). (2) There is no vertical displacement of tissue. (3) The nodal point z_n can move vertically, redistributing the mass in the cover. (4) The glottal area varies linearly from the bottom to top of the vocal folds.

The equations for glottal aerodynamics of this model have already been published (Titze, 2002) and will not be repeated here. Briefly, we assume Bernoulli flow from the lungs to the point of flow separation, at which point jet flow continues and the pressure remains constant in the jet, from flow detachment to glottal exit. Liljencrants' (1991) rules or those of Pelorson *et al.* (1994) for a slightly enlarged detachment area and a downstream detachment point have been incorporated. Furthermore, for this investigation, we have not included any vocal tract simulation. We realize that vocal fold vibration can be profoundly affected by the acoustic load of the vocal tract, but this will be treated in a follow-up investigation and does not impact on the rules developed here.

A. Equations of motion for three degrees of freedom

Consider first the motion of the vocal fold cover in the bar-plate version (right side of Fig. 1). This motion can be described with two degrees of freedom if we assume a rotation θ about the nodal point z_n and a translation ξ of the nodal point,

$$I_c\ddot{\theta}+B_c\dot{\theta}+\kappa\theta=T_a, \tag{1}$$

$$m\ddot{\xi}+b(\dot{\xi}-\dot{\xi}_b)+k(\xi-\xi_b)=F_a, \tag{2}$$

where T_a is the applied aerodynamic torque, I_c is the moment of inertia for rotation of the cover, B_c is the rotational damping, b is translational damping, F_a is the force at the nodal point, ξ_b is the displacement of the body, and other parameters are as previously defined. Similarly, the equation of motion for the body is written as

$$M\ddot{\xi}_b + b(\dot{\xi}_b - \dot{\xi}) + k(\xi_b - \xi) + K\xi_b + B\dot{\xi}_b = 0, \tag{3}$$

where B is the body damping.

The natural frequencies of this body-cover system can be obtained analytically by letting all damping forces and aerodynamic forces vanish (i.e., $b=B=B_c=T_a=F_a=0$) and assuming harmonic functions of the form $e^{i\omega t}$. The result is

$$\omega_2^2 = \kappa/I_c \tag{4}$$

for the shear mode, and

$$\omega_{1,3}^2 = \frac{k+K}{2M} + \frac{k}{2m} \pm \left[\left(\frac{k+K}{2M} + \frac{k}{2m} \right)^2 - \frac{kK}{mM} \right]^{1/2} \tag{5}$$

for the two compressional modes. As an asymptotic check, the compressional mode frequencies in Eq. (5) reduce to

$$\omega_{1,3}^2 = \frac{k}{m} \quad \text{for } M \gg m \tag{6}$$

$$= \frac{K}{M+m} \quad \text{for } k \gg K \tag{7}$$

$$= k\frac{M+m}{Mm} \quad \text{for } K=0 \tag{8}$$

$$= 0 \quad \text{for } k=0 \text{ or } m=\infty. \tag{9}$$

When the driving forces are not zero, Eqs. (1) and (2) are coupled by the fact that both the driving torque T_a and the driving force F_a on the cover are dependent on a common glottal flow.

The equations of motion for the three-mass model (left side of Fig. 1) are described in detail by Story and Titze (1995). We add only a variable mass distribution in the cover by letting m_1 and m_2 depend on the nodal point z_n,

$$m_1 = mz_n/T, \tag{10}$$

$$m_2 = m(1 - z_n/T), \tag{11}$$

$$k_1 = kz_n/T, \tag{12}$$

$$k_2 = k(1 - z_n/T), \tag{13}$$

where m and k are the bar-plate constants as defined above.

The relation between the vertical coupling stiffness k_c in the three-mass model and the rotational stiffness κ in the bar-plate model can be found by equating ω_2 (the normal mode frequency for shear) of the two models when the body mass M is rigid. This frequency is defined by Eq. (4) for the bar-plate model, but for the three-mass model a separate modal analysis is performed for the cover, with only the masses m_1 and m_2, and three springs k_1, k_2 and k_c. The result yields the following equivalence,

$$\omega_2^2 = \frac{\kappa}{I_c} = \frac{k_1+k_c}{2m_1} + \frac{k_2+k_c}{2m_2} + \left[\left(\frac{k_1+k_c}{2m_1} - \frac{k_2+k_c}{2m_2} \right)^2 + \frac{k_c^2}{m_1 m_2} \right]^{1/2}. \tag{14}$$

Substituting Eqs. (10)–(13) into Eq. (14) gives a relation for κ in terms of k_c,

$$\kappa = \frac{I_c}{m} \left[k + k_c \left(\frac{z_n}{T} \right)^{-1} \left(1 - \frac{z_n}{T} \right)^{-1} \right], \tag{15}$$

and its inverse

$$k_c = \left(\frac{m}{I_c}\kappa - k \right) \left(\frac{z_n}{T} \right) \left(1 - \frac{z_n}{T} \right). \tag{16}$$

We see that k_c in the three-mass model is a mixture of both shear (κ) and compression (k). This is because one of the three springs for a two-mass cover is essentially redundant.

When there is symmetry between the upper and lower mass, then

$$m_1 = m_2 = m/2, \tag{17}$$

$$k_1 = k_2 = k/2, \tag{18}$$

and the vertical coupling stiffness is related to the torsional stiffness as

$$k_c = \frac{1}{4} \left(\frac{m}{I_c}\kappa - k \right). \tag{19}$$

Equations (16) and (19) show that there are some restrictions on the choices of κ and k if the coupling stiffness k_c is to be a positive number. We will now explore these restrictions on the basis of continuum mechanics.

B. Relating lumped constants to elastic moduli

The purpose of adopting continuum-mechanical constants for modeling is that they are independent of geometry and, therefore, become the basic building blocks for both distributed and lumped element representations. For a negligible amount of fiber tension in the vocal folds, it is possible to express the translational and rotational stiffnesses in terms of one measurable isotropic shear modulus (Chan and Titze, 1997). Referring to continuum mechanics, the constitutive equations for transversely isotropic tissue with planar strain (Fung, 1993) can be written as

$$\sigma_x = \frac{2\mu}{(1-\nu)} \left(\frac{\partial \xi}{\partial x} + \nu \frac{\partial \zeta}{\partial z} \right) \tag{20}$$

and

$$\tau_{xz} = \mu \left(\frac{\partial \xi}{\partial z} + \frac{\partial \zeta}{\partial x} \right), \tag{21}$$

where σ_x is the normal stress in the x direction, τ_{xz} is the shear stress on a z plane in the x direction, μ is the shear modulus, ν is the Poisson ratio, ξ is horizontal displacement, and ξ is vertical displacement. The two independent elastic constants are μ and ν, but since we have already assumed $\zeta=0$ to reduce the number of degrees of freedom of the model, the tissue must now be treated as completely com-

pressible ($\nu=0$), leaving μ as the only elastic constant.

If we further assume that all the compression takes place linearly over a depth D, then the normal stress is

$$\sigma_x = 2\mu \frac{\partial \xi}{\partial x} = 2\mu \xi / D, \tag{22}$$

and the shear stress is

$$\tau_{xz} = \mu \frac{\partial \xi}{\partial z} = \mu \tan\theta \approx \mu\theta, \tag{23}$$

where θ is the torsional angle defined in Eq. (1), which is typically assumed to be small in linear elasticity theory.

The normal stress σ_x in Eq. (22) can be converted to a net restoring force by multiplying this stress by the cross sectional area of the tissue LT,

$$F_x = LT\sigma_x = k\xi. \tag{24}$$

When σ_x from Eq. (22) is now substituted into Eq. (24), the spring constant k in the cover becomes

$$k = 2\mu_c LT/D_c, \tag{25}$$

where μ_c is the shear modulus in the cover and D_c is the depth of the cover. For the body, we can similarly write

$$K = 2\mu_b LT/D_b, \tag{26}$$

where μ_b is the shear modulus of the body and D_b is the depth of the body. The constants k_1 and k_2 for the three-mass model are obtained by direct substitution of the above into Eqs. (12) and (13).

For the shear stress in the cover, τ_{xz} is multiplied by the horizontal shear area LD_c to get a shear restoring force

$$F_\theta = LD_c\mu_c\theta. \tag{27}$$

This shear force is then multiplied by an upper moment arm $(T-z_n)/2$ and a lower moment arm $z_n/2$ to get a restoring torque,

$$T_\theta = (LD_c\mu_c\theta)\left[\frac{T-z_n}{2} + \frac{z_n}{2}\right] = (LD_c\mu_c\theta)\frac{T}{2}. \tag{28}$$

This torque is divided by the angular strain θ to get the torsional stiffness:

$$\kappa = \tfrac{1}{2}\mu_c LTD_c. \tag{29}$$

Note that the lumped-element spring constants k, K, and κ are all directly proportional to the shear modulus of the tissue, but these spring constants also contain the geometrical factors length, thickness, and depth of the tissue layers. In particular, all stiffnesses increase linearly with vocal fold length and thickness, but rotational stiffness increases with tissue depth [Eq. (29)], whereas compressional stiffnesses decrease with tissue depth [Eqs. (25) and (26)]. Thus, tissue depths are critical for modeling when lumped-element parameters are used.

The effective masses of the body and cover are related to the tissue density ρ and the volume of tissue in vibration,

$$m = \rho LTD_c, \tag{30}$$

$$M = \rho LTD_b. \tag{31}$$

The moment of inertia I_c of the cover can also be derived from basic principles as

$$I_c = \int_0^{z_n} (z_n - z)^2\, dm + \int_{z_n}^{T} (z - z_n)^2\, dm \tag{32}$$

$$= \rho LD_c \int_0^T (z - z_n)^2\, dz \tag{33}$$

$$= \rho LD_c T^3 \left[\frac{1}{3} - \frac{z_n}{T}\left(1 - \frac{z_n}{T}\right)\right]. \tag{34}$$

With these derivations, the natural frequencies of the bar-plate model [Eqs. (4) and (5)] can be calculated analytically in terms of the elastic constants μ_c and μ_b, the tissue density ρ, and vocal fold geometry,

$$F_2 = \frac{1}{2\pi}\sqrt{\frac{\kappa}{I_c}} = \frac{1}{\pi T}\sqrt{\frac{3\mu_c}{4\rho}}\left[\frac{1}{3} - \frac{z_n}{T}\left(1 - \frac{z_n}{T}\right)\right]^{-1/2}, \tag{35}$$

$$F_{1,3} = \frac{1}{2\pi\sqrt{\rho}}\left\{\frac{\mu_c}{D_c^2} + \frac{\mu_c}{D_cD_b} + \frac{\mu_b}{D_b^2} \pm \left[\left(\frac{\mu_c}{D_c^2} + \frac{\mu_c}{D_cD_b} + \frac{\mu_b}{D_b^2}\right)^2 - \frac{4\mu_c\mu_b}{D_b^2D_c^2}\right]^{1/2}\right\}^{1/2}. \tag{36}$$

Note that the nodal point z_n and the vocal fold thickness T affect only the shear mode frequency F_2, whereas tissue depths D_c and D_b affect only the two compressional mode frequencies F_1 and F_3. This makes it easy for the user to adjust the normal modes with geometric factors.

Analytical calculation of the natural frequencies of the three-mass model is also possible, but involves cube-root terms that are mathematically quite cumbersome and take up excessive space.

Note that all three natural frequencies in Eqs. (35) and (36) are proportional to the square root of a ratio between a shear modulus and the tissue density, e.g., $(\mu_c/\rho)^{1/2}$, which dimensionally is a tissue wave velocity. The shear moduli μ_c and μ_b have been measured in the laboratory (Chan and Titze, 1997, 1999). For excised cover tissue $\mu_c \approx 500$ Pa and for excised thyroarytenoid muscle tissue, $\mu_b \approx 1000$ Pa, but these moduli can vary significantly depending on moisture content, temperature, and frequency. The density is less variable, $\rho = 1040$ kg/m^3, giving a tissue wave velocity $(\mu/\rho)^{1/2}$ in the range of 0.7–1.0 m/s, which agrees with measurement of the mucosal wave (Baer, 1975; Titze *et al.*, 1993).

Equation (16) gave a relation between the vertical coupling spring constant k_c of the three-mass model and the parameters of the bar-plate model. Substituting into Eq. (16) the mass of the cover m from Eq. (30), the moment of inertia I_c from Eq. (34), the torsional spring constant κ from Eq. (29), and the stiffness k from Eq. (25), the following inequality must hold for k_c to be a positive number,

$$D_c > 2T\left[\frac{1}{3} - \frac{z_n}{T}\left(1 - \frac{z_n}{T}\right)\right]^{1/2}. \tag{37}$$

As an example, this inequality simplifies to $D_c > 0.577T$ for $z_n/T = 0.5$, a fairly stringent requirement for the depth-to-thickness-ratio of the cover. It points out one of the limitations of the low-dimensional models in achieving a reconcili-

ation with continuum mechanics. As the nodal point shifts up and down, with z_n/T approaching 0 or 1, the restriction becomes even greater. The problem is that a solid bar mass M was chosen for the body. No shear can take place in this element, making the shear depth for the cover effectively too thin. Our rules will pay attention to this problem by limiting vocal fold thickness.

C. Inclusion of fiber stress in tissues

The foregoing equivalence between continuum mechanics and lumped-element mechanics applies only to the case where there is tissue isotropy, i.e., where there is a negligible anterior–posterior tension in the tissue fibers (muscle and ligament). But the spring constants can be augmented to account for anisotropic (stringlike) tension in the fibers by equating the natural frequency of a mass-spring mode of vibration to an ideal string mode of vibration,

$$\frac{1}{2\pi}\sqrt{\frac{K'}{M}}=\frac{1}{2L}\sqrt{\frac{\sigma}{\rho}}, \tag{38}$$

where σ is the fiber stress and K' is the added fiber stiffness. Solving for K' and letting $M=\rho LTD$, the added stiffness is

$$K'=\left(\frac{\pi}{L}\right)^2\sigma LTD=\pi^2\sigma DT/L. \tag{39}$$

For rotational movement, K' is replaced by κ' and M is replaced by I_c in the above. The augmented lumped-element parameters now become the following:

Bar-plate model (cover):

$$\kappa=\frac{1}{2}\mu_c LTD_c+\pi^2\sigma_c(D_c/L)T^3\left[\frac{1}{3}-\frac{z_n}{T}\left(1-\frac{z_n}{T}\right)\right], \tag{40}$$

$$k=2\mu_c(LT/D_c)+\pi^2\sigma_c(D_c/L)T, \tag{41}$$

$$m=\rho LTD_c. \tag{42}$$

Three-mass model (cover):

$$I_c=\rho LTD_cT^2\left[\frac{1}{3}-\frac{z_n}{T}\left(1-\frac{z_n}{T}\right)\right], \tag{43}$$

$$k_1=2\mu_c(LT/D_c)z_n/T+\pi^2\sigma_c(D_c/L)z_n, \tag{44}$$

$$k_2=2\mu_c(LT/D_c)(1-z_n/T)+\pi^2\sigma_c(D_c/L)\times T(1-z_n/T), \tag{45}$$

$$k_c=\left\{\frac{1}{2}\mu_c(LD_c/T)\left[\frac{1}{3}-\frac{z_n}{T}\left(1-\frac{z_n}{T}\right)\right]^{-1}-2\mu_c(LT/D_c)\right\}\frac{z_n}{T}\left(1-\frac{z_n}{T}\right), \tag{46}$$

$$m_1=\rho LTD_cz_n/T, \tag{47}$$

$$m_2=\rho LTD_c(1-z_n/T). \tag{48}$$

Both models (body):

$$K=2\mu_bLT/D_b+\pi^2\sigma_b(D_b/L)T, \tag{49}$$

$$M=\rho LTD_b. \tag{50}$$

Note that for this nonisotropic tissue, all stiffnesses now depend on all vocal fold dimensions. It is no longer a simple matter to adjust shear or compression stiffness by vocal fold geometry. The anatomical resting depths of the layers of vocal fold tissue are gleaned from Hirano (1975):

$D_{\text{muc}}=$ depth of mucosa
$=$ epithelium + superficial layer
$=0.2$ cm in males (0.15 cm in females),

$D_{\text{lig}}=$ depth of ligament
$=$ intermediate + deep layer
$=0.2$ cm in males (0.15 cm in females),

$D_{\text{mus}}=$ depth of TA muscle (thyrovocalic portion)
$=0.4$ cm in males (0.3 cm in females).

According to Hirano (1975), the body and the cover share the depth of the ligament. The medial half, known as the intermediate layer of the lamina propria, is assigned to the cover, whereas the lateral half, known as the deep layer of the lamina propria, is assigned to the body. The fiber stress in the body is then made up of both muscle stress and ligament stress. But stresses do not add in parallel layers; rather, forces do. Hence, we convert the stresses to forces by multiplying by cross-sectional areas, or (since thickness is common) by the respective depths,

$$\sigma_b=[0.5\sigma_{\text{lig}}D_{\text{lig}}+\sigma_{\text{mus}}D_{\text{mus}}]/D_b, \tag{51}$$

where σ_{lig} is the ligament stress and σ_{mus} is the muscle stress. Likewise, for the cover,

$$\sigma_c=[\sigma_{\text{muc}}D_{\text{muc}}+0.5\sigma_{\text{lig}}D_{\text{lig}}]/D_c, \tag{52}$$

where σ_{muc} is the mucosa stress.

The muscle stress has an active and a passive component (Alipour-Haghighi *et al.*, 1989)

$$\sigma_{\text{mus}}=a_{\text{TA}}\sigma_{\text{am}}\,\text{Max}[0,1-b(\epsilon-\epsilon_m)^2]+\sigma_p, \tag{53}$$

where σ_{am} is the maximum active stress in the TA muscle fibers, a_{TA} is the muscle activity (ranging from 0 to 1), and σ_p is the passive stress. Note that the active stress varies with vocal fold strain ϵ, falling off quadratically on both sides of an optimum sarcomere strain ϵ_m, with b being an empirically determined constant for a given muscle. The constants are given in Table I.

The cover has only a passive stress. This passive stress has been modeled with a combination of a linear and an exponential function and matched to vocal fold stress–strain curves (Alipour and Titze, 1991; Min *et al.*, 1995). The form of this function applies also to the passive stress of the body (muscle tissue and ligament tissue) and is written for all tissue types as

$$\sigma_p=\begin{cases}0 & \text{for } \epsilon<\epsilon_1,\\ -\dfrac{\sigma_0}{\epsilon_1}(\epsilon-\epsilon_1) & \text{for } \epsilon_1\leqslant\epsilon\leqslant\epsilon_2,\\ -\dfrac{\sigma_0}{\epsilon_1}(\epsilon-\epsilon_1)+\sigma_2[e^{C(\epsilon-\epsilon_2)}-C(\epsilon-\epsilon_2)-1] & \text{for } \epsilon>\epsilon_2.\end{cases} \tag{54}$$

TABLE I. Numerical constants used in Eqs. (53) and (54) to generate stress-strain curves for the cover, ligament, and muscle layers.

Layer	ϵ_1	ϵ_2	σ_0 (kPa)	σ_2 (kPa)	C	σ_m (kPa)	ϵ_m	b
mucosa	−0.5	0.35	0.5	30.0	4.4	⋯	⋯	⋯
ligament	−0.5	−0.00	0.4	1.393	17.0	⋯	⋯	⋯
TA muscle	−0.5	−0.05	1.0	1.50	6.5	105	0.4	1.07

In Eq. (54), σ_o is the stress when $\epsilon=0$, ϵ_1 is the strain where the linear portion goes to zero, σ_2 is a scale factor for the exponential portion, and ϵ_2 is the strain where the exponential portion begins. These constants have been fitted individually to each stress–strain curve and are reported in Table I.

This concludes the physics of the models for natural (undamped and undriven) oscillation. For damped oscillation, it is typical to assign a damping ratio of 0.1–0.2 to each mass (Kaneko *et al.*, 1981; Chan and Titze, 1999). But for self-sustained oscillation, the low-dimensional models require a higher degree of damping for the less dominant modes. This broadens their frequency spectrum and allows for better entrainment to the dominant mode, which is usually the compressional mode of the body. This was already discovered by Ishizaka and Flanagan (1972) with their two-mass model. They made the compressional mode of the lower mass dominant, assigning it a damping ratio of 0.1. The upper mass was given a damping ratio of 0.6 for better entrainment. We will also use this value for the upper mass, while keeping the lower mass and the body mass at a damping ratio of 0.1.

IV. DEVELOPMENT OF RULES FOR MUSCLE CONTROL

In this section we develop rules for controlling the geometric and elastic parameters of the body-cover model by muscle activation. These rules take the place of the physics of deformable bodies for vocal fold posturing (adducting and elongating). To reduce the number of control parameters, only the cricothyroid activity (a_{CT}), the thyroarytenoid activity (a_{TA}), and the lateral cricoarytenoid activity (a_{LC}) will be considered. The effect of the interarytenoid muscle is neglected and the effect of the posterior cricoarytenoid muscle is included by allowing a_{LC} to become negative. Some of the rules are not yet in their final stage of development, but are quite functional at this point. We begin with the elongation rule.

A. Elongation rule

Experimentation with excised larynges and *in vivo* animal preparations (Titze *et al.*, 1988, 1997) has shown that vocal fold elongation can be expressed in the form

$$\epsilon = G(Ra_{\mathrm{CT}} - a_{\mathrm{TA}}) - Ha_{\mathrm{LC}}, \tag{55}$$

where ϵ is the longitudinal vocal fold strain (elongation divided by the cadaveric rest length), a_{CT} is the normalized cricothyroid muscle activity (ranging from 0.0 to 1.0), a_{TA} is the normalized thyroarytenoid muscle activity (same range), and a_{LC} is the normalized lateral cricoarytenoid muscle activity (same range). The empirical constants in the equation are the gain of elongation G, the torque ratio R, and the adductory strain factor H. This rule has been modified from a previous rule (Titze *et al.*, 1988) to include the adductory strain Ha_{LC} that occurs for prephonatory posturing. This inclusion left the nature of the equation the same, but changes the coefficients slightly. For this study, we are letting $G=0.2$, $R=3.0$, and $H=0.2$. The previous rule was for canines, whereas the current version is intended for humans, for which the range of ϵ must be higher to achieve a larger pitch range. Thus, G was increased from 0.1 to 0.2 for humans. With this gain, the maximum elongation is 60% when $a_{\mathrm{CT}}=1.0$ and $a_{\mathrm{TA}}=a_{\mathrm{LC}}=0$. This replaces the maximumal superior nerve stimulation condition in the Titze *et al.* (1997) study, for which the canine vocal folds elongated a maximum of 45%. The maximum shortening in Eq. (55) occurs for $a_{\mathrm{CT}}=0$ and $a_{\mathrm{TA}}=a_{\mathrm{LC}}=1.0$. For this case, the rule gives −40%, as compared to the measured canine value of −17% under maximal recurrent nerve stimulation. With full contraction of all muscles, the rule gives 20% elongation as compared to the measured value of 26% for canines.

We cannot expect phonation to be realizeable over the full range of elongation. Some of the extreme muscle contractions stiffen the vocal folds too much or shut off the glottis. Nishizawa *et al.* (1988) observed an approximate 2:1 ratio between the shortest and the longest vocal folds when subjects phonated over wide pitch ranges. In the rule given above, the range of ϵ is −30% to 50% when a_{LC} is chosen to be 0.5. This corresponds to a length change of $0.7L_o$ to $1.5L_o$, where L_o is the resting length, a range quite similar to that reported by Nishizawa *et al.*

With the above rule, the vocal fold length can be written as

$$L = L_o(1+\epsilon) \tag{56}$$

$$= L_o[1 + G(Ra_{\mathrm{CT}} - a_{\mathrm{TA}}) - Ha_{\mathrm{LC}}], \tag{57}$$

where L_o is 1.6 cm in males and 1.0 cm in females.

B. Nodal point rule

When the TA muscle contracts, the bottom edge adducts more than the top edge, reducing the convergence angle (see convergence rule below). Concurrently, the nodal point z_n for the shear mode effectively moves up on the medial surface, suggesting that there is greater vibrational amplitude at the bottom relative to the top. Empirically, this is related to the point of mucosal upheaval (Yumoto *et al.*, 1993). We propose the rule

$$z_n = (1 + a_{\mathrm{TA}})T/3. \tag{58}$$

With this rule, the nodal point is one-third from the bottom for falsetto register (thin vocal folds and greatest amplitude at the top) and one-third from the top in modal register (thick vocal folds with greater amplitude at the bottom). This rule is one of the weaker ones and will probably undergo modifications when more data become available.

C. Thickness and depth rules

In a purely passive sense, vocal fold thickness increases with vocal fold shortening. The Poisson ratio of continuous elastic media determines this increase. For an incompressible isotropic medium, the Poisson ratio is 0.5, but since medial-lateral expansion or contraction is constrained by the arytenoid and thyroid cartilage boundaries, most (or all) of the length change is absorbed by thickness change. Thus, our thickness rule is

$$T = \frac{T_o}{1+0.8\epsilon}, \tag{59}$$

where T_o is the vibrating thickness at resting length. Since the low-dimensional models do not have a medial surface curvature, we limit T_o to be 0.30 cm for males and 0.20 cm for females, as discussed earlier with regard to Eq. (37).

The depth rules are then as follows:

$$D_b = \frac{a_{\mathrm{TA}} D_{\mathrm{mus}} + 0.5 D_{\mathrm{lig}}}{1+0.2\epsilon}, \tag{60}$$

$$D_c = \frac{D_{\mathrm{muc}} + 0.5 D_{\mathrm{lig}}}{1+0.2\epsilon}, \tag{61}$$

where the 0.2ϵ factor in the denominator represents the complement to the 0.8ϵ factor for thickness. In the numerator, the terms represent portions of the mucosa, ligament, and muscle. For $a_{\mathrm{TA}}=0$, the deep layer of the lamina propria (half of the ligament) is the sole depth of the body; whereas, for $a_{\mathrm{TA}}=1$, the entire depth of the muscle is added to the body depth.

D. Adduction rule

The following adduction rule for the glottal half-width ξ_{02} at the vocal fold processes (Fig. 1) is chosen on the basis of fiberscopic measurements:

$$\xi_{02} = 0.25 L_o (1 - 2.0 a_{\mathrm{LC}}), \tag{62}$$

where a_{LC} is the activation of the lateral cricoarytenoid muscle. This rule anchors the "just touching" ($\xi_{02}=0$) position at $a_{\mathrm{LC}}=0.5$. For $a_{\mathrm{LC}}>0.5$ the vocal processes are pressed together (actually overlapped). For $a_{\mathrm{LC}}<0.5$, the vocal processes are separated. We allow a_{LC} to include the range 0 to -1 for *abduction* of the vocal processes, thereby simulating the action of the posterior cricoarytenoid muscle as an antagonist in this range.

Data by Scherer (1995) suggest that oscillation is obtained over about 10% of the range of motion of the vocal process. This range of motion was expressed by Scherer as a vocal process gap to length ratio ($2\xi_{02}/L_o$). Scherer's data were obtained with fiberscopic measurements on human subjects. A range of 0–1.5 for $2\xi_{02}/L_o$ is achieved in our rule when a_{LC} ranges from 0.5 to -1. This follows Scherer's claim that the maximum glottal gap at the vocal processes is about 50% greater than the vocal fold length for rapid inhalation.

The cadaveric (rest) position occurs for $a_{\mathrm{LC}}=0$, which yields a vocal process gap to length ratio of 0.5 in our rule. The ratio reported by Scherer for quiet respiration was 0.68, which may not represent exactly zero muscle activation. In our rule, Scherer's ratio of 0.68 would be slightly abductory activation ($a_{\mathrm{LC}}=-0.18$). For phonation threshold, Scherer claims a process gap to a length ratio of 0.17, which by our rule would occur for $a_{\mathrm{LC}}=0.33$ and would be 11% of the range of 0 to 1.5.

E. Convergence rule

Glottal convergence and medial surface bulging are, to a large extent, governed by activation of the thyroarytenoid muscle (Hirano, 1975). The critical issue for self-sustained oscillation is that a portion of the medial surface of the vocal folds must always be nearly vertical to produce a section of the glottis that is nearly rectangular (Chan *et al.*, 1997). This nearly rectangular glottal section can be achieved over a large portion of the human vocal fold thickness in modal register, since the medial surface has convex curvature (bottom to top), but only near the top of the fold in falsetto register. It would require at least three masses to be stacked vertically in a model to approximate this quadratic surface and thereby obtain a partial rectangular glottis. With the low-dimensional models under consideration here, three vertically stacked masses is not an option, which severely restricts the range of convergence under which the models can operate. With reference to Fig. 1, we propose the following rule for convergence:

$$\xi_c = \xi_{01} - \xi_{02} = T(0.05 - 0.15 a_{\mathrm{TA}}). \tag{63}$$

With this rule, the range of $\tan\theta_0 = \xi_c/T$ is from 0.05 to -0.1, which produces shapes from slightly convergent ($+3°$ at $a_{\mathrm{TA}}=0$) to moderately divergent ($-6°$ at $a_{\mathrm{TA}}=1$). But even with this restricted range of convergence and divergence, not all of these shapes produce self-sustained oscillation, as we will show. A rectangular glottis, which was shown to be ideal for phonation in a physical model (Chan *et al.*, 1997), is obtained for $a_{\mathrm{TA}}=0.33$. We will show that the best configu-

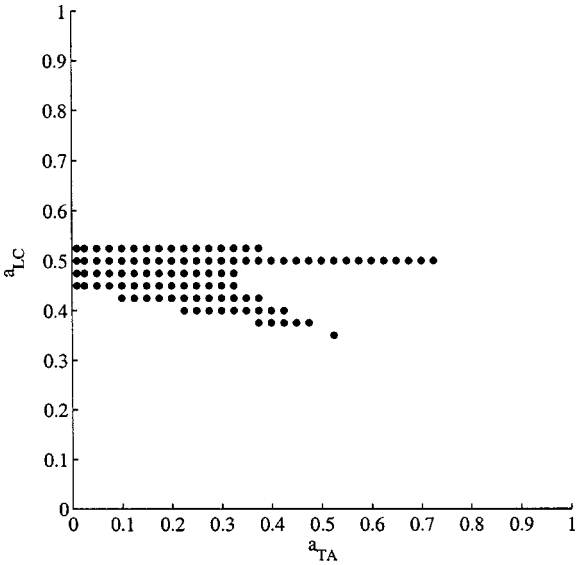


FIG. 2. Muscle activation plot (MAP) for lateral cricoarytenoid activity (a_{LC}) versus thryoarytenoid activity (a_{TA}). Data points show region of self-sustained oscillation with the convergence rule [Eq. (63)]. Cricothyroid activity (a_{CT}) was fixed at 0.2 and lung pressure was fixed at 0.8 kPa.

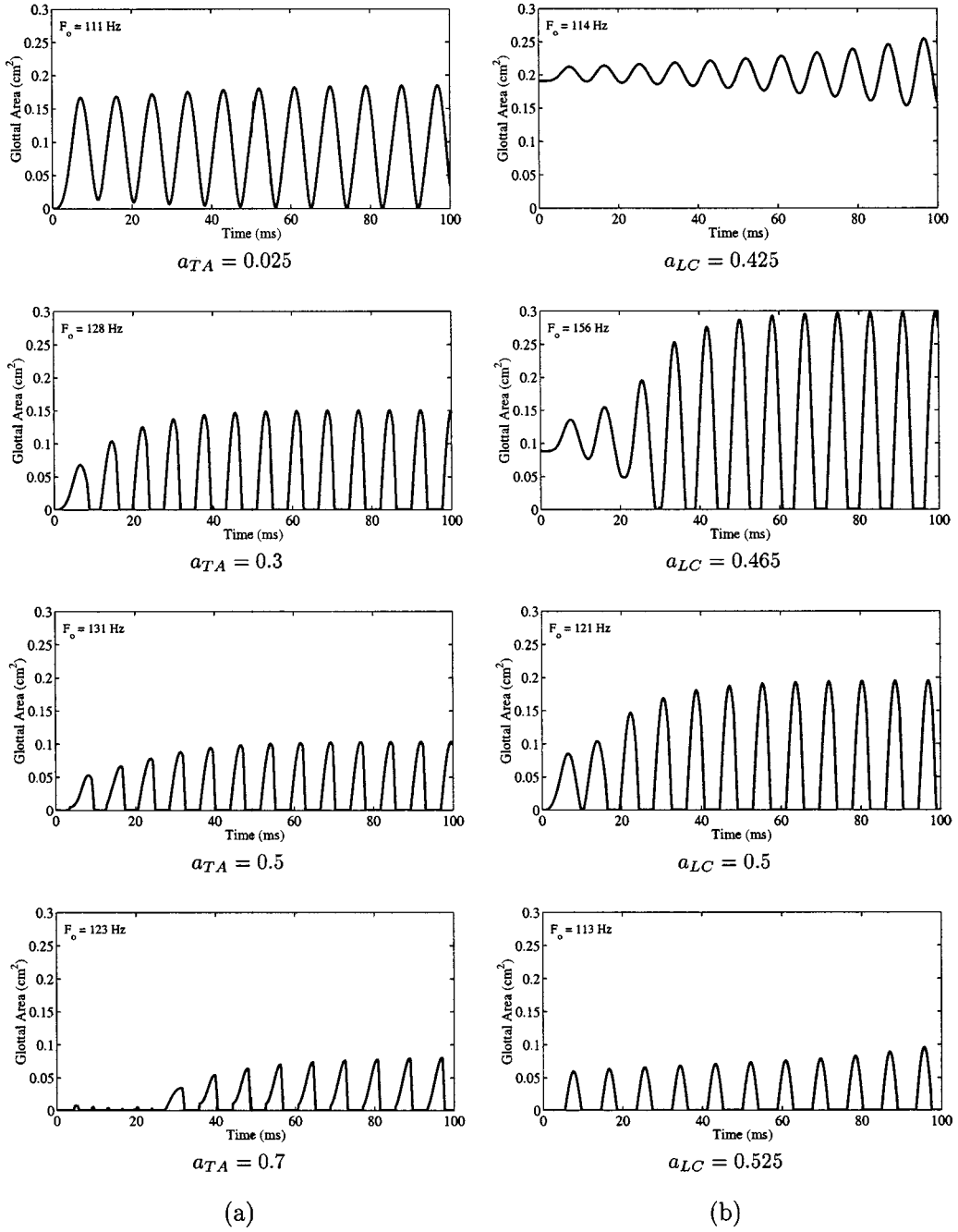


FIG. 3. Glottal area waveforms for selected sections through the adductory MAP of Fig. 2. (a) $a_{\mathrm{LC}}=0.5$ and (b) $a_{\mathrm{TA}}=0.25$. In both cases, a_{CT} was fixed at 0.2 and lung pressure was fixed at 0.8 kPa.

ration is an almost negligibly small convergence ($\theta \approx 0.0001$), which allows for complete adduction at the top ($\xi_{02}=0$) while full subglottal pressure is applied to the vocal folds when the bottom is barely abducted ($\xi_{01}=0.0005\,\mathrm{cm}$). We shall refer to this as the ''nearly rectangular'' case.

V. RESULTS

Figure 2 shows the region of oscillation of the bar-plate model in an adductory muscle activation plot (adductory MAP). The figure was created by performing separate 200 ms simulations for multiple pairs of TA and LCA activities. TA muscle activity is plotted horizontally, LCA activity is plotted vertically, and each dot denotes a pair of activities at which oscillation was sustained. Lung pressure was constant at 0.8 kPa (approximately 8 cm H_2O), a_{CT} was held constant at 0.2, and the convergence rule was used [Eq. (63)]. Existence of self-sustained phonation at each activation pair was determined by applying a zero-crossing detector to the last 100 ms of the simulated glottal area signal. Note that phonation is restricted to a small region of this MAP, as was expected from the results of Scherer (1995), on which our adduction rule was based. Oscillation occurs approximately along a constant horizontal line ($a_{\mathrm{LC}}=0.5$), but there appears to be a second (downward sloping) path for which an increase in a_{TA} must be counterbalanced with a slight decrease in a_{LC}.

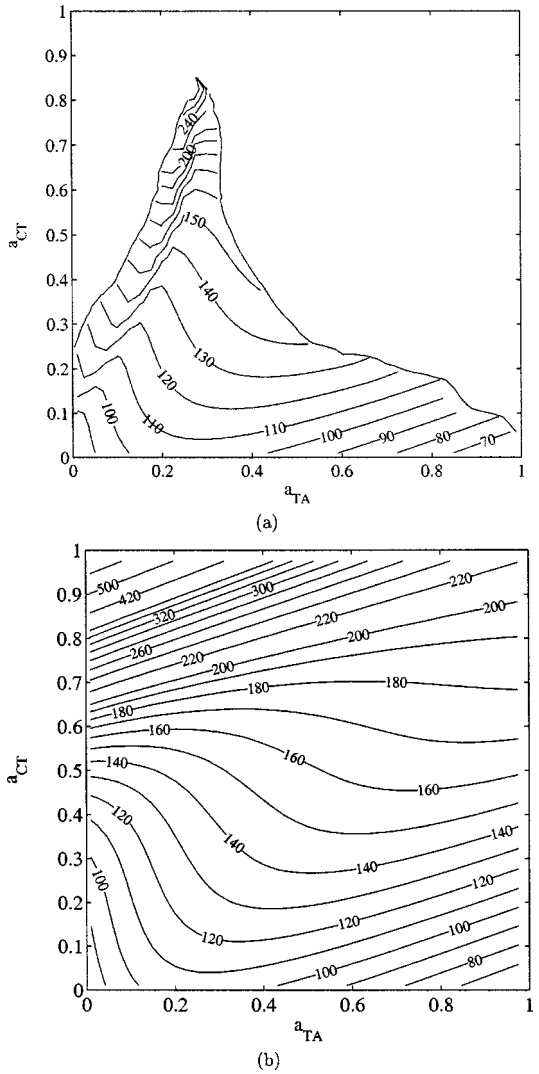


FIG. 4. Muscle activation plot (MAP) for cricothyroid activity (a_{CT}) versus thyroarytenoid activity (a_{TA}) with F_o contour lines. (a) The convergence rule [Eq. (63)] and (b) ''nearly rectangular'' convergence.

Figure 3 shows glottal area waveforms associated with the adductory MAP of Fig. 2. Part (a) shows waveforms for a *horizontal* section through the MAP at $a_{LC}=0.5$. Thyroarytenoid activity (a_{TA}) increases from top to bottom. Note that the glottal area is nearly sinusoidal at $a_{TA}=0.025$, but increases in both skewing and glottal closure when a_{TA} is increased to 0.7. Voice onset is delayed with $a_{TA}=0.7$. Note also that the amplitude of the glottal area decreases with a_{TA}, which is expected with greater adduction at the bottom of the fold and increased mass of the body (*vis-a-vis* our rules).

Figure 3(b) shows waveforms for a *vertical* section through the adductor MAP, namely for $a_{TA}=0.25$. Lateral cricoarytenoid activity (a_{LC}) now increases from 0.425 at the top to 0.525 at the bottom. Note that there is first an increase in glottal area amplitude (from very slow voice onset to rapid onset with large peak-to-peak variations), followed by a decrease in glottal area amplitude. This supports the notion that there is an optimal adduction for maximal power transfer in the ''pressed voice'' to ''breathy voice'' continuum (Titze, 2002).

Glottal airflow waveforms are not shown here because they are not remarkably different from the area waveforms. We have deliberately decoupled the vocal tracts (subglottal and supraglottal) for this investigation because its inclusion would have complicated the parameter set. Hence, for this decoupled vocal tract case, all of the glottal airflow waveforms looked similar to the area waveforms, with the typically larger skewing quotient of airflow (due to vocal tract inertance) being absent.

Figure 4 is a muscle activation plot (MAP) for elongating and tensing the vocal folds, showing a_{CT} on the vertical axis and a_{TA} on the horizontal axis. Contour lines are for constant F_o in Hz. Part (a) is for the convergence rule [Eq. (63)] while part (b) is for the ''nearly rectangular'' case. Note that the ''nearly rectangular'' configuration (part b) produces oscillation over the entire MAP (all values of a_{CT} and a_{TA}) whereas the rule-based configuration (part a) is severely restricted in the upper quadrants. In particular, for large values of a_{CT} (high frequencies), only a small range of values of a_{TA} (between 0.2 and 0.4) produces oscillation. It seems to be important to keep the glottis slightly convergent (below $a_{TA}=0.33$). This generates a strong asymmetry in the aerodynamic driving forces (a push-pull mechanism) for outward versus inward movement of the vocal folds (Titze, 1994, Chap. 4).

An important feature of these MAPs is the downward bending of the contour lines in the lower left quadrant, where most speech occurs. Because of this downward bending, F_o can be raised by increasing either a_{CT} or a_{TA}. In other quadrants (especially the upper left), an increase in a_{TA} will generally cause a decrease in F_o. (For a thorough review of this F_o control mechanism see Titze, 1994, Chap. 8.)

Figure 5 shows glottal area waveforms for vertical and horizontal sections through the MAP of Fig. 4(b). On the left [Fig. 5(a)] are waveforms for a vertical section (with a_{TA} constant at 0.25 and a_{CT} ranging from 0.05 to 0.9); on the right [Fig. 5(b)] are waveforms for a horizontal section (with a_{CT} constant at 0.25 and a_{TA} ranging from 0.05 to 0.9). Adduction is held constant throughout at $a_{LC}=0.5$ and convergence is the ''nearly rectangular'' case. Note that the effect of increasing a_{CT} (left side, top to bottom) is to increase F_o slightly at first (for values of $a_{CT}<0.5$) but more dramatically later (for values of $a_{CT}>0.5$). The amplitude of vibration decreases with increasing a_{CT} and there is a slight decrease of adduction with this amplitude change. The effect of increasing a_{TA} (right side, top to bottom) is to obtain a small increase in F_o at first, followed by a small decrease (when $a_{TA}>0.5$). This is a direct result of the curvature of the constant F_o lines in Fig. 4. The amplitude also decreases with increasing a_{TA}, and there is an increase in waveform skewing (seen best in the bottom right plot).

Figure 6 shows tensor MAPs with mode frequencies F_1, F_2, and F_3 as contour parameters. Note that for mode 1, the principal compressional mode, the MAP exhibits downward sloping F_1 curves in the lower left quadrant. In fact, the MAP is strikingly similar to the self-oscillation MAP of Fig. 4(b), suggesting that F_o is controlled by F_1, the compressional mode in the body. Again, the downward sloping curves allow for maximal increase in frequency with a_{TA},

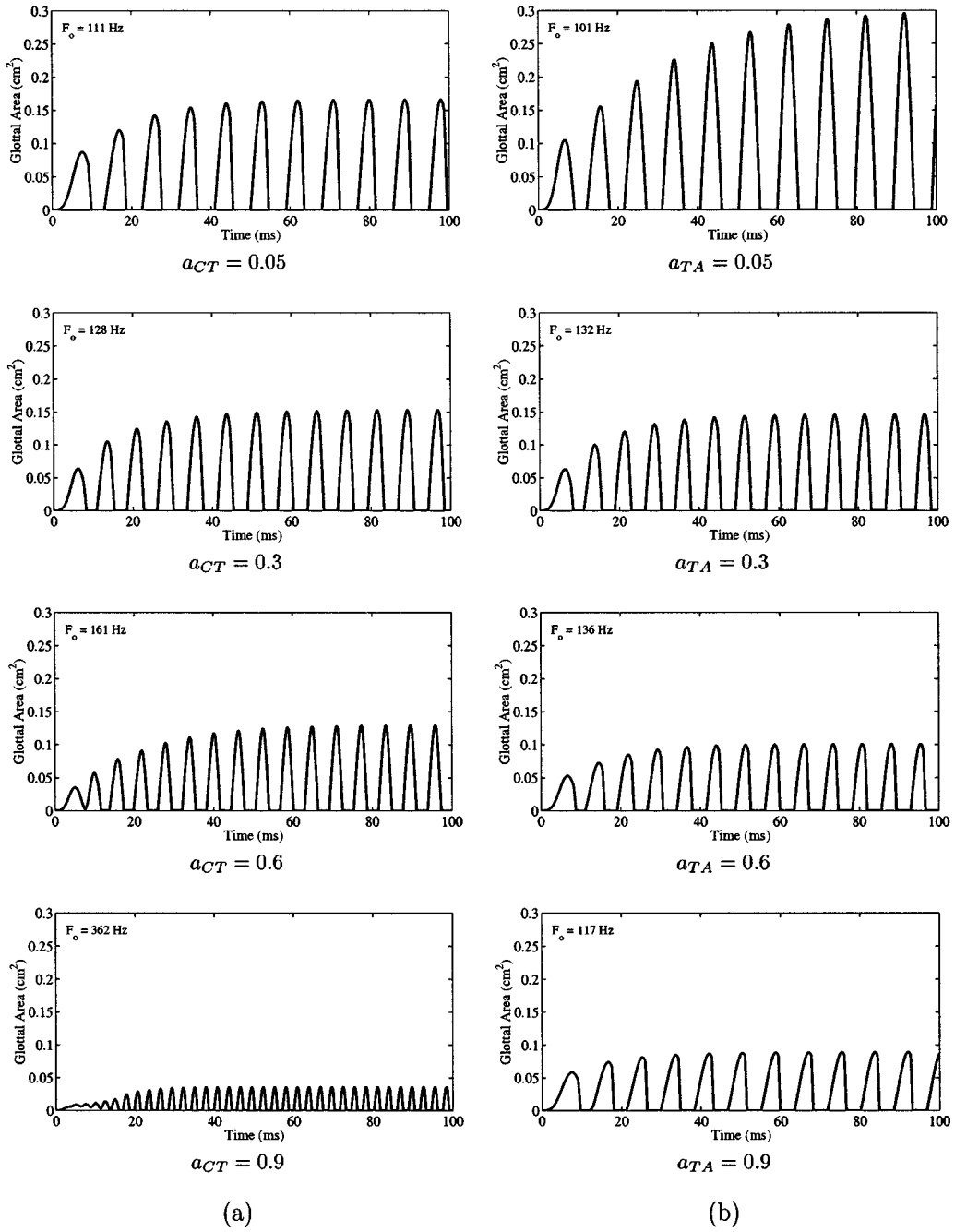


FIG. 5. Glottal area waveforms for selected sections through the tensor MAP of Fig. 4(b). (a) $a_{\mathrm{TA}}=0.25$ and (b) $a_{\mathrm{CT}}=0.25$. In both cases, a_{LC} was fixed at 0.5 and lung pressure was fixed at 0.8 kPa.

since frequency increase is always perpendicular to the contour lines. Downward sloping contour lines for F_o were shown to exist in human subjects (Titze *et al.*, 1989). For large a_{CT} [top half of Fig. 6(a)], maximum F_1 increase occurs primarily in the direction of increasing a_{CT} (upward), especially in the upper left quadrant.

The shear mode [mode 2, Fig. 6(b)] and the higher compressional mode [mode 3, Fig. 6(c)] show monotonic frequency increases with a_{CT} and monotonic decreases with a_{TA}. These modes are generally entrained to mode 1 in self-sustained oscillation as discussed previously. Note that F_2 is generally about twice F_1 for the same muscle activities, providing a likely 2:1 entrainment. Mode 3 is likely to be entrained in a 3:1 ratio to mode 1.

Figure 7 shows how the stiffnesses k_1, k_2, k_c, and K, vary with the muscle activities. Note that K in Fig. 7(d) (the body stiffness) is dominant in controlling the mode 1 frequency, given that the contour lines are most similar to those of the natural frequency of mode 1, except in the lower right corner. The cover stiffnesses k_1 and k_2 increase monotonically with a_{CT} and decrease monotonically with a_{TA}. The coupling stiffness k_c first decreases with a_{TA} and then increases again. In particular, k_c increases for $a_{\mathrm{TA}}<0.5$ and decreases for $a_{\mathrm{TA}}>0.5$. This is a consequence of the nodal rule [Eq. (58)], which gives the torsional stiffness in the cover the largest value when the nodal point is at the center.

In Fig. 8, we show the tensor MAPs once again, but this time the contour parameter is tissue mass. Figure 8(a) is for

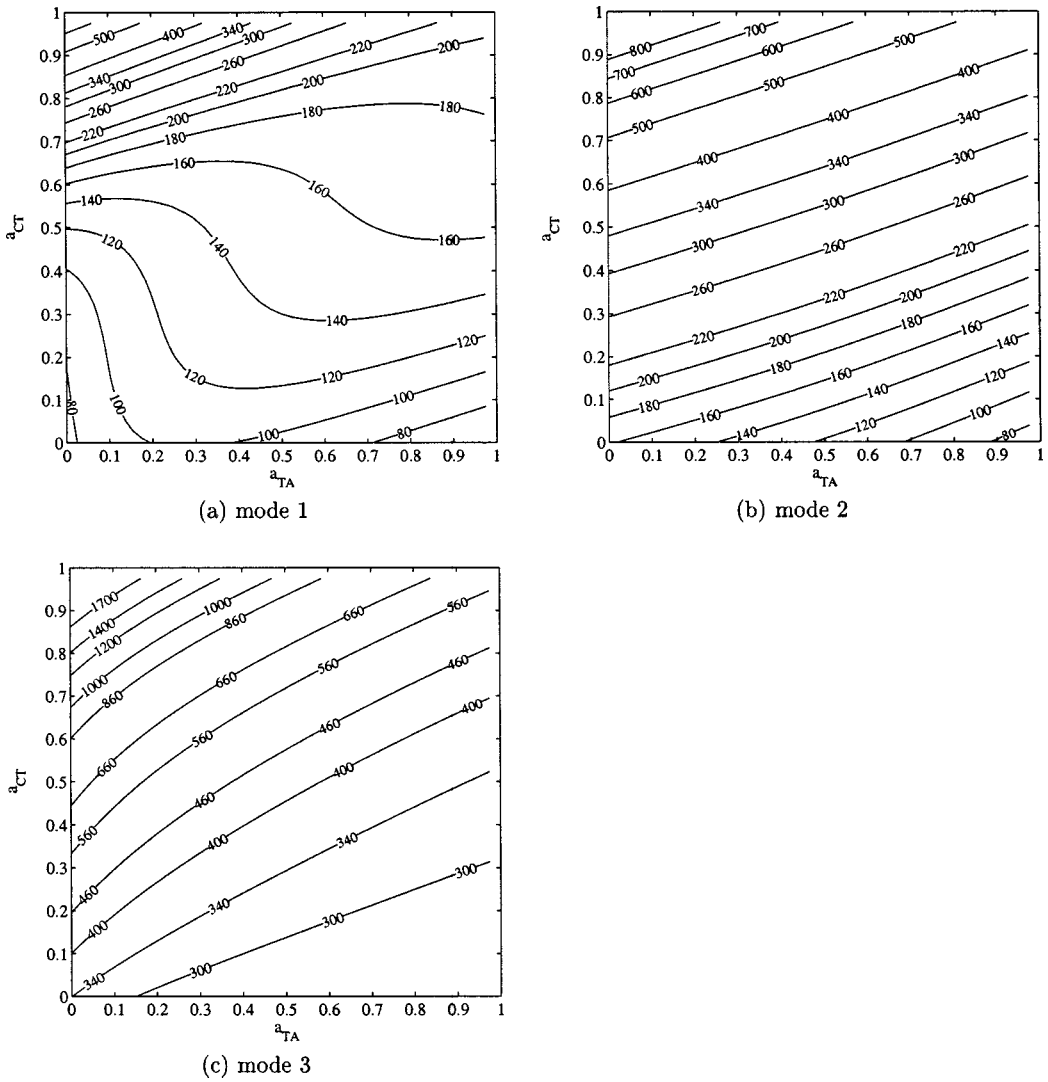


FIG. 6. Muscle activation plots (MAPs) for cricothyroid activation (a_{CT}) versus thyroarytenoid activation (a_{TA}) with normal mode frequencies as contour parameter. (a) Mode 1, the principal compressional mode, (b) mode 2, the shear mode, and (c) mode 3, the higher compressional mode. These plots are based on the rules developed in this paper.

the lower cover mass m_1, Fig. 8(b) is for the upper cover mass m_2, and Fig. 8(c) is for the body mass M. Note that all masses increase monotonically with a_{TA} and are nearly unaffected by a_{CT} as a result of our rules.

VI. CONCLUSIONS

Low-dimensional models of the vocal folds can be made to oscillate in a semi-realistic way with muscle activations, but rules are necessary to capture the covariation between parameters. In particular, the medial surface of the vocal folds, which is quantized by only two vertical compartments, must be nearly rectangular for self-sustained oscillation to occur over a wide region of muscle activities.

In order to incorporate a fibrous (stringlike) elastic restoring force as well as a nonfibrous (gellike) restoring forces, the spring constants can be constituted of multiple terms that incorporate measurable elastic moduli and fiber tensions. Both active and passive muscle properties can be incorporated. Co-contraction between the cricothyroid and thyroarytenoid muscles produce oscillation regions (and fundamental frequency contours) that are quite comparable to those that have been measured. The overall frequency range is also comparable to what is expected for humans (about 100–500 Hz). A deliberate omission in this work has been the acoustic loading of the vocal tract (both subglottal and supraglottal). It is well known that the oscillation regions of any interactive model are affected by acoustic loading, but this complexity was not warranted for this study because the focus was on vocal fold tissue characteristics.

It is conceivable that improvements could be made by letting the depths of the muscle tissue be a function of lung pressure. Such a rule is presently being investigated, but little data are available for vibration depth as a function of driving pressure. At this point, the control of depth is by intrinsic muscle activation only.

One of the most severe restrictions of low-dimensional models is the exclusion of vertical movement of the tissue. Much more variability in the medial shaping of the glottis (e.g., more convergence and divergence) is possible if the tissue is simultaneously driven upward and outward, forming elliptical trajectories. But that would double the degrees of freedom and would minimize the interpretive power that

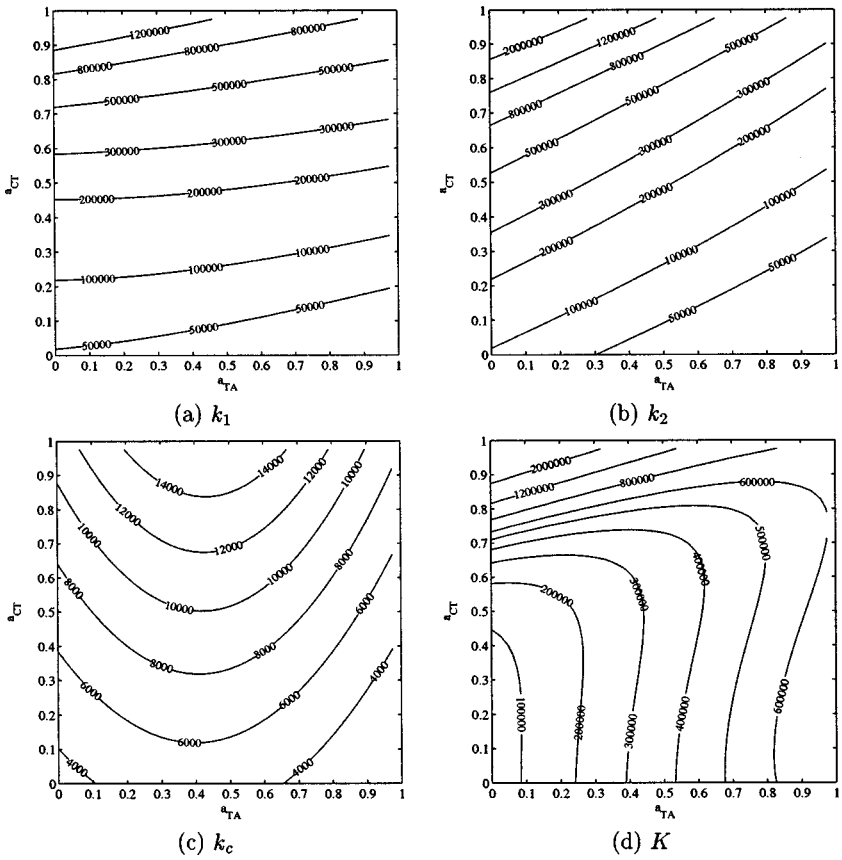


FIG. 7. Muscle activation plots (MAPs) for cricothyroid activation (a_{CT}) versus thyroarytenoid activation (a_{TA}) with the lumped-element spring constants as contour parameters. (a) Lower spring of cover, (b) upper spring of cover, (c) coupling spring of cover, and (d) spring of body.

nonlinear dynamics can offer. This paper has focused on low-dimensional vocal fold models (three degrees of freedom) and its rules for physiologic control. It cannot be overstated that pushing these rules, and the general use of low-dimensional models, much beyond this point may be counter-productive. Specifically, to investigate many aspects of vocal quality and for modeling pathology, we recommend higher-dimensional models and fewer rules.

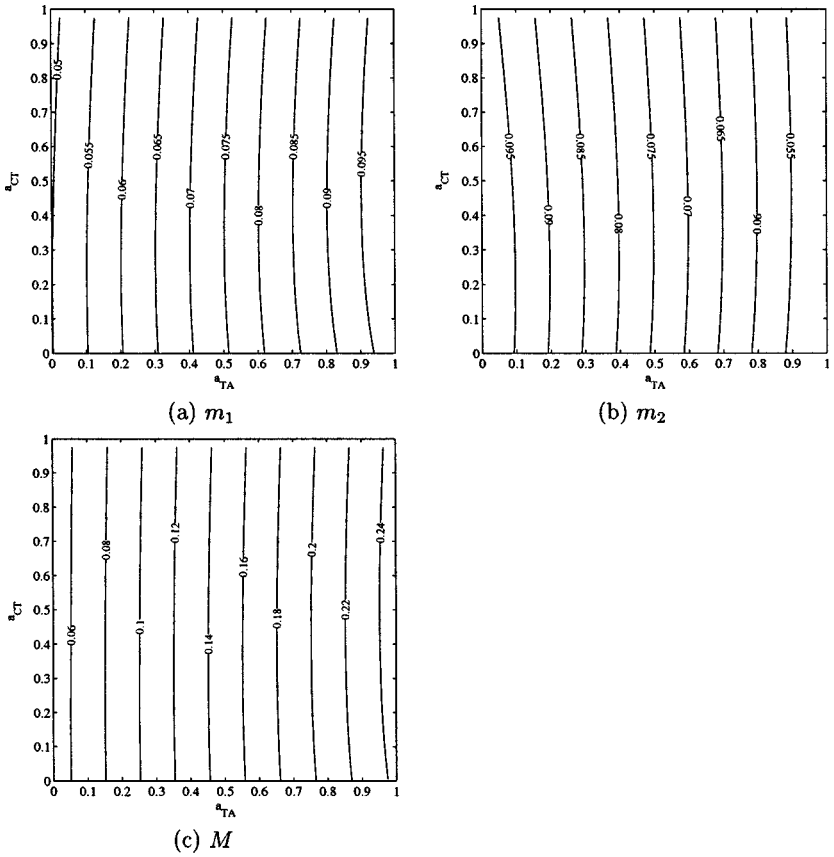


FIG. 8. Muscle activation plot (MAP) for cricothyroid activity a_{CT} versus thyroarytenoid activity a_{TA} with lumped element masses as contour parameters. (a) Lower mass of cover, (b) upper mass of cover, and (c) mass of body.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01 DC04347 from the National Institute on Deafness and Other Communication Disorders. The authors also thank Sarah Klemuk for assistance in manuscript preparation.

Alipour-Haghighi, F., and Titze, I. R. (**1983**). ''Simulation of particle trajectories of vocal fold tissue during phonation,'' in *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*, edited by I. Titze and R. Scherer, Denver Center for the Performing Arts, pp. 183–190.

Alipour-Haghighi, F., and Titze, I. R. (**1991**). ''Elastic models of vocal fold tissues,'' J. Acoust. Soc. Am. **90**, 1326–1331.

Alipour-Haghighi, F., and Titze, I. R. (**1996**). ''Combined simulation of two-dimensional air and vocal fold vibration,'' in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. Davis and N. Fletcher (Singular, San Diego), pp. 17–29.

Alipour-Haghighi, F., Titze, I. R., and Perlman, A. L. (**1989**). ''Tetanic contraction in vocal fold muscle,'' J. Speech Hear. Res. **32**, 226–231.

Baer, T. (**1975**). ''Investigation of phonation using excised larynges,'' doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Berry, D. A., and Titze, I. R. (**1996**). ''Normal modes in a continuum model of vocal fold tissues,'' J. Acoust. Soc. Am. **105**, 3345–3354.

Berry, D. A., Moon, J. B., and Kuehn, D. P. (**1999**). ''A finite element model of the soft palate,'' Cleft Palate Craniofac J. **36**(3), 217–223.

Chan, R., and Titze, I. R. (**1997**). ''Dynamic shear modulus of vocal fold tissues and phonosurgical biomaterials,'' J. Acoust. Soc. Am. **101**, 3179(A).

Chan, R., and Titze, I. R. (**1999**). ''Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results,'' J. Acoust. Soc. Am. **106**, 2008–2012.

Chan, R., Titze, I. R., and Titze, M. (**1997**). ''Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa,'' J. Acoust. Soc. Am. **101**, 3722–3727.

Flanagan, J. L., and Landgraf, L. L. (**1968**). ''Self-oscillating source for vocal-tract synthesis,'' IEEE Trans. Audio Electroacoust. **AU-16**(1), 57–64.

Fung, Y. C. (**1993**). *Biomechanics: Mechanical Properties of Living Tissue*, 2nd ed. (Springer Verlag, New York).

Hirano, M. (**1975**). ''Phonosurgery: Basic and clinical investigations,'' in 78th annual convention of the Oto-Rhino-Laryngological Society of Japan, 78th annual convention of the Oto-Rhino-Laryngological Society of Japan.

Ishizaka, K., and Flanagan, J. L. (**1972**). ''Synthesis of voiced sounds from a two-mass model of the vocal cords,'' Bell Syst. Tech. J. **51**(6), 1233–1268.

Kaneko, T., Uchida, K., Suzuki, H., Komatsu, K., Kanesaka, T., Kobayashi, N., and Naito, J. (**1981**). ''Mechanical properties of the vocal fold: Measurements *in vivo*,'' in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (Tokyo U.P., Tokyo), pp. 365–376.

Liljencrants, J. (**1991**). ''A translating and rotating mass model of the vocal folds,'' STL Quarterly Progress and Status Report, 1, Speech Transmission Laboratory (Royal Institute of Technology (KTH), Stockholm, Sweden), pp. 1–18.

Min, Y., Titze, I. R., and Alipour-Haghighi, F. (**1995**). ''Stress-strain response of the human vocal ligament,'' Ann. Otol. Rhinol. Laryngol. **104**(7), 563–569.

Nishizawa, N., Sawashima, M., and Yonemoto, K. (**1988**). ''Vocal fold length in vocal pitch change,'' in *Vocal Physiology: Voice Production, Mechanisms and Functions*, edited by O. Fujimura (Raven, New York), pp. 75–82.

Pelorson, X., Hirschberg, A., van Hassel, R., Wijnands, A., and Auregan, Y. (**1994**). ''Theoretical and experimental study of quasi-steady separation within the glottis during phonation. Application to a modified two-mass model,'' J. Acoust. Soc. Am. **96**, 3416–3431.

Scherer, R. (**1995**). ''Laryngeal function during phonation,'' in *Diagnosis and Treatment of Voice Disorders*, edited by J. Rubin, R. Sataloff, G. Korovin, and W. Gould (IGAKU-SHOIN, New York).

Story, B. H., and Titze, I. R. (**1995**). ''Voice simulation with a body cover model of the vocal folds,'' J. Acoust. Soc. Am. **97**, 1249–1260.

Titze, I. R. (**2002**). ''Regulating glottal air in phonation: application of the maximum power transfer theorem,'' J. Acoust. Soc. Am. **111**, 67–76.

Titze, I. R., and Strong, W. (**1975**). ''Normal modes in vocal cord tissues,'' J. Acoust. Soc. Am. **57**, 736–744.

Titze, I. R., and Talkin, D. T. (**1979**). ''A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation,'' J. Acoust. Soc. Am. **66**, 60–74.

Titze, I. R., Jiang, J., and Druker, D. (**1988**). ''Preliminaries to the body-cover theory of pitch control,'' J. Voice **1**(4), 314–319.

Titze, I. R., Luschei, E. S., and Hirano, M. (**1989**). ''The role of the thyroarytenoid muscle in regulation of fundamental frequency,'' J. Voice **3**(3), 213–224.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice–Hall, Englewood Cliffs, NJ).

Titze, I. R., Jiang, J. J., and Lin, E. (**1997**). ''The dynamics of length change in canine vocal folds,'' J. Voice **11**(3), 267–276.

Titze, I. R., Jiang, J. J., and Hsiao, T. Y. (**1993**). ''Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration,'' Ann. Otol. Rhinol. Laryngol. **102**(1), 58–63.

Wilhelms-Tricarico, R. (**1995**). ''Physiological modeling of speech production: Methods for modeling soft-tissue articulators,'' J. Acoust. Soc. Am. **97**, 3085–3098.

Yumoto, E., Katota, Y., and Kurokawa, H. (**1993**). ''Tracheal view of vocal fold vibration in excised canine larynxes,'' Arch. Otolaryngol. Head Neck Surg. **119**(1), 73–78.

Developing an anatomical model of the human laryngeal cartilages from magnetic resonance imaging

W. Scott Selbie
Laryngeal and Speech Section, Medical Neurology Branch, National Institute of Neurological Disorders and Stroke, 10 Center Drive, MSC 1416, Bethesda, Maryland 20892-1416

Sally L. Gewalt
Center for In Vivo Microscopy, Department of Radiology, Duke University, Durham, North Carolina 27710

Christy L. Ludlow
Laryngeal and Speech Section, Medical Neurology Branch, National Institute of Neurological Disorders and Stroke, 10 Center Drive, MSC 1416, Bethesda, Maryland 20892-1416

(Received 9 June 2000; revised 30 January 2002; accepted 9 May 2002)

The purpose of this work was to construct a three-dimensional anatomical framework of the cartilages of the human larynx. The framework included representative surface models of the four laryngeal cartilages and estimated attachment points for the intrinsic laryngeal muscles. High-resolution magnetic resonance imaging (MRI) was used to scan one female and four male human cadaveric larynges. The cartilages were segmented manually from the MRI volume for analysis. Two of these larynges were subsequently dissected and the landmark distances on the cartilages measured for comparison with the MRI measures and previous studies. The MRI measures were 8% smaller than the anatomical measures and 12% smaller than data reported in the literature. A laryngeal coordinate system was defined using the plane of symmetry of the cricoid cartilage. Measures of cricoid cartilage symmetry had less than 3% difference between the two sides for a series of measures. An algorithm for registering larynges that minimized the root-mean-square distance between the surface of a reference cricoid cartilage and the surfaces of nonisotropically scaled candidate cricoid cartilages was evaluated. This study provided an anatomical framework for registering different larynges to the same coordinate space. [DOI: 10.1121/1.1501586]

PACS numbers: 43.70.−h, 43.70.Aj, 43.70.Bk [AL]

I. INTRODUCTION

Dynamic and precise laryngeal opening and closing gestures are required for voice onset and offset during speech and airway protection for swallowing and coughing. To improve our understanding of vocal-fold movement control, several components need to be understood including the three-dimensional structure and mass of the laryngeal cartilages. A geometrical description of the laryngeal cartilages could define the orientation of the structures and locate the muscle attachments to the cartilages. Muscle kinematics will depend, in part, on the origin, insertion, and paths for the individual muscles as they wrap around the cartilages (Zajac and Gordon, 1989) in addition to their passive and mechanical properties (Alipour-Haghighi and Titze, 1985, 1991, 1999).

Several options were explored for developing a three-dimensional reconstruction of the laryngeal cartilages. Available collections of laryngeal sections lacked the fiducial markings required to realign contiguous sections. Excessive tissue deformation caused by histological processing and paraffin embedding also rendered these specimens unusable (Brizzi *et al.*, 1994; Olofsson, 1985; Glanz, 1984). Recent articles have used plastination for preserving horizontal cross sections of the human larynx (Eckel and Sittel, 1995; Eckel *et al.*, 1999). These sections display high contrasts between tissues, and can be preserved for repeated measurement and analyses (Eckel *et al.*, 1994). The horizontal sectioning process, however, requires that slices be spaced 4 mm apart, which is insufficient resolution for three-dimensional reconstructions of the cartilages given that the arytenoid cartilage is between 10 and 20 mm in most measurable dimensions (Hicks, 1981). Others have measured excised cartilages, for example: Maue and Dickson (1971), Hicks (1981), Sprinzl *et al.* (1999), and Tayama *et al.* (2001). Such measures, however, do not reference the cartilages to the same three-dimensional coordinates within the larynx. To generate data for creating three-dimensional anatomical models of the larynx, therefore, we explored use of high-resolution MRI, which is accurate to the dimensions of the tissues being scanned (Engstrom *et al.*, 1991) and preserves the three-dimensional structure.

Previous MRI studies have indicated that the laryngeal cartilages contrast adequately with the surrounding tissue to allow for segmentation of the cartilages from the other tissues (Castelijns *et al.*, 1985; Kikinis *et al.*, 1989; Sakai *et al.*, 1990). One study demonstrated that the surfaces of the laryngeal cartilages and intrinsic laryngeal muscles could be reconstructed from MRI for interactive display on a computer (Rubin *et al.*, 1998).

To assess the accuracy of the manual segmentation of laryngeal cartilages from the MRI, we compared our MRI results with direct measures of the same cartilages after dissection. We also compared our MRI measures with landmark distances reported in two studies in the literature (Hicks,

1981; Eckel *et al.*, 1994). Although many anatomical dissection studies have provided data on the laryngeal cartilages (Malinowski, 1967; Maue and Dickson, 1971; Kahane, 1978; Hicks, 1981; Kahane, 1982; Ajmani, 1990; Eckel and Sittel, 1995; Friedrich and Lichtenegger, 1997; Sprinzl *et al.*, 1999; Tayama *et al.*, 2001), the two studies selected provided measures for the same anatomical landmarks in a large number of subjects which were reproducible on MRI and were used in a similar fashion by others as standards for comparison (Tayama *et al.*, 2001).

A three-dimensional anatomical representation of the larynx could allow for the use of kinematic data from different laboratories, and integrate data collected using different visual orientations (Selbie *et al.*, 1998). Currently, a precisely defined coordinate system for the human larynx does not exist, although a tentative description was proposed by Broad (1968a) and was adapted for studying the cricoarytenoid joint (Selbie *et al.*, 1998). Laryngeal anatomy varies in the degree of lateral asymmetry (Sprinzl *et al.*, 1999). Hirano *et al.* (1989) found the cricoid cartilage was symmetric about the sagittal plane, while the thyroid cartilage was not. Therefore, we defined a coordinate system using the symmetry of the cricoid cartilage around the cricoid lumen based on the surface vertices. The resulting three-dimensional coordinate system was then used to examine the degree of symmetry within each larynx. Registration of anatomical structures to one atlas has provided an invaluable tool in the field of brain imaging (Talairach and Tournoux, 1998). This accommodates individual differences in brain size and symmetry when combining subjects to determine group results (Raichle, 1994). We explored using a similar approach to laryngeal anatomy. By registering human larynges to a reference three-dimensional anatomical space, an individual's experimental data can be compared with model predictions in the future.

The inertial properties and distribution of mass of the cartilages are needed for future dynamical simulations. These could be estimated from the segmented MRI volume data, and were used to test the assumption that three-dimensional volume elements (voxels) have a constant mass. The geometrical surface landmarks of the cartilages can be quantified and reconstructed from the segmented MRI data. Rubin *et al.* (1998) used MRI data to demonstrate how surface data for individual cartilages and intrinsic muscles could be generated for the human larynx; our purpose here is to use similar segmentations to measure and compare individual cartilages between larynges for the purpose of defining a standard laryngeal anatomical model. Others have used intrinsic laryngeal muscles dissected from the dog to determine the three-dimensional anatomic location and orientation of the intrinsic laryngeal muscles (Mineck *et al.*, 2000). Here, we used high-resolution MRI imaging in the intact human larynx to examine the three-dimensional orientation of the laryngeal cartilages and their muscle attachments.

The three-dimensional data were examined to: (1) determine the lateral asymmetry of the laryngeal cartilages; (2) define a coordinate system; (3) measure the accuracy of anatomical measures made from the MRI images against dissection data; (4) evaluate the use of an image registration system for the human larynx; and (5) specify the origin and insertion of muscles on the cartilages within the human larynx.

II. METHODS

A. MRI scanning protocol

Larynges from four male (65–75 years old), M1–M4, and one female (67 years old), F1, were excised when cadavers, previously rapidly frozen immediately after death, were unfrozen and made available by the Uniformed Services University of the Health Sciences, Department of Anatomy. Two of these larynges, M1 and M2, were used in a previous study of the cricoarytenoid joint (Selbie *et al.*, 1998). Larynges were used only if autopsy records and gross examination revealed no glottal or anatomical abnormalities. The specimens were fixed in 5% formalin, secured rigidly, and oriented axially inside acrylic cylinders. Each cylinder was inserted into a 2.5-in.-i.d. search coil and scanned with a 2.0 Tesla Omega scanner (GE-NMR Instruments, Freemont, California) using a TR of 200 ms and a TE of 6.7 ms. This scanner was selected because it was the best available to us to meet our resolution needs at that time for identifying and segmenting the major cartilages of the human larynx. A 3D Fourier transform acquisition was used to provide isotropic resolution (Suddarth and Johnson, 1991). One larynx, M1, was scanned at a voxel resolution of 256×256 pixels for 128 slices over a field of view of $90\times90\times45$ mm (yielding cubic voxels with 0.35-mm sides), while the other four larynges were scanned at 128×128 pixels for 256 slices over a field of view of $55\times55\times110$ mm (yielding cubic voxels with 0.43-mm sides).

The MRI scans were stored as isotropic three-dimensional matrices of voxel intensities and analyzed using VOXELVIEW_ULTRA (Vital Images, Inc., Fairfield, Iowa) on a Silicon Graphics R4000 Indigo computer. Cartilage, muscle, and airways were distinguished by their contrasting voxel intensities (Fig. 1).

B. Segmenting cartilages in the MRI volumes

An automatic segmentation method, the ''seed'' fill algorithm in VOXELVIEW_ULTRA, was initially evaluated. This employed a threshold value for acceptance of neighboring voxels from an initial seed voxel location. This was rejected, however, based on visual inspection of the data. Errors included: (1) the omission of low-intensity cartilage voxels that were clearly part of the cartilages, as a later dissection found no holes in the cartilages; and (2) the inclusion of voxels that were part of adjacent structures such as the trachea.

A manual segmentation approach was then used by author W.S.S. to identify and tag individual voxels and group them by structure using the GREASE PENCIL tool in VOXEL VIEW (Fig. 2). The individual voxels then remained tagged in the modified data set. During the segmentation process, the

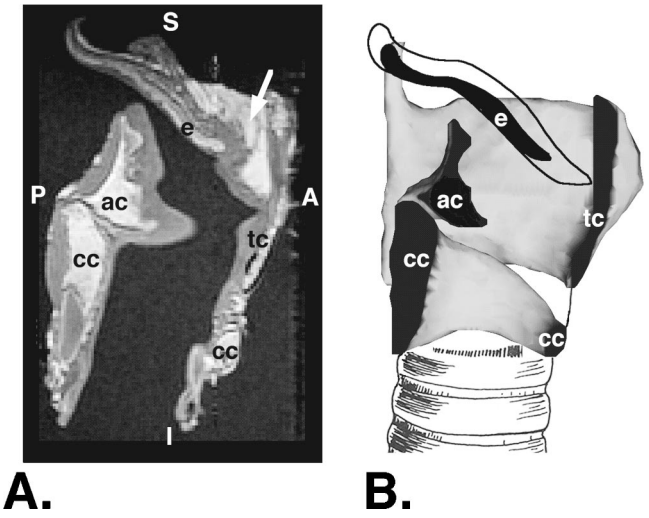


FIG. 1. (A) A sagittal MRI image, lateral to the midline, of larynx M3 (2 Tesla MRI, TR=200, TE=6.7, voxel size 0.43 mm on each side). Higher voxel intensities, in this case indicating cartilage, are colored white. The arrow indicates high-intensity voxels which are not cartilage but tissue. The intrinsic laryngeal muscles were identified as the gray voxels surrounding each cartilage. The dark speckled region is the formalin fixative. Anterior (A), posterior (P), superior (S), and inferior (I) are indicated. The cricoid cartilage (cc), arytenoid cartilage (ac), thyroid cartilage (tc), and epiglottis (e) are labeled in both A and B. (B) Surface rendering of larynx M3 cut at the plane of the image shown in (A). The parts of the structures intersecting the plane are colored black. For completeness, line drawings of the trachea, cricothyroid ligament, and projected epiglottis have been included.

data were not digitally filtered because filtering could (1) modify the size of the segmented structures and (2) reduce tissue boundary contrasts. The intensity histogram was modified interactively to enhance the contrast between tissues. Because the data sets were isotropic, and cross sections of some edges were challenging to identify, structural elements were marked by examining and editing the modified data set on each of the three series of principal planes. Voxels associated with each structure were assigned voxel intensities unique to that structure. These new voxel intensities defined our scheme for classifying each structure. These structures were then saved in data files for measurement (Fig. 3). The segmentation was an iterative process with tags modified

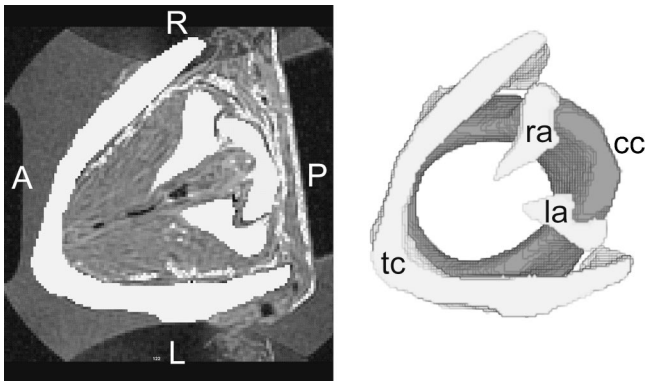


FIG. 2. (A) An axial MRI image (left) about the level of the vocal folds from one of the larynges (2 Tesla MRI, TR=200 ms, TE=6.7 ms, voxel size 0.43 mm on each side). Anterior (A), posterior (P), left (L), and right (R) are indicated. (B) The image at the right shows the cricoid cartilage (cc), right arytenoid cartilage (ra), left arytenoid cartilage (la), and thyroid cartilage (tc) segmented from the images of the whole slice and labeled.

during successive evaluations on different days. *Post hoc* evaluation was based on superimposing the original MRI data set and the segmented cartilages in the same spatial location with no interleaving. The data were then visualized by varying the opacity of each data set independently.

To estimate the reliability of the segmentation process, it was repeated for all cartilages several months later by the same individual (W.S.S.). The surface voxels of the two data sets were compared by calculating the distance from a voxel on one surface to the nearest voxel on the other surface (measured in voxels). The cricoid cartilage for specimen M1 contained the largest number of voxels of the five data sets. In this data set, less than 1% of the surface voxels from one cartilage were more than a one-voxel distance from those of the other data set. The root-mean-square error of the distance between surface voxels of the two data sets was 0.4 voxels. The cricoid cartilage for specimen F1, which contained the smallest number of voxels, had less than 6% of the surface voxels from one cartilage more than one voxel distance from the second data set. The root-mean-square error of the distance between surface voxels of the two data sets was 0.6 voxels.

The volume of each structure was determined by counting the total number of voxels. The voxels on the surface of each structure were identified after manual segmentation using an edge-detection algorithm, based on nearest neighbors, in VOXELVIEW_ULTRA. The surface area was determined by counting the total number of voxels on the outer surface of each structure.

C. Measuring the lateral symmetry of the cricoid and thyroid cartilages

For biomechanical purposes, the important details of the cricoid cartilage lie on the surface. We could not use the symmetry measures of Sprinzl *et al.* (1999) because of difficulties with identifying the joint capsules on MRI. The symmetry of the cartilages was calculated, therefore, based on a surface enveloping the vertices rather than on the individual segmented vertices. The sagittal plane of symmetry was calculated by minimizing both the volume of the convex hull enveloping voxels of a cartilage, and its reflection about the plane of symmetry. The convex hull was the closed polygonal surface, of minimum volume, that contained all of the voxels. To do this, one voxel of the cartilage near the superior aspect of the right cricoarytenoid joint (V1) was selected manually [Fig. 4(A)]. All of the voxels were transformed spatially such that the vector between the chosen voxel (V1) and another voxel (V2) was perpendicular to the sagittal plane and intersected the sagittal plane at the midpoint of the vector. A copy of the volume data set was then reflected about the sagittal plane. The volume of both the convex hull enclosing the union of the original data set, and the reflected copy of the data set [Figs. 4(C), (D)], were calculated (Watson, 1981.) The plane of greatest symmetry was defined algorithmically as the plane with the minimum convex hull of the combined data sets. This algorithm was implemented in C on a Silicon Graphics Indigo R4000. A multi dimensional downhill simplex method (Press *et al.*, 1988) was used to determine the optimal second voxel (V2).

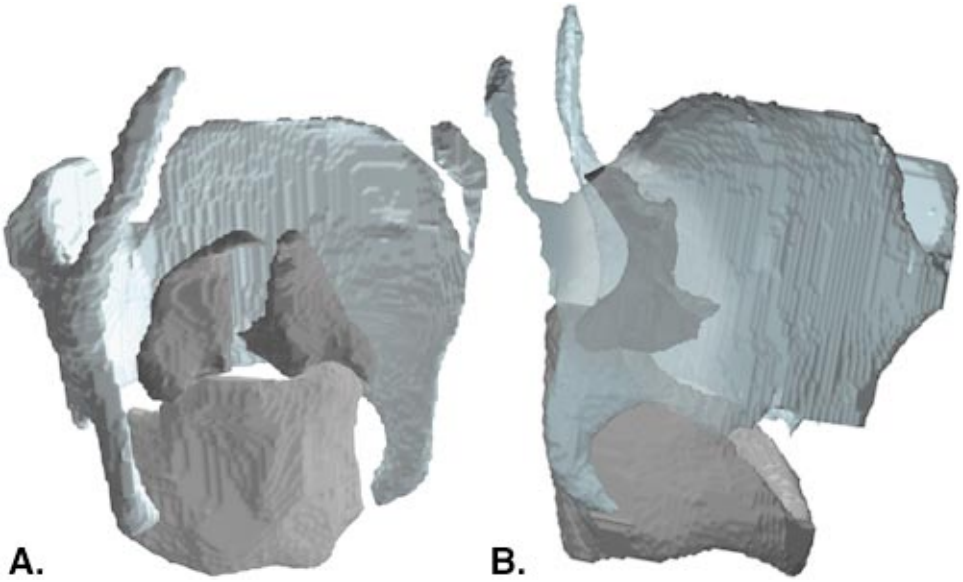


FIG. 3. (A) Volume rendering of a posterior view of the cricoid (medium shade), the thyroid cartilage (light shade), and the two arytenoid cartilages (darker shade). (B) Volume rendering of a side view of the same laryngeal cartilages extracted from one of the MRI scans.

The distribution of volume, surface area, the first moment ($\Sigma\ z$/Number of Points) and second moment ($\Sigma\ z^2$/Number of Points) (where z = the perpendicular distance from the plane of symmetry to each voxel with respect to the plane of greatest symmetry), were calculated as a test of the algorithm.

Because digitization errors inevitably occurred, we tested the sensitivity of the symmetry calculation to errors in segmentation. Based on the repeatability of segmenting the cricoid cartilage to a resolution of one voxel, the symmetry calculation was repeated on each data set with all vertices of the convex hull perturbed randomly between 0 and 1 voxel. The calculations were repeated for 50 perturbed data sets for each cricoid cartilage. If the symmetry calculation was sensitive to the location of all of the vertices, then the plane of symmetry should differ dramatically for each calculation.

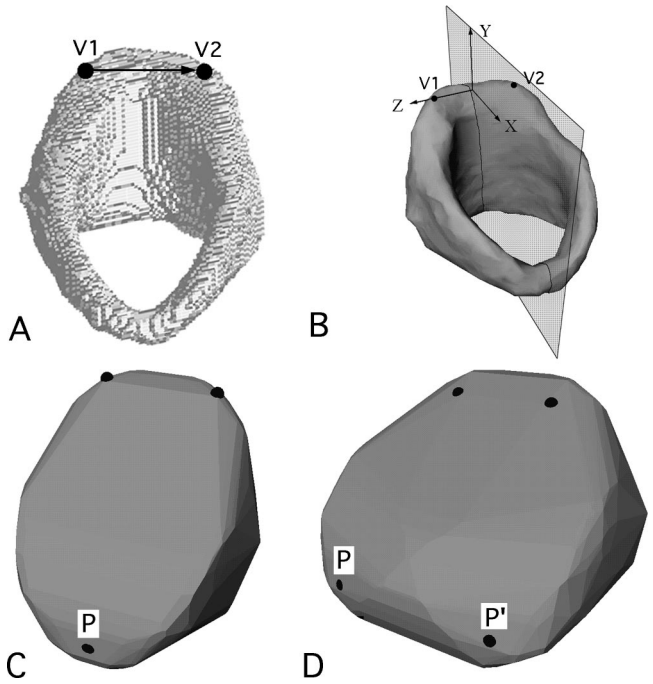


FIG. 4. (A) A volume rendering of a cricoid cartilage with vertices V1 and V2 identified. The vector from V1 to V2 is perpendicular to the axis of symmetry (S) which passes through the midpoint between V1 and V2. See methods for explanation. (B) An oblique view of a surface rendered image of the same cricoid cartilage identifying the coordinate system. The $X-Y$ plane is the plane of symmetry. (C). The convex hull of the cartilage in (A) and its reflection about the plane of symmetry for the optimal value of V2. The vertex P is unchanged by the reflection. (D) The convex hull of the cartilage in (A) and it reflection about the plane of symmetry for a poor value of V2. The point P is reflected to the point P′ for this choice of V2.

D. Defining the coordinate system

We used the same specification for a coordinate system attached to the cricoid cartilage [Fig. 4(B)] as in a previous study of the cricoarytenoid joint (Selbie *et al.*, 1998). The $(x-y)$ plane was defined by the sagittal plane of symmetry. The $(x-z)$ plane was parallel to the plane where the airway opening through the cricoid cartilage was maximum. We chose not to define this plane relative to a line through the cricothyroid joints as in Broad (1968b), because the joint was not well defined in the MRI images. A golden section search method (Press *et al.*, 1988) calculated the angle of rotation of the data set that defined the $(x-z)$ plane. The origin of the coordinate system was placed at the center of the airway opening in the $x-z$ plane and at the most superior point on the cricoid cartilage. The x axis was directed anteriorly, the y axis was directed rostrally, and the z axis was directed to the right. This coordinate system becomes closer to Broad's (1968b) when it is rotated so that the $(x-z)$ plane is perpendicular to the line of sight of a rigid 90° laryngoscope (Selbie *et al.*, 1998) and can be used to compare the model predictions with laryngoscopic data. Because the line of sight is based on the position and optical angle of the lens of a laryngoscope, it provides an anatomical coordinate that can be used in future clinical studies.

E. Orientation of the thyroid relative to the cricoid

Two sets of symmetry measures were calculated for the thyroid cartilages. First, assuming that the larynx as a whole

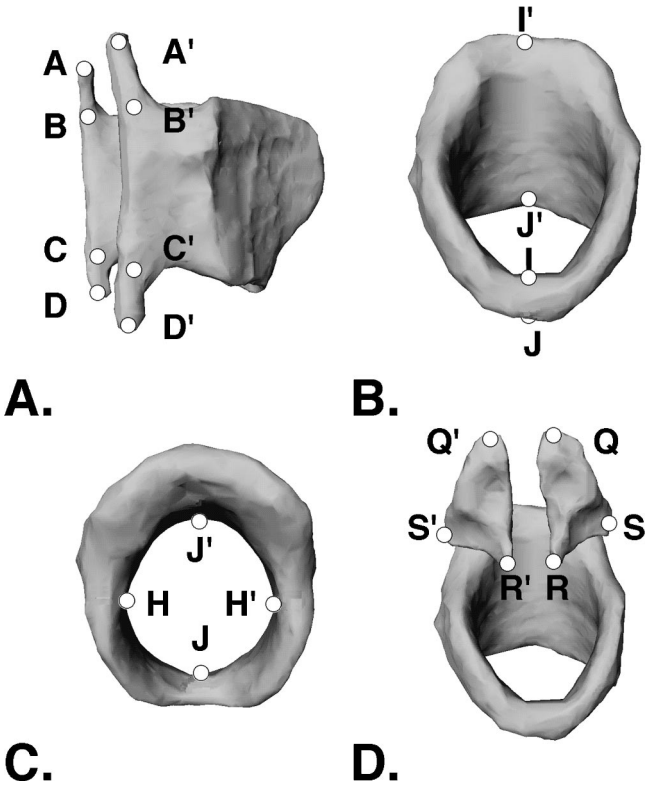


FIG. 5. (A)–(D) Surface renderings of cartilage M3 are displayed to illustrate the locations of the landmark measurement points (Maue and Dickson, 1971; Hicks, 1981) used for both the MRI and direct measures in this investigation. Corresponding points on each side are designated by the point and a prime (e.g., A and A′). (A) Illustrates the superior–inferior measures of the thyroid cartilages. (B) Illustrates the superior–inferior measures of the cricoid cartilage. (C) Illustrates the internal anterior to posterior and lateral medial measures of the cricoid. (D) Illustrates the three measurement points of one arytenoid cartilage on each side for the superior tip (Q), the muscular process (S), and the vocal process (R).

is symmetrical, the symmetry was calculated with respect to the plane of greatest symmetry of the cricoid cartilage. Second, using the same value of V1 as defined earlier on the cricoid, V2 was selected using the same algorithm to determine the plane of greatest symmetry for the thyroid cartilage.

F. Anatomical measurements

Two male larynges (M1 and M2) were removed from the cylinders; the cartilages were dissected and then measured using a protocol similar to those reported in the literature (Maue and Dickson, 1971; Kahane, 1978; Hicks, 1981; Kahane, 1982; Eckel *et al.*, 1994) (Fig. 5). These two larynges were selected because their MR images showed the best contrast between cartilage and other tissues. All measures of these two sets of laryngeal cartilages were made from structures free of contiguous soft tissue. Previously described linear measures (Kahane, 1982) were made with a vernier caliper accurate to 0.05 mm. Weight measurements were made with a Sartorius balance accurate to the nearest 0.01 gm.

We initially used VOXELVIEW to identify voxels within the image slices in the MR volume homologous to the anatomical landmark points defined in the literature (Maue and Dickson, 1971; Kahane, 1978; Hicks, 1981; Kahane, 1982; Eckel *et al.*, 1994). Because the orientations of the specimens differed to some degree, it was challenging to choose homologous voxels consistently. The landmarks were located by displaying the estimated locations on the surface models. Using software that allowed easy manipulation of the object in space made it possible to locate homologous landmarks in the images and on the anatomical specimens. These models were sufficiently detailed to allow the precise placement of the landmarks.

Several landmark-based distances translated readily into measures of the volume data, while others could not translate to the MRI voxel measurement. For example, measures of height and width were made after transforming the voxels into the coordinate system of Sec. II D.

G. Registration

For future tests of biomechanical hypotheses, it will be important to scale an individual larynx to a reference framework for testing model predictions. This would allow for comparisons between model simulations and vocal-fold movements recorded from videolaryngoscopy. In our first attempt to register the cartilages, scaling factors were calculated for the three principal planes of the cartilages and registered to our local coordinate system (Sec. II D). We calculated the registration of one cricoid cartilage to a reference cricoid cartilage. We then scaled the thyroid and arytenoid cartilages using the scaling factors calculated for the cricoid cartilage, to register the entire larynx. The appropriateness of the scaling for all of the laryngeal cartilages using scaling factors of the cricoid cartilage was examined by comparing the anatomical measures before and after scaling (Sec. II F).

An optimization strategy was implemented that minimized the rms distance between the surface voxels of one cricoid cartilage along the vertex normal and the intersection of the nearest polygonal surface of the reference cricoid cartilage. We used MATLAB (Mathworks, Inc.) to calculate an isosurface for each extracted cartilage. These isosurfaces were calculated for the full data set. Because of the large numbers of polygon counts created by the isosurface procedure, we used the MATLAB polygon reduction routines to reduce the number of surface polygons.

H. Specifying the origin and insertion of the intrinsic laryngeal muscles

We used M1, the MRI specimen that had the best contrast and was used previously in a study of the cricoarytenoid joint properties (Selbie *et al.*, 1998), to estimate a single voxel as the point of attachment for each muscle from the MRI volume. Although the region of attachment is certainly greater than a single voxel, a single point of attachment can be used as one component for biomechanical modeling purposes (Zajac and Gordon, 1989). This was based on reviewing the major fascicle direction for each muscle in the MRI image, estimating the point of attachment on the cartilages, and examining the dissected cartilages from M1 to visually confirm the decisions (Fig. 6). It must be realized, however, that muscle force vectors cannot be estimated by

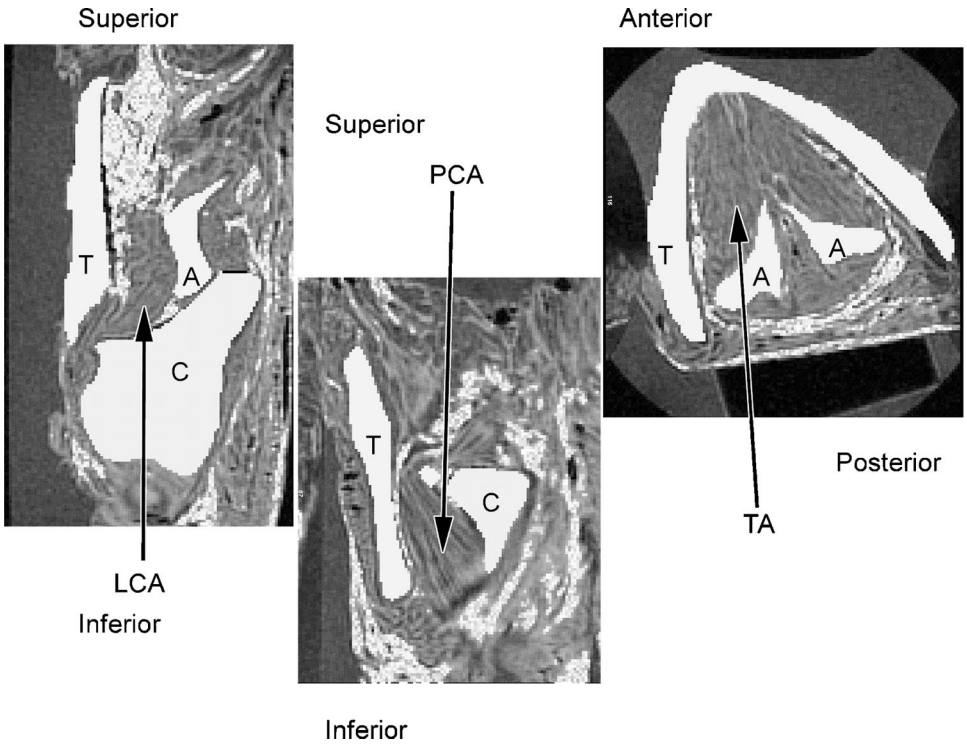


FIG. 6. Three MRI sections demonstrating the line of action of the intrinsic muscles. Note that the line of action is parallel to the fascicle direction of the muscles (light streaks in the MRI image.) TA = thyroarytenoid muscle, PCA = posterior cricoarytenoid muscle, LCA = lateral cricoarytenoid muscle, Sup = superior, Inf = inferior, A = arytenoid cartilage, T = thyroid cartilage, C = cricoid cartilage.

TABLE I. Comparison of the anatomical measurements, as described by Maue and Dickson, between the two sets of dissected cartilages (DIS), M1 and M2, and the same landmarks located in the MRI images (all measurements in mm) that were identified in Fig. 5.

	M1				M2			
	MRI	DIS	DIS/MRI	% diff	MRI	DIS	DIS/MRI	% diff
Cricoid								
IJ	6.04	6.29	1.04	4.25	6.55	7.92	1.21	20.85
I′J′	22.95	22.33	0.97	−2.71	23.05	23.90	1.04	3.69
HH′	18.52	19.48	1.05	5.21	21.07	21.44	1.02	1.74
JJ′	22.89	24.00	1.05	4.85	19.50	23.54	1.21	20.72
Arytenoid								
QR	16.58	15.78	0.95	−4.83	18.15	18.27	1.01	0.66
Q′R′	13.03	16.26	1.25	24.82	16.66	18.04	1.08	8.28
QS	15.04	15.81	1.05	5.08	17.29	17.73	1.03	2.56
Q′S′	15.07	15.53	1.03	3.05	15.59	18.04	1.16	15.72
RS	14.70	14.51	0.99	−1.31	14.22	15.53	1.09	9.17
R′S′	14.07	13.60	0.97	−3.36	13.56	15.42	1.14	13.72
Thyroid								
AD	47.62	47.57	1.00	−0.11	50.91	51.17	1.01	0.51
A′D′	48.39	49.22	1.02	1.71	43.35	48.03	1.11	10.79
AB	14.38	16.41	1.14	14.11	18.52	19.15	1.03	3.40
A′B′	18.00	18.72	1.04	4.00	14.07	16.49	1.17	17.24
CD	11.40	8.42	0.74	−26.12	12.65	12.89	1.02	1.91
C′D′	10.97	8.47	0.77	−22.75	10.89	11.95	1.10	9.76
AA′	21.78	22.88	1.05	5.09	39.81	41.63	1.05	4.58
BB′	33.59	35.39	1.05	5.37	45.02	45.54	1.01	1.16
CC′	29.81	31.33	1.05	5.11	37.84	41.18	1.09	8.81
DD′	27.51	27.75	1.01	0.91	30.18	30.52	1.01	1.14
W	5.28	5.89	1.11	11.47	5.18	5.07	0.98	−2.01
W′	5.04	5.96	1.18	18.30	5.18	4.79	0.93	−7.40
Mean				7.93				7.54

TABLE II. Comparison of the anatomical measurements of cartilages extracted from all five of the MRI images and the values reported by Hicks (1981) and from Eckel *et al.* (1994) (all measurements in mm) using the landmarks identified in Fig. 5. The mean values for the MRI are for the four male larynges followed by Hicks' values for males, the female data followed by Hicks' values for females and the Eckel *et al.* (1994) values for males. The numbers in parentheses are the numbers used by Eckel *et al.* (1994) to denote the same locations for measurement as were used by Hicks (1981) and in this study.

	M1	M2	M3	M4	MRI		Hicks $N=21$		% diff	F1	Hicks $N=10$		Eckel *et al.* $N=28$	
Cricoid					Mean	s.d.	Mean	s.d.			Mean	s.d.	Mean	s.d.
IJ	6.04	6.55	4.10	5.11	5.45	1.08	6.15	1.04	12.83	5.18	5.69	0.99		
I′J′	22.95	23.05	19.82	24.95	22.69	2.12	23.62	1.50	4.10	18.91	19.36	1.02		
HH′ (60)	18.52	21.07	16.75	21.24	19.39	2.16	24.64	2.11	27.04	17.29	19.66	1.91	18.2	2.18
JJ′ (56)	22.89	19.50	21.04	27.57	22.75	3.50	23.67	2.03	4.06	15.12	18.87	0.81	20.1	2.30
Arytenoid														
QR (77)	16.58	18.15	11.00	18.21	15.98	3.41	17.73	1.45	10.91	9.37	13.16	1.02	16.2	2.57
Q′R′ (76)	13.03	16.66	11.26	17.84	14.70	3.07	17.58	1.50	19.59	9.37	12.68	0.97	17.0	3.67
QS (75)	15.04	17.29	13.47	17.22	15.76	1.85	17.45	1.14	10.74	11.98	14.25	0.76	16.8	2.75
Q′S′ (74)	15.07	15.59	13.52	15.85	15.01	1.04	17.65	1.37	17.64	10.62	14.05	0.99	17.1	2.44
RS (79)	14.70	14.22	9.65	14.45	13.26	2.41	14.10	0.64	6.32	11.63	10.82	0.71	13.3	1.45
R′S′ (78)	14.07	13.56	10.54	13.95	13.03	1.67	13.89	0.89	6.61	10.40	10.74	0.81	13.5	1.89
Thyroid														
AD (15)	47.62	50.91	44.87	43.75	46.79	3.19	45.21	5.49	−3.36	36.05	33.76	3.43	45.1	4.18
A′D′ (14)	48.39	43.35	43.74	39.57	43.76	3.61	44.32	5.08	1.28	40.17	34.32	2.59	43.9	4.63
AB (25)	14.38	18.52	14.13	9.67	14.18	3.62	16.36	3.99	15.37	8.96	12.50	1.58	14.5	3.04
A′B′ (24)	18.00	14.07	13.71	11.02	14.20	2.88	16.97	3.66	19.48	12.12	13.59	1.65	14.9	2.94
CD (27)	11.40	12.65	5.86	6.91	9.20	3.32	7.06	1.27	−23.29	8.77	6.07	1.22	8.9	1.13
C′D′ (26)	10.97	10.89	6.56	7.34	8.94	2.32	6.48	1.48	−27.54	7.89	6.15	1.17	8.3	1.27
AA′	21.78	39.81	25.52	27.47	28.64	7.81	35.84	7.09	25.13	43.63	42.19	5.69		
BB′ (42)	33.59	45.02	31.04	35.64	36.32	6.10	38.71	3.66	6.58	45.13	36.50	2.24	42.6	6.08
CC′	29.81	37.84	26.96	33.63	32.06	4.73	31.60	2.36	−1.44	32.42	26.59	1.91		
DD′	27.51	30.18	24.63	33.09	28.85	3.62	33.81	2.97	17.18	29.32	27.92	2.44		
W	5.28	5.18	7.05	7.75	6.32	1.29	6.40	5.79	1.35	6.72	4.39	0.74		
W′	5.04	5.18	6.11	7.75	6.02	1.25	6.66	5.56	10.54	6.39	4.57	0.64		
Mean									12.38					

straight lines because muscle wrapping occurs around and between cartilages such as over the surface of the cricoid cartilage, between the cricoid and thyroid, and around the muscular process. Greater resolution is needed than was available here to establish the changes in fascicle direction as muscle fibers are wrapped around the cartilages. Therefore, our purpose was only to estimate the location of the muscles as they insert on the cartilages but not their force vectors for biomechanical modeling purposes.

III. RESULTS

A. Anatomical measurements

The signal intensities of the cartilages were higher than for muscle (Fig. 1). While the contrast between structures varied among specimens, most cartilaginous tissues were distinguishable. Voxels located at landmarks (Fig. 5) were tagged in the MRI volume for comparison with the anatomical measures made from the two dissected cartilages (Table I). For both larynges, 16 out of the 22 MRI measures lay within 10% of the anatomical measures (average difference: M1 = 7.93% and M2 = 7.54%). The majority of the MRI measures were smaller than the corresponding anatomical measures to a similar degree for both cartilages (M1 16 out of 22; M2 20 out of 22).

Table II contains the measures from the MRI scans of the five larynges along with the data reported by Hicks (1981) and Eckel *et al.* (1994). Mean values were calculated for the MRI scans of the four male larynges. The majority of the measures from the MRI data set were smaller than the corresponding anatomical measures (20 out of 23). The mean percentage error for the MRI data set compared with the Hicks' (1981) data set was 12.4%, with 13 of the 23 measures being less than the mean of Hicks' measures.

B. Segmentation

The major challenge during segmentation was identifying the articular surfaces of the cricoarytenoid and the cricothyroid joints because the joint gap was about one voxel and formalin may have distorted the joint capsule. Another challenge was segmenting the arytenoid cartilage from the vocal process and the vocal ligament. This was also difficult to distinguish during anatomical dissection.

The weight, volume, and estimated density of the two dissected cartilages were examined (Table III). The density was more similar within a larynx than between larynges: larynx M1 (1.11; 0.97; 0.87; 1.29 g/cm^3) and larynx M2 (1.46; 1.63; 1.63; 1.73 g/cm^3).

The volume and surface areas measured from the MRI in all five larynges are also provided in Table III. The volumes of the male cartilages spanned a wide range of sizes; cricoid (3.04–5.87 cm^3), arytenoids (0.18–0.44 cm^3), and thyroid (6.13–8.81 cm^3). The volumes of the female cartilage were smaller than all of the corresponding male cartilages except for the arytenoid cartilages of M3 (Table III).

TABLE III. Comparing the volume and surface area of the cartilages calculated from the number and size of the voxels comprising the structure. The weights were measured only for the dissected larynges. For comparison the data of Hicks (1981) has been included.

	M1	M2	M3	M4	F1	Hicks ($N=21$)	
Cricoid						mean	s.d.
Volume (cm^3)	4.22	4.39	3.04	5.87	2.38		
Surface area (cm^2)	20.4	19.8	13.8	25.3	12.0		
Weight (g)	4.67	6.40	⋯	⋯	⋯	5.26	0.73
Density (g/cm^3)	1.11	1.46	⋯	⋯	⋯		
Arytenoid (L)							
Volume (cm^3)	0.44	0.37	0.20	0.42	0.21		
Surface area (cm^2)	3.36	3.0	1.84	3.23	1.91		
Weight (g)	0.43	0.60	⋯	⋯	⋯	0.39	0.05
Density (g/cm^3)	0.97	1.63	⋯	⋯	⋯		
Arytenoid (R)							
Volume (cm^3)	0.42	0.39	0.18	0.43	0.22		
Surface area (cm^2)	3.31	3.0	1.71	3.26	1.95		
Weight (g)	0.36	0.63	⋯	⋯	⋯		
Density (g/cm^3)	0.87	1.63	⋯	⋯	⋯	0.39	0.05
Thyroid							
Volume (cm^3)	7.54	6.25	6.13	8.81	5.23		
Surface area (cm^2)	47.96	42.54	34.53	45.42	27.69		
Weight (g)	9.73	10.92	⋯	⋯	⋯	8.52	1.37
Density (g/cm^3)	1.29	1.73	⋯	⋯	⋯		

M3 was the smallest male larynx and because of poor contrast between tissues in the MRI, it was more difficult to segment the arytenoid cartilages.

C. Measured symmetry of the cricoid and thyroid cartilages

Table IV contains the four measures of symmetry. The difference in the number of right-to-left side voxels was normalized to the total number of voxels by computing percent difference as $[(R-L)/(R+L)]*100$. This measure allowed us to estimate the percentage of volume, surface area, first moment, and second moment that would vary due to side-to-side asymmetry. The measures of each cricoid cartilage were made with respect to the plane of symmetry calculated for that cricoid cartilage. The volume on the right was greater than on the left in four of the five larynges (range of −1.83% to 2.21%). On the other hand, the surface area on the left was greater than the right in four of the five larynges (range of −1.18% to 2.75%). The first moment on the right was greater than on the left in two of the five larynges (range of −1.89% to 0.31%), and the second moment on the right was greater than on the left in two of the five larynges (range of −2.79% to 0.18%). These asymmetries were not consistent and did not demonstrate that either side tended to be greater than the other.

Perturbing all of the vertices of the convex hull tested the sensitivity of the symmetry calculation for the cricoid cartilage by random values between 0 and 1 voxel in any direction. The difference in the number of right-to-left side voxels for all cartilages had a standard deviation of less than 0.52% (M1 = 0.34%; M2 = 0.33%; M3 = 0.44%; M4 = 0.49%; F1 = 0.52%) with a total range less than 2.44%

TABLE IV. Percentage differences (e.g., $[(R-L)/(R+L)]*100$) about the plane of symmetry of the cricoid cartilage (columns 2 and 3) and thyroid cartilage (columns 4 and 5) in five larynges examined, including those between the volume (V), surface area (SA), first moment (FM), and second moment (SM) of the right side of the cartilage to the left side.

	Using cricoid coordinate system				Using thyroid coordinate system			
	V	SA	FM	SM	V	SA	FM	SM
Cricoid								
M1	1.55	−0.61	−0.02	−0.22				
M2	−1.83	−1.18	−1.89	−2.79				
M3	0.49	−0.12	0.22	0.18				
M4	1.29	−1.03	−1.09	−1.98				
F1	2.21	2.75	0.31	0.17				
Thyroid								
M1	−12.34	−10.87	−8.37	−14.92	−5.97	−3.98	−2.06	−1.85
M2	−5.03	1.28	−7.68	−15.06	5.01	−1.36	−2.48	−3.46
M3	1.66	−0.42	5.08	11.09	2.06	0.20	−0.59	−1.13
M4	−4.63	−1.95	7.16	13.91	−3.86	−0.78	−1.23	−2.81
F1	−0.71	−1.96	0.80	1.90	0.25	−0.48	1.21	1.79

TABLE V. The scaling parameters and rms error of the distance along the normal vector of each vertex on the cricoid cartilage for M1 to the intersection of a polygon on the scaled cricoid cartilages for M2, M3, M4, and F1.

	x-scale	y-scale	z-scale	rms error (mm)
M2	1.110 891 27	1.019 124 70	0.879 730 68	0.0167
M3	1.191 917 88	1.390 924 95	1.089 294 62	0.0124
M4	0.954 624 63	0.983 394 04	0.869 810 29	0.0084
F1	1.438 527 90	1.398 986 32	1.128 316 01	0.0128

(M1 = 1.43%; M2 = 1.48%; M3 = 2.13%; M4 = 2.44%; F1 = 2.35%). The difference in the surface area of the right side to the left side for all cartilages had a standard deviation less than 0.48% (M1 = 0.24%; M2 = 0.28%; M3 = 0.26%; M4 = 0.43%; F1 = 0.48%), with a total range less than 2.61% (M1 = 1.07%; M2 = 1.27%; M3 = 1.40%; M4 = 2.03%; F1 = 2.61%).

Measures of thyroid cartilage symmetry were then made using the laryngeal coordinate system defined by the symmetry of the cricoid cartilage (Table IV). The volume of the left was greater than the right in four of the five larynges (range of −12.34% to +5.03%). The surface area on the left was greater than on the right in four of the five larynges (range of −10.87% to +1.28%). The first moment on the right was greater than on the left side in three of the five larynges (range of −8.37% to +7.16%), and the second moment on the right was greater than on the left side in three of the five larynges (range of −15.06% to +13.91%).

Symmetry measures for the thyroid cartilage were then compared when based on the symmetry of the thyroid cartilage (Table IV). For the resulting plane of symmetry of the thyroid cartilage, the volume on the right was greater than on the left in three out of five larynges (range of −5.97% to +5.01%). The surface area on the left side was greater than on the right in four of the five larynges (range of −3.98% to +0.20%). The first moment on the right side was greater for the left side in one of the five larynges (range of −2.48% to +1.21%). The second moment on the left was greater than on the right in four of the five larynges (range of −3.46% to +1.79%). Because the directions of asymmetries differed for particular larynges between the different measures of mass, volume, and density, there was no clear tendency for the right or left side to systematically be greater.

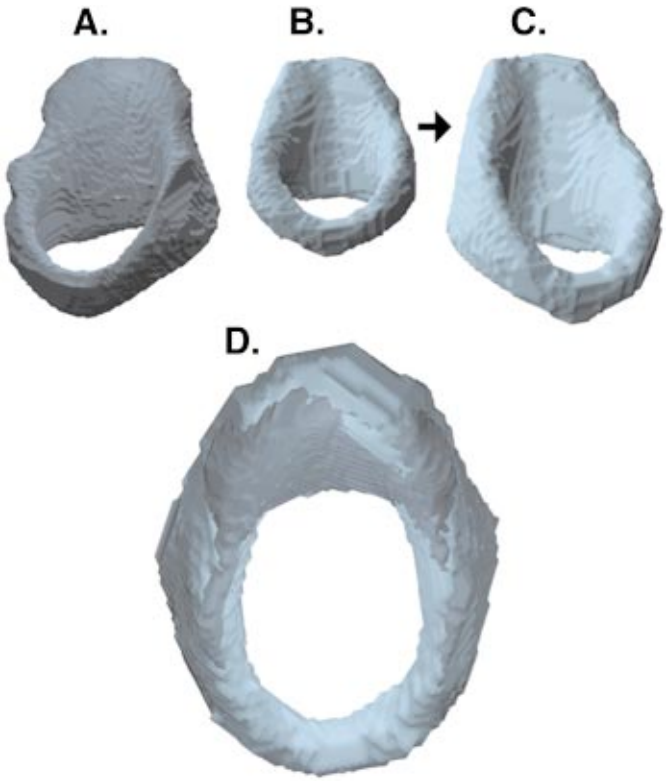


FIG. 7. (A) The male cartilage (M1 in gray) (B) The female cartilage (F1 in white) in the same coordinate system. (C) The female cartilage after being scaled. (D) The M1 and F1 cartilages combined after scaling F1.

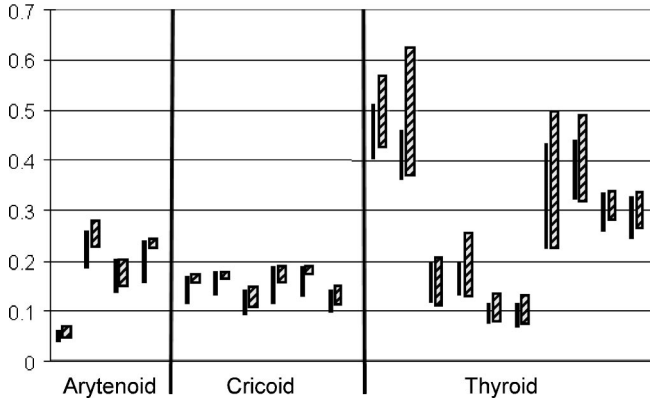


FIG. 8. The distance between landmarks are displayed as a bar graph where the size of the graph indicates the range of values for the five cartilages. The hatched wide line represents the scaled cartilage values and the dark thin line the unscaled cartilage values.

We elected to define the coordinate system with respect to the cricoid cartilage instead of the thyroid cartilage, because the cricoid was usually symmetrical. The differences we observed in measuring the symmetry of the thyroid cartilage relative to the cricoid plane of symmetry may reflect a tilting of the thyroid cartilage relative to the cricoid cartilage.

D. Registration

The cricoid cartilages, M2, M3, M4, and M5, registered accurately to cartilage M1 (Table V). The rms error of the distance along the normal vector of each vertex of M1 to the intersecting polygon on the comparison cartilage was less the 0.02 mm for all four cartilages. This is illustrated in Fig. 7, taking F1, the larynx with the cricoid cartilage most different from M1, and showing that with scaling the two surfaces correspond.

Registration was performed only on the cricoid cartilage. To test the effect of the registration on the other cartilages, we calculated the anatomical distances as defined by Kahane *et al.* (1978, 1982) and then recalculated these distances after scaling. Figure 8 displays a bar graph showing the range of values for all five larynges for each measured distance before and after scaling. Note that after scaling (the hatched bars) the ranges were reduced for the cricoid and arytenoid cartilages but not for the thyroid cartilages. This demonstrated that the size of the thyroid cartilage was not linked directly to the size of the cricoid cartilage.

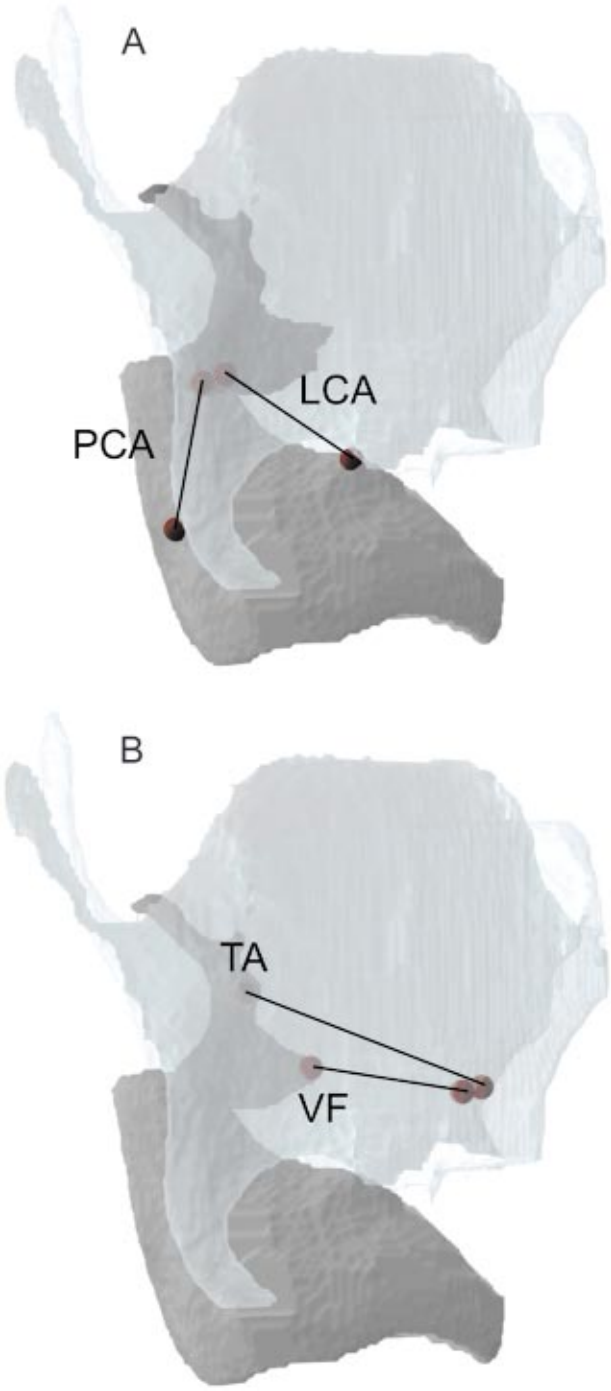


FIG. 9. The attachment points and the line of action of the intrinsic laryngeal muscles. TA = thyroarytenoid muscle, VF = line of the vocalis portion of the thyroarytenoid muscle, PCA = posterior cricoarytenoid muscle, LCA = lateral cricoarytenoid muscle.

E. Location of the origin and insertion of the intrinsic muscles

Figure 9 displays the position of the intrinsic laryngeal muscles on a model of the cricoid cartilage. Unfortunately, the MRI technique only allowed us to identify the fascicle direction of one of the two compartments of the posterior cricoarytenoid muscle. We were, however, able to identify two different fascicle directions for the thyroarytenoid muscle; the lateral and the more medial compartments were seen. The medial portion was along the direction of the vocal fold.

IV. DISCUSSION

Our rationale for developing a three-dimensional anatomical framework of the larynx is the assumption that the functional characteristics of the larynx may be defined, in part, by geometry. If there are invariant anatomical relationships between cartilages, this could be of importance in future predictions of the behavior of the larynx. Detailed anatomical measures from laryngeal cartilages were previously reported using surface landmarks (Maue and Dickson, 1971; Kahane, 1978; Malinowski, 1967; Hicks, 1981). Such measures, however, do not relate the form within an individual larynx nor define the relationships between larynges. Other measures, such as those used by Lele (1991), calculated a set of matrices based on the distance between homologous landmarks on the surface. These do not describe the surface geometry in a form suitable for registration because of limitations in finding homologous landmarks on the cartilages. For this reason, we examined the relationships between asymmetries in the cricoid and thyroid cartilages, using each as a reference for the other. Registration to a reference structure will also allow the integration of anatomical data between individuals.

With future improvements in MRI resolution, similar studies may eventually become realistic in living persons. Several technological obstacles must first be addressed because of the sensitivity of MRI scanning to movement artifact. In particular, scanning will need to be closely synchronized within the respiratory cycle because the laryngeal movement during breathing and swallowing epochs must be extracted.

A. Scanning protocol

The comparison between measures made from the MRI and the dissection data confirmed the accuracy of the MRI data. Comparisons with previous values in the literature demonstrated that the cartilages were of normal size and proportion and suitable for the developing an anatomical framework.

Our results confirmed the feasibility of quantifying laryngeal cartilage geometry using MRI. With respect to the MRI scanning protocol, a short spin echo time (TE = 6.7 ms) with a T1 weighted scan provided good T2 contrast for muscles and cartilages. Although this was not a quantitative regime for either T1 or T2 measurements, it provided adequate contrast between tissues for identifying the cartilages. The spatial resolution of the images was adequate; even the smallest cartilage, one of the arytenoids, comprised 2283 voxels. Therefore, the MRI scanning protocol was adequate for the purpose of this study. Such a scanning protocol would not be adequate for other purposes, however, such as quantifying the articular facets of the cricoarytenoid joint which was done effectively using a milling machine (Selbie *et al.*, 1998).

B. Anatomical measurements

The MRI measures were compared with anatomic measures reported in the literature. Previously reported measurement protocols were difficult to replicate by locating the

landmark voxels using VOXELVIEW_ULTRA on the MRI data set. Tagged voxels, therefore, corresponding to previous anatomical landmarks were compared between larynges for consistency. For example, the majority of these MRI measurements were within 10% of the dissected data. Two factors contributed errors: first, difficulties discriminating the landmarks which were at extreme locations on a cartilage in the MRI volume, and second, difficulties determining the edge of a structure because of the fuzzy boundaries between tissues. An error of one voxel in each of the two landmarks could account for the majority of the observed differences in length (20 of the 23 measures for M1; 16 of the 23 measures for M2). Other differences were attributed to difficulties in distinguishing locations such as the vocal process from the vocal ligament, the extent of the thyroid cornua, the juxtaposition between the arytenoid and cuneiform cartilages, and the inferior edge of the cricoid cartilage from the trachea. These differences cannot be attributed to scaling problems in the MRI images, however, because MRI preserves the anatomical scale (Engstrom *et al.*, 1991).

The landmarks defined in the laryngeal literature were not always well suited for use within an MRI data set because they were specified at extremes of the cartilages. We mimicked previous protocols (Maue and Dickson, 1971; Hicks, 1981; Kahane, 1982) to test the viability of using MRI for obtaining the same anatomical measures. The results were within the normal range and comparable to the measures reported by Hicks (1981). As MRI becomes an anatomical standard, new measures can be based on tissue interface and address theoretical needs.

The data published by Eckel and his colleagues (Eckel *et al.*, 1994; Eckel and Sittel, 1995; Eckel *et al.*, 1999) provide extensive descriptions of the cross-sectional anatomy of the larynx. If specimens were scanned in a high-resolution MR scanner before plastination, then segmentation of the MRI images could be compared with their plastinated counterparts. However, the MRI method is still preferred for determining the three-dimensional space coordinates and location of structures in the human larynx because of greater resolution between slices in the MRI.

C. Segmentation

To extract and measure the cartilages, the MRI images needed to be segmented into the individual cartilages (Udupa and Herman, 1991). Our focus was on using the segmented laryngeal cartilages for reconstruction of the surfaces, not on the segmentation procedures *per se*. There is an abundant literature on automatic and semiautomatic segmentation of MRI images (Kohn *et al.*, 1991; McInerney and Terzopoulos, 1995). Because our purpose was to develop an accurate anatomical framework, we chose not to implement these algorithms. If future research depends on the speed of segmentation, these algorithms will need to be considered. Although manual segmentation can involve tracing errors and limits in the number of specimens analyzed because it is labor intensive, it is the accepted standard for identifying structures (Collins *et al.*, 1998). Manual segmentation serves as the standard for testing automatic and semiautomatic segmentation methods in future applications (Collins *et al.*, 1998).

The intensity distribution of the laryngeal cartilages varied between specimens (Sakai *et al.*, 1990). Some variability can be attributed to the different tissues such as fatty and hemotopoietic marrow, and the degree of ossification among the major laryngeal cartilages. This variability limits the use of procedural algorithms for distinguishing the structures automatically.

The care used in the segmentation process minimized errors. By overlaying the segmented structures simultaneously in the same voxel locations as the original MRI, and displaying the overlaid images in different planes, sufficient accuracy was obtained. Within each of the two cricoid cartilages where segmentations were repeated, M1 and F1, the root-mean-square error was 0.4 voxels, with 99% of the surface voxels of one set having a voxel of the second set as a nearest neighbor. Using an automated algorithm would not have eliminated the existence of artifacts or insufficient contrast between some voxels.

The sizes of the five larynges varied substantially, with a twofold difference in cartilage volume between the largest and the smallest specimen. The female larynx was smaller than the four male larynges, but one male larynx was of comparable size to the female larynx with the arytenoids being slightly smaller. Despite these differences we were able to register all of the cricoid cartilages to **M1**.

The two dissected cartilages, M1 and M2, had different densities, although the densities were relatively uniform within each larynx. Different degrees of ossification, particularly across age groups, may account for this and indicates that a common mass cannot be assumed because densities will differ between larynges.

A major feature of the MRI scanning protocol was the use of a 3D Fourier transform acquisition to provide isotropic resolution. This was especially useful because of the subtle changes in tissue contrast between the cartilages and surrounding tissue. These changes appeared as fuzzy boundaries in the MRI images, making some cross sections challenging to identify from only a single series of planar slices. For example, a thin slice of cartilage may be indistinguishable from one plane but was readily identified from a perpendicular plane. We exploited the isotropy by tagging all of the voxels in each structure and viewing and editing these structures from all three principal planes.

D. Polygon model of the laryngeal cartilages

Rubin *et al.* (1998) developed polygonal models for each of the cartilages and intrinsic muscles contained in one human cadaveric larynx. Our focus here extended that approach by rendering the surfaces of the cartilages in an interactive environment that allowed the user to make measures. The interactive display of the cartilages would allow users to manipulate the view of the model to mimic that of a rigid laryngoscope. This would allow laryngoscopic-derived measures to be compared on the same plane as the anatomical framework.

E. Symmetry of the cricoid and thyroid cartilages

Our results confirmed the symmetry of the cricoid cartilage about a sagittal plane passing through the middle of the cartilage reported by others (Hirano *et al.*, 1989). Here, the plane of symmetry was estimated by minimizing the volume of the convex hull enclosing the larynx and its reflection about the plane of symmetry. When the symmetry between the right and left sides was compared (Table IV), the percentage differences between the two sides were modest (3% or less).

The results also confirmed the asymmetry of some thyroid cartilages (Hirano *et al.*, 1989). The percentage differences between the two sides of the thyroid cartilages were less than 6% with respect to the thyroid cartilage itself, but as high as 12.3% if the plane of symmetry of the cricoid cartilage was used. This difference in asymmetry of the thyroid cartilage for the two coordinate systems reflects a twisting of the thyroid cartilage relative to the cricoid cartilage (Table IV). It could be hypothesized that the movement of the vocal folds must accommodate for this twisting between the cartilages. That is, the vocal fold on the left would need to move a different amount from the vocal fold on the right side. Alternatively, some cartilage asymmetries may result from differences in the muscle forces between the two sides affecting development because fewer asymmetries were found in the infant larynx (Hirano *et al.*, 1989).

The algorithm for identifying the plane of symmetry for the cricoid cartilage was based on calculating the volume of the convex hull enclosing both the cartilage and its reflection. While this is not a direct anatomical measure of symmetry, it provided a robust performance index for the optimization code and resulted in identifying a transformation by which more intuitive measures of symmetry could be made. The measured sensitivity (s.d. $<0.5\%$) reflects the robustness of the method.

Lele (1991) discussed the comparison of shape based on landmark data. Lele's analysis has been used to describe the asymmetry in the human face landmark data (Ferrario *et al.*, 1995). We found only limited numbers of homologous identifiable landmarks that were out of the sagittal plane on the cricoid cartilage. When examining for landmarks on the lateral surface, for example, we found the shape of the cricothyroid joint capsule was not well defined in the MRI scans, and that the joints were not located symmetrically with respect to the airway.

Although landmarks were identified after segmentation, it would not normally be necessary to complete segmentation in order to locate landmarks and determine if a specimen was within the normal range. The segmentation of the convex hull may not require segmentation of the entire structure. While complete segmentation was considered important for these initial tests, future segmentations could be used which cover most of the surface but have missing data points.

Coordinate-free geometrical descriptions may be useful for morphological applications. Landmark distances for the individual laryngeal cartilages are usually specified in terms of height, length, and depth with respect to a standard anatomical orientation. Because such measures cannot determine spatial relationships between the laryngeal cartilages in the intact larynx, we defined homologous landmarks independent of a three-dimensional coordinate system.

F. Registration

We chose a modest set of scaling parameters for registration for the following reasons. Because the size of the laryngeal cartilages in small infants is linearly related to external measures of crown–rump length, crown–heel length, and head circumference (Schild, 1984), internal laryngeal measures may be linearly related as well. Next, we wanted parameter values that could be collected experimentally. Because the shape of the laryngeal airways can be estimated from x-ray fluoroscopy (Collett *et al.*, 1986) this might be a practical method for determining the size of the larynx.

Superimposition of images for registration or asymmetry measures can be determined most easily using both landmark data (Lele, 1991) and two-dimensional outline data (McAlarney, 1995). Finite-element method models (Lewis *et al.*, 1980) are essentially landmark models containing a substantial number of landmarks. By identifying conformation changes in node positions, structures can be compared. An important assumption is that the nodal positions identified in both structures are homologous and support the calculation of the deformation of the nodal locations.

Most methods of scaling structures, while maintaining form, are based on homologous anatomical landmarks (Lele, 1991). The landmarks of the laryngeal structures based on the MRI images can be ambiguous. While it would have been possible to use the Euclidean distance measurement technique of Lele (1991) to test for the comparison between structures, we considered the errors in landmark measures to be too great and used all the surface vertices of the object instead. McAlarney's (1995) method of scaling does not require landmarks, because it fits predetermined parallel slices of data. Our MRI data, however, were not all oriented in the same coordinate system. Unless the data were transformed first, the contours could not be aligned.

We also used the MRI images to locate the sites of attachment and location of the intrinsic laryngeal muscles. It was a shortcoming that we were not able to reliably distinguish the more superior medially directed portion of the posterior cricoarytenoid muscle; however, the vertically going belly is the largest and our location for this muscle was similar to that found in the canine using muscle dissection by Mineck *et al.* (2000). Our direction for the lateral cricoarytenoid muscle was in more of a superiorly going direction than found in the dog (Mineck *et al.*, 2000). This may be because of differences in the arytenoid cartilage size and orientation in the canine from the human. Also, we measured the intrinsic muscle fiber direction while the Mineck *et al.* (2000) study may have measured the more extrinsic fibers, which they suggest might be more curved. Further work, with higher-resolution MRI, should help to better define changes in muscle fiber directions in the human larynx, particularly as the muscles wrap around and between the cartilages which will be of importance for any biomechanical model (Bryant *et al.*, 1996).

G. Study limitations

Our study has several limitations that need to be addressed in future research. We only studied a small number of larynges. Due to availability, all were geriatric and therefore would have ossification of the laryngeal cartilages in contrast with those of younger adults. The greater ossification may have aided our segmentation process, while studies of younger larynges without ossification may be more difficult because of segmentation difficulties. Ongoing improvements in MRI scanning resolution may help to overcome these difficulties in the future. This study used a 2 T scanner; 7 T scanners are now becoming available for experimental use.

The use of only geriatric larynges limits the relevance of the results to this population, although our data were similar to previous studies (Hicks, 1981; Eckel *et al.*, 1994). Changes in laryngeal dimensions take place with aging, although how this would alter the three-dimensional framework is yet to be determined.

The use of fresh frozen human tissues also has several limitations; there may have been some differences in time between death and freezing allowing for tissue changes. A fixative was used to reduce signal drop-off at the edge of the tissues and preserve the specimens during the long MRI scan times. Kikinis *et al.* (1989) reported that the MRI intensity contrast within the larynx is unaffected by the fixative. The use of fixative, however, likely caused some tissue shrinkage, although Freidrich and Lichtenegger (1997) found no significant differences in their measures of the laryngeal cartilages and soft tissues between 20 fresh larynges and 30 larynges preserved in formalin. Therefore, we do not know the exact amount of shrinkage of the specimens due to the fixative. However, such shrinkage may have been less than the resolution limits of the MRI technique employed here where one voxel side was 0.43 mm. The most likely distortion caused by the fixative would have been in the cricoarytenoid joint capsule, which was examined with a great deal more accuracy in a previous study using a milling machine (Selbie *et al.*, 1998).

The segmentation process used here was highly labor intensive and only performed by one examiner. Therefore, we do not know the extent of possible interexaminer error. Examiner fatigue was not an issue, however, because segmentation was performed over several different sessions. With future improvements in MRI resolution, the segmentation should become more automatized, producing greater ease of use of this technique.

V. CONCLUSIONS

A three-dimensional anatomical framework could be used in future work on laryngeal modeling of vocal-fold opening and closing actions for speech, swallowing, and coughing (Wilhelms-Tricarico, 1995; Sanguineti *et al.*, 1998) and voice production (Titze *et al.*, 1988; Titze, 1994). Our attempt to register laryngeal anatomy, as has been performed in other disciplines (An *et al.*, 1979; Sommer *et al.*, 1982), was only the first step in this process. With improvements in MRI scanning, semiautomatic or automatic segmentation algorithms for segmenting MRI images will be possible. We attempted to provide a normalizing routine for the laryngeal structures that would support comparisons in shape between different larynges. By registration to a standard, a structure scanned at a lower resolution MRI could then be used in conjunction with videolaryngoscopy to relate vocal-fold movement to the underlying anatomy. Future use of sequential time-series MRI data could provide dynamic data for images with limited resolution. By having a registration standard, such data could be referenced and averaged. Finally, a particular patient's larynx could be scaled to the model after registration using anatomical landmarks. In this way, the model could be used to explore the behavior of pathological larynges for opening and closing the glottis.

ACKNOWLEDGMENTS

This work was supported by the Division of Intramural Research of the National Institute of Neurological Disorders and Stroke, Project Number IZ01 NS02980-01 and by NIH/NCRR No. P41 RR05959 to the Center for *In Vivo* Microscopy, Duke University.

Ajmani, M. L. (**1990**). ''A metrical study of the laryngeal skeleton in adult Nigerians,'' J. Anat. **171**, 187–191.

Alipour-Haghighi, F., and Titze, I. R. (**1985**). ''Viscoelastic modeling of canine vocalis muscle in relaxation,'' J. Acoust. Soc. Am. **78**, 1939–1943.

Alipour-Haghighi, F., and Titze, I. R. (**1991**). ''Elastic model of vocal fold tissues,'' J. Acoust. Soc. Am. **90**, 1326–1931.

Alipour-Haghighi, F., and Titze, I. R. (**1999**). ''Active and passive characteristics of the canine cricothyroid muscles,'' J. Voice **13**, 1–10.

An, K. N., Chao, E. Y., Cooney, W. P., and Linscheid, R. L. (**1979**). ''Normative model of human hand for biomechanical analysis,'' J. Biomech. **12**, 775–788.

Brizzi, E., Sgambati, E., Capaccioli, L., Giurovich, E., and Montagiani, L. (**1994**). ''A radiological-anatomical comparison between formalin-preserved organs and 'plastinated' ones,'' Ital. J. Anat. Embryol. **99**, 145–155.

Broad, D. J. (**1968a**). ''Kinematic considerations for evaluating laryngeal cartilage motions,'' Folia Phoniatr. **20**, 269–284.

Broad, D. J. (**1968b**). ''Some physiological parameters for prosodic description,'' SCRL Monograph no. 3, Speech Communications Research Laboratory, Santa Barbara, CA.

Bryant, N. J., Woodson, G. E., Kaufman, K., Rosen, C., Hengesteg, A., Chen, N., and Yeung, D. (**1996**). ''Human posterior cricoarytenoid muscle compartments: Anatomy and mechanics,'' Arch. Otolaryngol. Head Neck Surg. **122**, 1331–1336.

Castelijns, J. A., Doornbos, J., Verbeeten, B., Vielvoye, G. J., and Bloem, J. L. (**1985**). ''MR imaging of the normal larynx,'' J. Comput. Assist. Tomogr. **9**, 919–925.

Collett, P. W., Brancatisano, A. P., and Engel, L. A. (**1986**). ''Upper airway dimensions and movement in bronchial asthma,'' Am. Rev. Respir. Dis. **133**, 1143–1149.

Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., and Evans, A. C. (**1998**). ''Design and construction of a realistic digital brain phantom,'' IEEE Trans. Med. Imaging **17**, 463–468.

Eckel, H. E., Sittel, C., Zorowka, P., and Jerke, A. (**1994**). ''Dimensions of the laryngeal framework in adults,'' Surg. Radiol. Anat. **16**, 31–36.

Eckel, H. E., and Sittel, C. (**1995**). ''Morphometry of the larynx in horizontal sections,'' Am. J. Otol. **16**, 40–48.

Eckel, H. E., Kebke, J., Sittel, C., Sprinzl, G. M., Pototschnig, C., and Stennert, E. (**1999**). ''Morphology of the human larynx during the first 5 years of life studied on whole organ serial sections,'' Ann. Otol. Rhinol. Laryngol. **108**, 232–238.

Engstrom, C. M., Loeb, G. E., Reid, J. G., Forrest, W. J., and Avruch, L. (**1991**). ''Morphometry of the human thigh muscles. A comparison between anatomical sections and computer tomography and magnetic resonance imaging,'' J. Anat. **176**, 139–156.

Ferrario, V. F., Sforza, C., Miani, A., and Serrao, G. (**1995**). ''A three-dimensional evaluation of human facial asymmetry,'' J. Anat. **186**, 103–110.
Friedrich, G., and Lichtenegger, R. (**1997**). ''Surgical anatomy of the larynx,'' J. Voice **11**, 345–355.
Glanz, H. K. (**1984**). ''Carcinoma of the larynx. Growth, p-classification and grading of squamous cell carcinoma of the vocal cords,'' Adv. Oto-Rhino-Laryngol. **32**, 1–123.
Hicks, D. M. (**1981**). ''A morphological study of the aged human larynx,'' Doctoral dissertation, Vanderbilt University.
Hirano, M., Kurita, S., Yukizane, K., and Hibi, S. (**1989**). ''Asymmetry of the laryngeal framework: A morphologic study of cadaver larynges,'' Ann. Otol. Rhinol. Laryngol. **98**, 135–140.
Kahane, J. C. (**1978**). ''A morphological study of the human prepubertal and pubertal larynx,'' Am. J. Anat. **151**, 11–20.
Kahane, J. C. (**1982**). ''Growth of the human prepubertal and pubertal larynx,'' J. Speech Hear. Res. **25**, 446–455.
Kikinis, R., Wolfensberger, M., Boesch, C., and Martin, E. (**1989**). ''Larynx: MR imaging at 2.35 T,'' Laryngoscope **171**, 165–169.
Kohn, M. I., Tanna, N. K., Herman, G. T., Resnick, S. M., Mozley, P. D., Gur, R. E., Alavi, A., Zimmerman, R. A., and Gur, R. C. (**1991**). ''Analysis of brain and cerebrospinal fluid volumes with MR Imaging. I. Methods, reliability, and validation,'' Radiology **178**, 115–122.
Lele, S. (**1991**). ''Some comments on coordinate-free and scale-invariant methods in morphometrics,'' Am. J. Phys. Anthropol. **85**, 407–417.
Lewis, J. L., Lew, W. D., and Zimmerman, J. R. (**1980**). ''A nonhomogeneous anthropometric scaling method based on finite element principles,'' J. Biomech. **13**, 815–824.
Malinowski, A. (**1967**). ''The shape, dimensions and process of calcification of the cartilaginous framework of the larynx in relation to age and sex in the Polish population,'' Folia Morphol. (Prague) **26**, 118–128.
Maue, W. M., and Dickson, D. R. (**1971**). ''Cartilages and ligaments of the adult human larynx,'' Arch. Otolaryngol. **94**, 432–439.
McInerney, R., and Terzopoulos, D. (**1995**). ''A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4d image analysis,'' Comput. Med. Imaging Graph. **19**, 69–83.
McAlarney, M. E. (**1995**). ''Use of the boundary element method for biological morphometrics,'' J. Biomech. **28**, 609–616.
Mineck, C. W., Tayama, N., Chan, R., and Titze, I. R. (**2000**). ''Three-dimensional anatomic characterization of the canine laryngeal abductor and adductor musculature,'' Ann. Otol. Rhinol. Laryngol. **109**, 505–513.
Olofsson, J. (**1985**). ''Aspects on laryngeal cancer based on whole organ sections,'' Auris Nasus Larynx **12** (Suppl 2), S166–171.
Press, W. H., Flannery, B. T., Teukolsky, S. A., and Vetterling, W. T. (**1988**). *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge).
Raichle, M. E. (**1994**). ''Images of the mind. Studies with modern imaging techniques,'' Annu. Rev. Psychol. **45**, 333–356.
Rubin, J. S., Summers, P., and Harris, T. (**1998**). ''Visualization of the human larynx: A three-dimensional computer modeling tool,'' Auris Nasus Larynx **25**, 303–308.
Sakai, F., Gamsu, G., Dillon, W. P., Lynch, D. A., and Gilbert, T. J. (**1990**). ''MR imaging of the larynx at 1.5 T,'' J. Comput. Assist. Tomogr. **14**, 60–71.
Sanguineti, V., Laboissiere, R., and Ostry, D. J. (**1998**). ''A dynamic biomechanical model for neural control of speech production,'' J. Acoust. Soc. Am. **103**, 1615–1627.
Schild, J. A. (**1984**). ''Relationship of laryngeal dimensions to body size and gestational age in premature neonates and small infants,'' Laryngoscope **94**, 1284–1292.
Selbie, W. S., Zhang, L., Levine, W. S., and Ludlow, C. L. (**1998**). ''Determining the properties of the cricoarytenoid joint based on the anatomy of the articular facets,'' J. Acoust. Soc. Am. **103**, 1115–1127.
Sommer, H. J., Miller, N. R., and Pijanowski, G. J. (**1982**). ''Three-dimensional osteometric scaling and normative modeling of skeletal segments,'' J. Biomech. **15**, 171–180.
Sprinzl, G. M., Eckel, H. E., Sittel, C., Pototschnig, C., and Koebke, J. (**1999**). ''Morphometric measurements of the cartilaginous larynx: An anatomic correlate of laryngeal surgery,'' Head Neck **21**, 743–750.
Suddarth, S. A., and Johnson, G. A. (**1991**). ''Three-dimensional MR microscopy with large arrays,'' Magn. Reson. Med. **18**, 132–141.
Talairach, J., and Tournoux, P. (**1998**). *Co-planar Stereotaxic Atlas of the Human Brain* (Thieme Medical, New York).
Tayama, N., Kaga, K., Chan, R. W., and Titze, I. R. (**2001**). ''Geometric characterization of the laryngeal cartilage framework for the purpose of biomechanical modeling,'' Ann. Otol. Rhinol. Laryngol. **110**, 1154–1161.
Titze, I. R., Jiang, J., and Drucker, D. G. (**1988**). ''Preliminaries to the body-cover theory of pitch control,'' J. Voice **1**, 314–319.
Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).
Udupa, J. K., and Herman, G. T. (**1991**). *3D Imaging in Medicine* (CRC Press, Boca Raton, FL).
Watson, D. F. (**1981**). ''Computing the *n*-dimensional Delaunay tessellation with application of Voronoi polytopes,'' Computer **24**, 167–172.
Wilhelms-Tricarico, R. (**1995**). ''Physiological modeling of speech production: Methods for modeling soft-tissue articulators,'' J. Acoust. Soc. Am. **97**, 3085–3098.
Zajac, F. E., and Gordon, M. E. (**1989**). ''Determining muscles force and action in multi-articular movement,'' in *Exercise and Sport Science Reviews*, edited by K. Pandolf (Williams & Wilkins, Baltimore), pp. 187–230.

Perception of synthesized voice quality in connected speech by Cantonese speakers

Edwin M-L. Yiu[a)]
Voice Research Laboratory, Department of Speech and Hearing Sciences, The University of Hong Kong, 5/F Prince Phillip Dental Hospital, Sai Ying Pun, Hong Kong

Bruce Murdoch
Department of Speech Pathology & Audiology, The University of Queensland

Kathryn Hird
School of Psychology, Curtin University of Technology

Polly Lau
Voice Research Lab, Department of Speech and Hearing Sciences, The University of Hong Kong

(Received 20 July 2000; accepted for publication 18 June 2002)

Perceptual voice analysis is a subjective process. However, despite reports of varying degrees of intrajudge and interjudge reliability, it is widely used in clinical voice evaluation. One of the ways to improve the reliability of this procedure is to provide judges with signals as external standards so that comparison can be made in relation to these ''anchor'' signals. The present study used a Klatt speech synthesizer to create a set of speech signals with varying degree of three different voice qualities based on a Cantonese sentence. The primary objective of the study was to determine whether different abnormal voice qualities could be synthesized using the ''built-in'' synthesis parameters using a perceptual study. The second objective was to determine the relationship between acoustic characteristics of the synthesized signals and perceptual judgment. Twenty Cantonese-speaking speech pathologists with at least three years of clinical experience in perceptual voice evaluation were asked to undertake two tasks. The first was to decide whether the voice quality of the synthesized signals was normal or not. The second was to decide whether the abnormal signals should be described as rough, breathy, or vocal fry. The results showed that signals generated with a small degree of aspiration noise were perceived as breathiness while signals with a small degree of flutter or double pulsing were perceived as roughness. When the flutter or double pulsing increased further, tremor and vocal fry, rather than roughness, were perceived. Furthermore, the amount of aspiration noise, flutter, or double pulsing required for male voice stimuli was different from that required for the female voice stimuli with a similar level of perceptual breathiness and roughness. These findings showed that changes in perceived vocal quality could be achieved by systematic modifications of synthesis parameters. This opens up the possibility of using synthesized voice signals as external standards or ''anchors'' to improve the reliability of clinical perceptual voice evaluation. [DOI: 10.1121/1.1500753]

PACS numbers: 43.71.Bp, 43.71.Gv [CWT]

I. INTRODUCTION

Voice quality measurements are important in characterizing or describing a voice signal. The measures provide a severity index of dysphonic voice. Despite the rapid development of instrumentation in clinical voice assessment, perceptual voice evaluation is still a popular clinical procedure in documenting the severity of abnormal voice quality (Gerratt *et al.*, 1991). The major disadvantage of perceptual voice evaluation is that it is a subjective process and reliability is an issue. A review of the literature by Kreiman *et al.* (1993) showed that the reliability and agreement in voice quality rating could be as low as 18%, although it could improve with normal or extremely deviant qualities (see Murry *et al.*, 1987). It has been suggested that individuals develop mental (internal) standards for different voice quality through their previous exposure to voice samples (Kreiman *et al.*, 1993, 1992). These internal standards, however, are unstable and vary from one individual to another (Kreiman *et al.*, 1993). It has been demonstrated that when listeners were given explicit references (external anchors) during the rating tasks, the reliability of their judgments improved (Gerratt *et al.*, 1993; Kreiman and Gerratt, 1996). For example, Gerratt *et al.* (1993) demonstrated that the agreement in rating ''roughness'' improved from 50% (with no anchor) to 70% when anchors were provided. It is now generally accepted that the use of explicit external anchors would suppress the variable influence of the internal standards that different raters might have.

Currently, there are two possible types of external anchors that can be used to facilitate perceptual voice evaluation. One is natural occurring pathological voices and the other is synthesized signals. Synthesized signals have several advantages over natural occurring voice samples. With syn-

a)Electronic mail: edwinyiu@hku.hk

thesized signals, the number of signals that can be created is theoretically unlimited and is only restricted by the specificity of the synthesis parameters. With natural voice, a large set of pathological voice samples must exist first from which the appropriate anchors can be selected. Furthermore, it is relatively difficult to find a specific natural pathological voice which varies from other voice samples in a particular way. For example, finding a voice which is ''twice'' as breathy as another voice sample would be very difficult unless there is a large database from which one can choose. A third limitation of using natural pathological voice is that they rarely exhibit a single abnormal perceptual quality, but, instead, usually show combinations of several perceived qualities. Synthesized signals, however, do not suffer from this limitation. It is almost possible to systematically vary one particular parameter to achieve different degrees of abnormality in the synthesized signals. Other advantages of synthesized signals include simplicity and reproducibility. In natural pathological voice, acoustic properties are often complex. Many studies have attempted to extract the acoustic characteristics of these ''complex'' signals and to investigate how they affect perceptual judgment (for example, Deal and Emanuel, 1978; Hirano *et al.*, 1988; Kreiman *et al.*, 1990; Martin *et al.*, 1995; Wolfe *et al.*, 1997). Although conflicting results are shown by different studies, it is generally agreed that the two most commonly rated perceptual qualities, breathiness and roughness, are indeed multidimensional. In other words, both of these two perceptual qualities are found to correlate significantly with more than one acoustic property. For example, jitter, shimmer, and noise component have all been shown to correlate with the perception of rough and breathy quality. The reported correlation coefficients were generally of moderate strength (0.4 to 0.7). Since the acoustic properties of synthesized signals are determined by the synthesis parameters, a manipulation of the specific synthesis parameter will, in theory, produce comparatively fewer acoustically complex signals than natural voice samples. This may make it easier to study the relationship between acoustic properties and perceptual quality. In summary, provided all the synthesis parameters are detailed, these signals are relatively easy to reproduce. The ease of reproducibility of synthesized signals also facilitates replication of studies.

Although synthesized voice signals have advantages over natural voice samples in many ways, there are several limitations that investigators have to overcome. The first limitation is the naturalness of the synthesized signals. Due to the difficulty in synthesizing signals that sound natural when the speech materials get longer, perceptual voice quality studies which made use of synthesized signals used only single vowels (Bangayan *et al.*, 1997; Gerratt *et al.*, 1993; Martin and Wolfe, 1996) and avoided using connected speech. Several studies have provided some general guidelines in synthesizing natural sounding signals (Karlsson, 1991, 1992; Klatt and Klatt, 1990; Price, 1989). However, these techniques are not of much use for synthesizing connected speech.

The second limitation is related to the synthesis parameters available in the synthesizer. Generally, it has been shown that a noise component is necessary to model breathiness (Childers and Ahn, 1995; Childers and Lee, 1991; Hillenbrand, 1988; Klatt and Klatt, 1990; Martin and Wolfe, 1996) while a jitter component is needed to model roughness or aperiodicity (Hillenbrand, 1988; Klatt and Klatt, 1990). For example, it is claimed that the commercially available Klatt synthesizer (Klatt and Klatt, 1990) can change the perceived breathiness (by adjusting the aspiration noise, spectral tilt, open quotient, and increased bandwidths of first and second formants) and roughness (by adjusting the flutter). Whether these parameters are sufficient to synthesize signals that could be perceived as different degrees of pathological deviation has been questioned by some investigators (e.g., Bangayan *et al.*, 1997).

The present study had two objectives. First, it aimed to investigate whether a commercially available Klatt parallel/cascade speech model synthesizer could be used to create different pathological voice qualities using its available parameters. Second, it aimed to determine how the acoustic properties of the synthesized signals, as measured by jitter, shimmer, and noise to harmonic ratio, would affect perceptual voice quality judgment. If pathological voice quality could be synthesized successfully using a Klatt synthesizer, and was shown to correlate with perceptual ratings, this could ultimately provide a framework for creating ''reference'' voice qualities for evaluation and documenting abnormal voices.

The present study attempted a further step by synthesizing connected speech. The investigators of the present study, like other researchers (e.g., Hammarberg *et al.*, 1980; Kreiman and Gerratt, 2000), questioned the degree to which sustained vowels were representative in describing voice quality. We believe that connected speech should be used in perceptual voice evaluation because it is more representative of the voice used by speakers in daily speech tasks. Therefore, if one is to synthesize perceptual anchors with different voice qualities, connected speech should be used. In this study, we chose a simple subject–verb–object structure as the target connected speech. The Klatt synthesizer was chosen as it is commercially available and can be run on a personal computer with either a Macintosh or Window platform. This choice therefore makes it possible to allow other investigators to further explore this area without requiring more sophisticated instrument or special programming skills (cf. Hillenbrand, 1988).

II. METHODS

A. Preparation of the prototype stimuli

Synthesized signals based on a Cantonese sentence were created to simulate male and female voices. The signals were created using Sensimetrics' HLSyn Speech Synthesis System in a Microsoft Window platform. The prototype sentence used was

/baba	da	bɔ/	
father	hit	ball	(''father hits the ball'')

The HLSyn system is essentially a Klatt synthesizer (Klatt and Klatt, 1990) with the addition of some ''high-level'' synthesis parameters. In the present study, only the original, or

TABLE I. Percentage of stimuli within each synthesis parameter that were perceived as roughness, breathiness, and vocal fry. AH—amplitude of aspiration, AV—amplitude of voicing, DI—Diplophonia, FL—Flutter, OQ—Open quotient, and TL—spectral tilt. In some cases, more than one descriptor was used for the same stimulus; therefore, they may add up to more than 100%.

	AH	AV	DI	FL	TL	OQ	AV+DI
Rough	55%	29%	75%	52%	16%	15%	64%
Breathy	66%	53%	94%			19%	78%
Vocal fry			60%	8%			42%

''low-level,'' synthesis parameters were used. The average values of the synthesis parameters for the male and female prototype sentences were determined from analyzing sentences produced by six native Cantonese speakers (three males and three females) using fast Fourier transform (FFT) and linear predictive coding (LPC) analyses in the Kay Elemetric's Computerized Speech Lab 4300B system. The average values of the fundamental frequency (f_0), the first four formant frequencies (F1, F2, F3, and F4), and the duration of the vowels were used to synthesize the two prototype sentences. The fundamental frequency of the female signal was between 181 and 270 Hz, while that of the male signal was between 92 and 133 Hz. The variation in the fundamental frequency was due to the fact that the third word (/da_2/) of the sentence is a falling-rising tone. The values of these synthesis parameters were varied slightly by trial and error so that natural sounding prototype sentences, as determined by two native Cantonese speakers (authors EY and PL), were synthesized.

B. Pilot study

After the male and female prototype sentences were generated, seven synthesis parameters associated with voice qualities were varied independently with nine levels of severity to create 63 stimuli for each gender voice (a total of 126 stimuli). These seven parameters included amplitude of aspiration (AH) in dB, amplitude of voicing (AV) in dB, diplophonic double pulsing % (DI), flutter % (FL), open quotient % (OQ), spectral tilt of voicing source (TL) in dB, and amplitude of voicing in dB mixed with diplophonic double pulsing % (AV+DI). When one synthesis parameter was varied, the other parameters were all held constant at the Klatt's recommended default values. A pilot experiment was carried out using these 126 stimuli to determine (1) what perceptual voice qualities were to be included in the main study, and (2) which synthesis parameters were to be used in varying these voice qualities in the synthesized signals.

Five speech pathologists, each with at least two years of experience in assessing and treating voice disorders, were asked to serve as judges to listen to these synthesized stimuli. The judges were told that the stimuli were synthesized signals which represented different voice qualities. They were asked to label each signal with a descriptor which would best represent the voice quality. No specific instruction was given to the judges as to what descriptors were to be used.

Roughness, breathiness, and vocal fry were the three descriptors used overwhelmingly by the judges to describe the 126 stimuli. More than 75% of the stimuli were covered under these three descriptors. This was taken as an indicator that the Klatt synthesis parameters for voice quality could create signals primarily perceived as rough, breathy, or fry. These three descriptors were therefore used in the main study.

The data were further examined to determine which synthesis parameters were primarily responsible for signaling these three perceptual voice qualities. Table I lists the percentage of stimuli (with the male and female stimuli combined) within each synthesis parameter group. It was decided that the synthesis parameter which had 50% or more of its stimuli being perceived as rough, breathy, or fry were to be used in the main study to create stimuli with varying degree of roughness, breathiness, and fry. Therefore, the amplitude of aspiration (AH), diplophonia (DI), flutter (FL), amplitude of voicing (AV), and amplitude of voicing mixed with diplophonia (AV+DI) parameters were chosen to be used in the main study.

C. Main study

The objective of the main study was to investigate how the perception of different voice quality was determined by the synthesis parameters and the corresponding acoustic properties.

1. Preparation of stimuli with varying degree of abnormal voice quality

Based on the results of the pilot study (see Table I), the parameters AH, DI, FL, AV, and AV+DI were varied independently to synthesize different degree of voice quality. The incremental steps were 5 dB for the AH and AV, 10% for DI, and 20% for FL. For the stimuli which were varied in both AV and DI, each incremental step for AV was 5% (with the DI value set at 0%) until it reached the maximum value, i.e., 80%. From then onwards, the DI value was varied with 0.5% steps. Together with the prototype stimulus, this resulted in a total of 36 stimuli for each gender voice. Table II lists the synthesis parameters and the range of manipulation. When one synthesis parameter was varied, the other parameters were held constant at the Klatt's recommended default values.

Acoustic measures of jitter, shimmer, and noise-to-harmonic ratio using Kay's Computerized Speech Lab 4300B and Multidimensional Voice Program were carried out on extracted segments of these signals. Each extracted signal included the onset of the first word (/ba/) and the offset of the last word (/bɔ/). The Computerized Speech Lab has been shown to be tolerant to the fluctuation in acoustic properties in connected speech and provide valid acoustic results (Yiu *et al.*, 2000).

TABLE II. Incremental steps and range of manipulation of the values of synthesis parameters. (Default values for prototype stimulus: DI-0, AH40, AV60, FL0).

Parameter modified	Incremental steps	Range of manipulation
Amplitude of aspiration (AH)	8 Steps: AH5 for each step	AH45 to AH80
Diplophonia (DI)	10 Steps: DI-10 for each step	DI-10 to DI-100
Flutter (FL)	6 Steps: FL20 for each step	FL20 to FL100
Amplitude of voicing (AV)	4 Steps: AV5 for each step	AV65 to AV80
Amplitude of voicing at 80 dB plus diplophonia (AV80+DI)	7 Steps: DI-0.5 for each step	AV80DI-1 to AV80DI-4

2. Subjects

Twenty speech pathologists (17 females and 3 males) participated in the main study. They were all native Cantonese speakers. All had at least three years of experience in assessing and treating voice disorders on a daily basis.

3. Procedure

The synthesized stimuli were presented using a program written in Microsoft Visual Basic. The hardware system used included a Creative Sound Blaster Gold sound card and a pair of Sony SRS-PC51 speakers. The stimuli were presented in a random order to the listeners in a quiet room. Each stimulus was repeated twice, resulting in a total of 144 trials (72 female and 72 male stimuli). Precautions were taken, however, to prevent the same stimulus from being presented in a sequential manner. Half of the subjects were presented with the male stimuli first and the other half were presented with the female stimuli first. The subjects were asked whether the voice quality of each stimulus was normal, rough, breathy, or fry. Definitions of the three descriptors for abnormal quality were given to the subjects in writing during the procedure (see Table III). Subjects were given three trial items as practice before each set of stimuli was presented. The subject could choose to listen to each stimulus as many times as they would like in practice as well as in all trials.

III. RESULTS

For the acoustic measures of the female signals, the fundamental frequency was around 240 to 250 Hz, with the exception of the DI signals, which showed a frequency of around 127 Hz. This was approximately half the values of the other signal series. This halving of fundamental frequency, as pointed out by Klatt and Klatt (1990), could happen in signals where the alternate pulses disappear in extreme cases. The female AH and FL series showed a steady stepwise increase in all five acoustic measures (see Fig. 1). The female DI series also showed a general stepwise increase in the jitter (RAP and PPQ) and shimmer (Shim% and APQ) values, with the exception of DI-10 (which showed higher values in the jitter and shimmer measures when compared to those of the DI-20 signal) and DI-100 (which showed smaller values than those of DI-90). The female AV+DI series also demonstrated a general increase in the jitter (RAP and PPQ), shimmer (Shim% and APQ), and NHR, but the increase was not even for the whole AV+DI signal series. The last three signals showed relatively higher jitter and shimmer values.

For the acoustic measures of the male signals, the fundamental frequency was around 113 Hz, with the DI signals showing also about half of the values at 65 Hz. A general stepwise increase in all five acoustic measures was noticed in the male AH and FL series (see Fig. 1). Interestingly, the acoustic analysis of the male DI signal series showed a general decrease from the signal DI-20 to DI-90 in the jitter (RAP and PPQ) and shimmer (Shim% and APQ) values. This unusual finding may be due to the fact that by lowering the fundamental frequency below 60 Hz, the signal pulses contained less perturbation with the alternate pulses gone. A gentle and steady increase in RAP was noticed with the male AV+DI series.

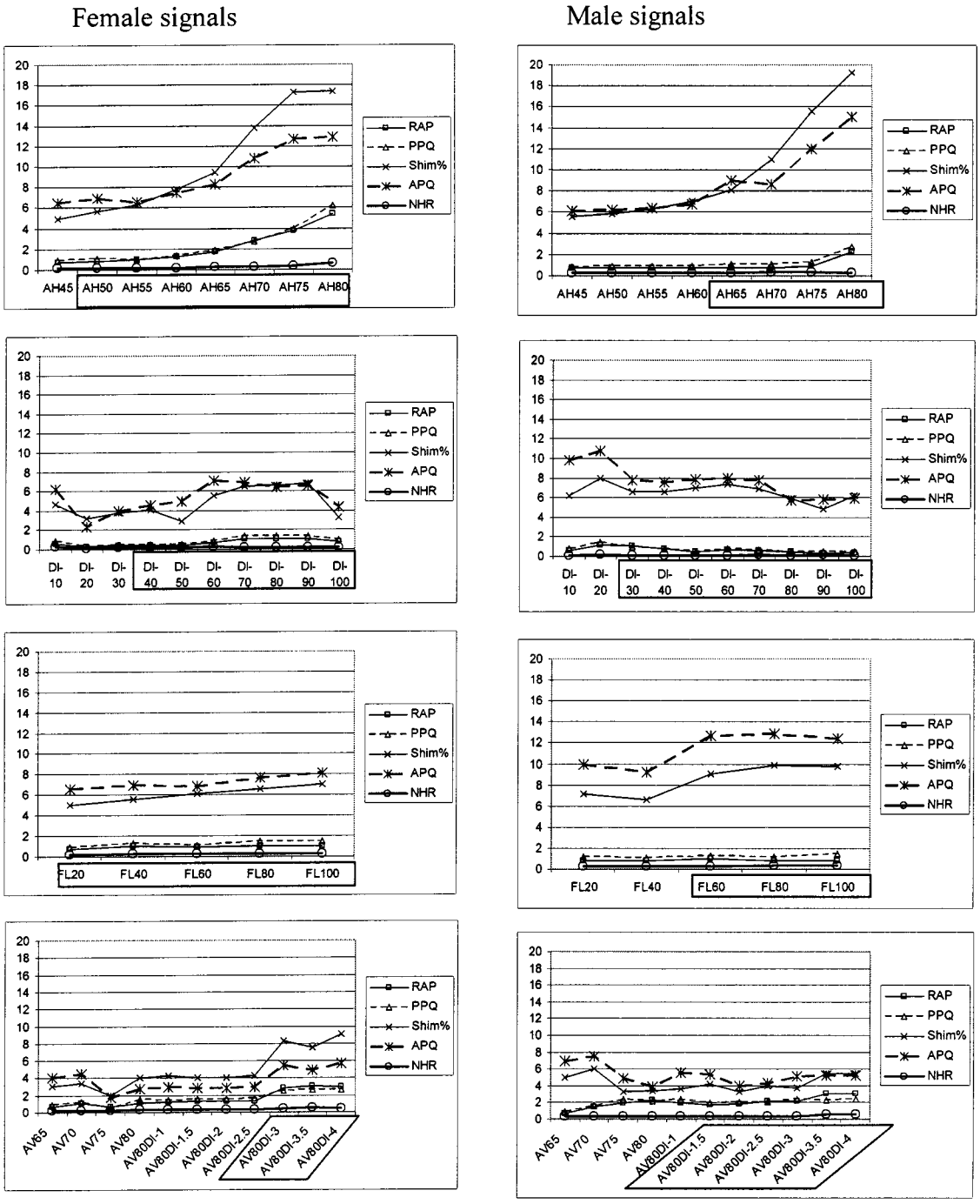
The responses of the subjects on each set of stimuli are given in Figs. 2–5. The figures show clearly that the number of subjects who perceived the stimuli as normal decreased

TABLE III. Definitions of abnormal voice qualities.

Rough
- Synonymous with "Harshness" or "Hoarseness"
- Perceptual correlates:
  - (1) Irregular quality
  - (2) Random fluctuations of glottal pulse
  - (3) Lack of clarity
  - (4) Uneven quality
- Acoustic correlates:
  - (1) Aperiodic mode of vibration
  - (2) Perturbation of the spectrum

Breathy
- Synonymous with "Whispery voice" or "Whisperiness"
- Perceptual correlates:
  - (1) Audible sound of expiration
  - (2) Audible air escape
  - (3) Audible friction noise
- Acoustic correlates:
  - (1) Related to a significant component of noise due to turbulence

Fry
- Synonymous with "Creaky"
- Perceptual correlates:
  - (1) Creaky, sounds like a creaking door
  - (2) Also sounds rough and low in pitch
- Acoustic correlate
  - (1) A complex pattern of subharmonics and modulations



RAP- Relative Average Perturbation,
PPQ- Pitch Perturbation Quotient, Shim%- Shimmer Percent,
APQ- Amplitude Perturbation Quotient, NHR- Noise to Harmonic Ratio

Synthesized values enclosed in boxes refer to signals being perceived by 75% of listeners as abnormal

FIG. 1. Acoustic measurements of synthesized signals.

with increasing values of the synthesis parameters. In other words, the higher the values of the synthesis parameters, the higher the number of subjects who perceived the signals as abnormal.

In order to determine the cutoff point for a set of stimuli to be perceived as abnormal, a binomial distribution was employed using a 95% confidence level. Since each subject had to decide first whether the stimulus was normal or abnormal, the chance level of making any judgment is 0.5. With a total of 20 listeners, a binomial distribution table indicated that at least 15 of them had to agree on the judgment in order to reach the 95% confidence level (Runyon *et al.*, 1996).

The signals which were determined by at least 15 or

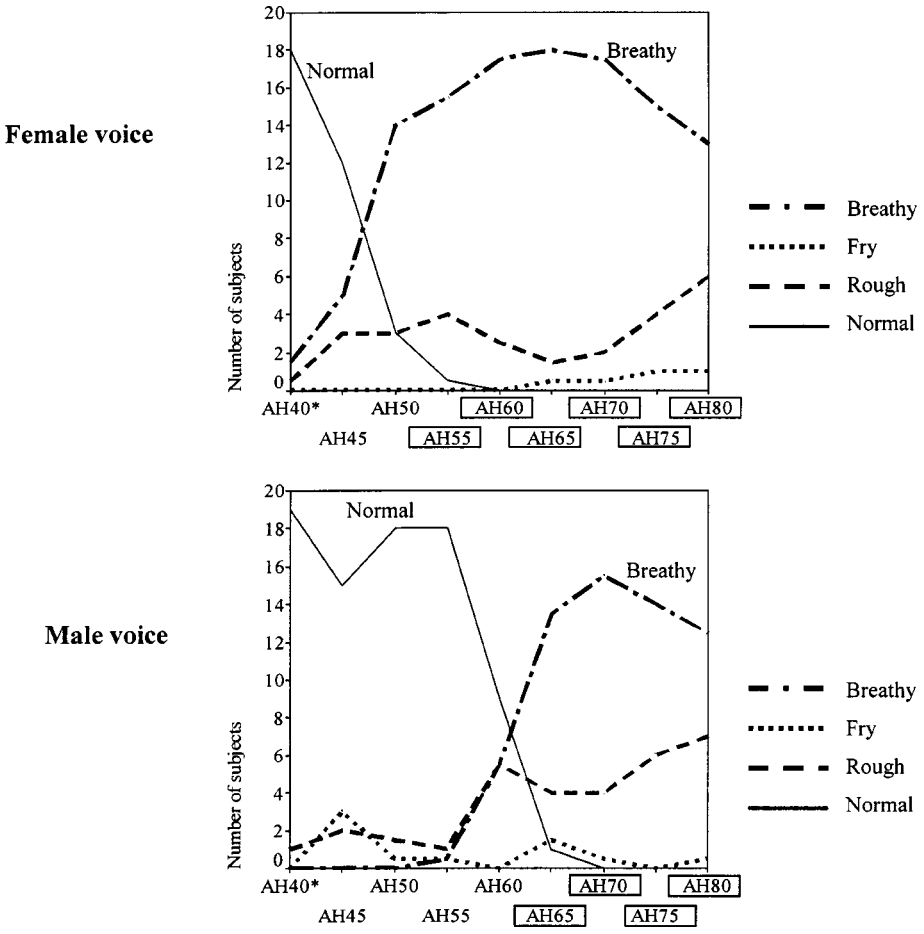


FIG. 2. The use of different descriptors in labeling stimuli with varying degree of amplitude of aspiration (AH).

* is the prototype stimulus

Synthesized values enclosed in boxes refer to signals being perceived by 75% of listeners as abnormal

more listeners to be abnormal are marked in Figs. 1–5 inside boxes. We further assumed that subjects would have one out of three chances to label the stimulus as either breathy, rough, or fry after having identified a stimulus as abnormal. By using the binomial distribution again, it was determined that at least 10 subjects would have to agree on a particular voice quality descriptor with a confidence level of 95%, assuming, of course, that at least 15 subjects had decided that a particular stimulus was abnormal.

For the stimuli that varied in AH and FL, the results clearly showed that they were perceived as breathy and rough, respectively, in both the female and male stimuli (see Table IV). For the DI parameter, male stimuli with high values of DI were all perceived as having a fry quality, while for the female stimuli the results were less clear. Female stimuli with DI values of 60 and 70 were perceived as rough, but the descriptor changed to vocal fry when the DI value increased to 100 (Table IV). When the DI parameter was varied in conjunction with a high AV, the male stimuli were generally perceived as rough. However, when the DI value increased up to 3%, a vocal fry quality was perceived. When the DI value increased to 4%, almost as many listeners perceived the stimuli as vocal fry as perceived them as rough (Table IV). For the female stimuli, only the AV80DI-3 stood out as having a rough quality. When the DI values increased, both rough and fry qualities were reported (Table IV).

It should be noted that female stimuli did not require so much aspiration (AH) as did the male stimuli to be perceived as breathy. The female stimuli were perceived as breathy starting at AH50 while the male stimuli were not perceived as breathy until AH had risen to 65 or above. When DI was added to the male stimuli, they were perceived as vocal fry when the DI value reached 30%. However, when DI was added to female stimuli, the perceived quality was less distinct. Apart from being perceived as vocal fry, roughness was also reported. Only when the DI was increased to 100% were the stimuli perceived distinctively as fry.

For the stimuli with variation of DI in combination with high levels of AV, the male stimuli were perceived as rough when the DI value started at 1.5% while the female required a DI of 3% to be perceived as rough. Once the DI values increased further to 3.5%, the fry quality began to appear in the perception of some listeners.

Table V shows the significant correlation coefficients between the number of judges that used a particular perceptual voice quality descriptor and the acoustic properties of the

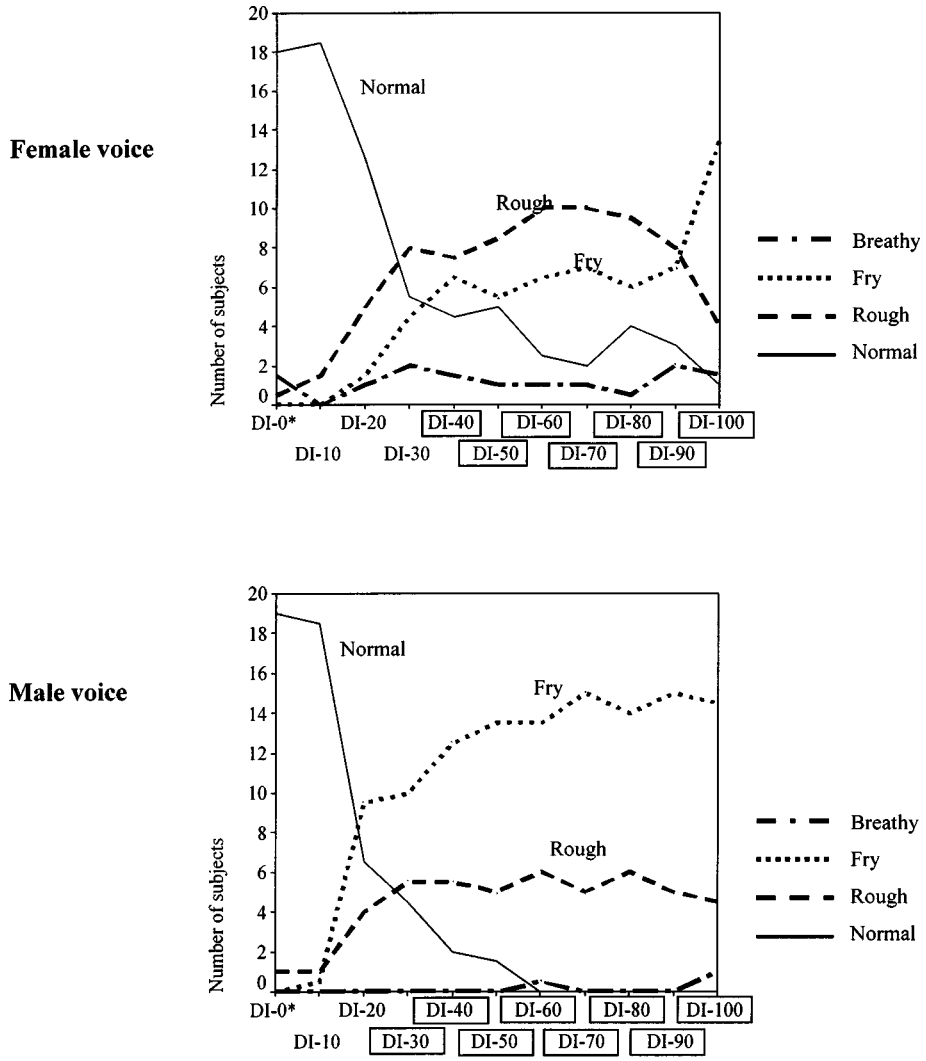


FIG. 3. The use of different descriptors in labeling stimuli with varying degree of diplophonia (DI).

* is the prototype stimulus

Synthesized values enclosed in boxes refer to signals being perceived by 75% of listeners as abnormal.

synthesized signals. The female AH stimuli showed no significant correlation while the male AH stimuli showed significant correlation between breathiness and three acoustic parameters (PPQ, Shim%, and APQ). The female DI stimuli showed a significant correlation between perceptual roughness and APQ whereas the male DI stimuli showed a significant negative correlation between perceptual roughness and shimmer percent. With the female FL stimuli, no significant correlation was found between perceptual roughness and any of the acoustic variables. The male FL stimuli, however, showed a significant correlation between roughness and shimmer percent. For the AV+DI stimuli, the female set showed significant correlation between perceptual rough and fry qualities with the RAP, whereas the male stimulus set demonstrated significant correlations between fry quality and the RAP as well as the NHR.

IV. DISCUSSION

This study shows the Klatt synthesizer can be used to create signals with different perceptual voice qualities by varying the synthesis parameters. It also shows that different degrees of synthesis values would be required to create a similar degree of perceptual quality in voices of different genders.

Synthesizing speech using a high values of the amplitude of aspiration (AH) parameter, as Klatt and Klatt (1990) contended, results in the perception of a breathy quality. In Fig. 1, it is clearly shown that increasing AH values resulted in a sharper increase in shimmer values (Shim% and APQ) and a moderate increase in jitter values (RAP and PPQ). These changes in acoustic properties appeared to account for the perception of breathiness. Klatt and Klatt (1990) suggested the default AH value be set at 40 dB so that a synthesized stimulus would sound natural. This study shows that relatively higher aspiration noise (AH) is needed in the male stimuli (at AH65) than in the female stimuli (at AH50) in order to produce a similar degree of perceptual breathiness. A closer examination of the acoustic properties of the signals (see Fig. 1) showed that the jitter (RAP and PPQ) had increased to a higher degree in the female signals than in the male signals with identical AH values. Therefore, this higher

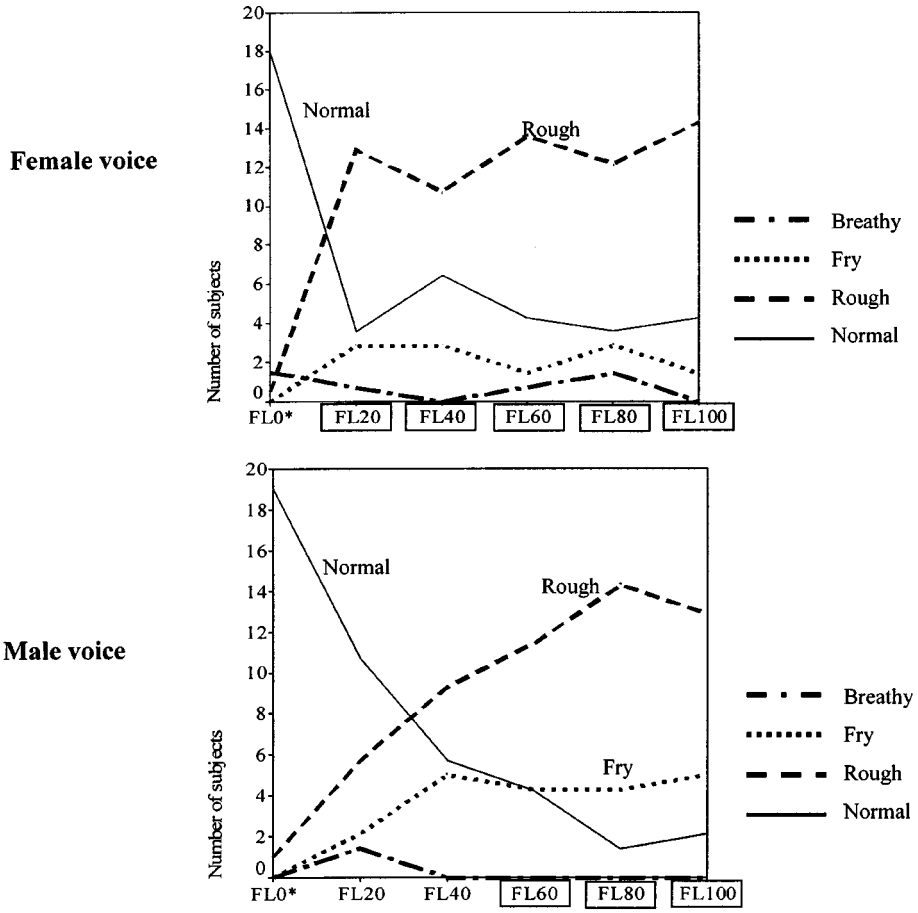


FIG. 4. The use of different descriptors in labeling stimuli with varying degree of flutter (FL).

* is the prototype stimulus

Synthesized values enclosed in boxes refer to signals being perceived by 75% of listeners as abnormal

degree of jitter in the female signals might have resulted in more breathy signals. The difference between the male and female signals was basically in the fundamental and formant frequencies. Therefore, the source of jitter might have come from the interaction of the frequency parameter with the aspiration parameter of the Klatt synthesizer. The correlation between the AH values in the male signals and the three acoustic measures (PPQ, Shim%, and APQ) are rather high. They are at least 0.78 or higher (see Table V). For the female AH signals, the ceiling effect might have accounted for the lack of correlation between the perceptual breathiness and any of the acoustic variables. This could be attributed to the high number of judges which perceived the female AH stimuli as breathy even with an AH value as low as 50.

The diplophonia (DI) parameter, according to Klatt and Klatt (1990), uses two glottal pulses in slightly different phases. In the present study, it was demonstrated that high values of this parameter are associated with perceptual roughness and vocal fry quality (see Fig. 3). When the DI value was increased beyond 5%, signals were primarily perceived as vocal fry in the male stimuli. However, in the female stimuli, an increase in the DI values was equally perceived as roughness or vocal fry quality. With the female DI signal series, one of the shimmer measures, APQ, correlated significantly with the perception of roughness (Spearman rho $= 0.73$, $p = 0.04$). However, no significant correlation was found between the acoustic measures and vocal fry quality. Indeed, the apparent negative correlation between the perceptual roughness and RAP might have been due to the error in extracting the perturbation measurements within the male DI stimuli as a result of the disappearance of alternate pulses in the signals.

Increasing values of the flutter (FL) parameter were found to produce a rough quality (see Fig. 4). Relatively higher flutter value was needed in the male stimuli (FL60) than in the female stimuli (FL20) in order to make the stimuli sound rough. Indeed, the male FL signal series already demonstrated relatively higher Shim% and APQ values than the female signals with the same FL values. The male FL stimuli showed a significant correlation between perceptual roughness and Shim% while there is a lack of correlation of any kind in the female stimuli. This lack of correlation appeared to be attributed to the ceiling effect of the subjects perceiving the female FL stimuli as perceptually rough. A number of subjects reported that the perception of roughness due to high FL values was very different from that produced by high AV + DI values. They reported a trembling quality in the stimuli synthesized with increased flutter (FL) values. Indeed, a reexamination of the data from the pilot

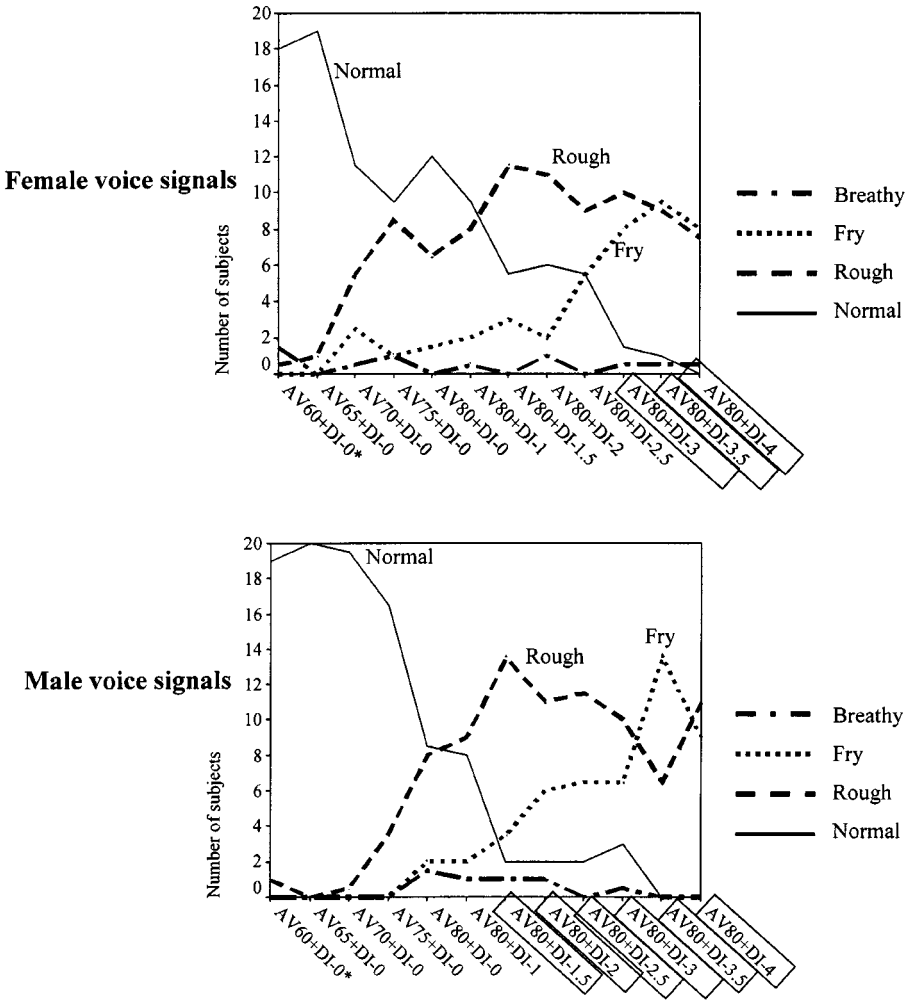


FIG. 5. The use of different descriptors in labeling stimuli with high amplitude of voicing and varying degrees of diplophonia (AV+DI).

* is the prototype stimulus

Synthesized values enclosed in boxes refer to signals being perceived by 75% of listeners as abnormal.

study showed that 46% of stimuli with increased FL values were perceived as showing tremor. However, as ''tremor'' was not an option given in the labeling task in the main study, the subjects might have been forced to choose roughness as the closest descriptor.

When the amplitude of voicing (AV) was increased to 80 dB and a few percent of DI was added, roughness was perceived. It should also be noted that none of the stimuli that varied only in the AV parameter was perceived as abnormal. Only when the DI was varied (even in small degree in the order of 1.5% to 3%) in combination with a high value of AV were the stimuli perceived as abnormal. Relatively higher degrees of AV plus DI are needed in the female stimuli (AV80DI-3) than in the male stimuli (AV80DI-1.5) in order to produce a rough or fry quality. Stimuli with high values of DI were perceived distinctively as vocal fry when synthesized as a male voice but equivocally as vocal fry and rough when synthesized as female voice. When AV80 was used to synthesize the signals, a relatively higher degree of DI was needed in the female signal (3% of DI) than in the male signals (1.5% of DI) in order to make the signals perceptually rough. The correlation of RAP and NHR with perceptual roughness and fry in these stimuli (see Table IV) showed the multidimensional nature of perceptual voice qualities and acoustic properties.

In summary, the Klatt synthesizer was found to be capable of synthesizing different degrees of breathiness, vocal fry, and roughness. Signals generated with a small degree of aspiration noise (AH) were perceived as breathy while small degrees of double pulsing (DI) or flutter (FL) were perceived as roughness. When the double pulsing (DI) and flutter (FL) increased, vocal fry was perceived instead of roughness.

Although some investigators (e.g., Klatt and Klatt, 1990; Bangayan *et al.*, 1997) contend that the Klatt synthesizer is better at synthesizing male voices than female voices, the present study demonstrated that it is possible to synthesize female voice with reasonably high quality. Nevertheless, the amount of AH, DI, or FL required to produce the perception of a similar level of pathological voice qualities was different for male and female voice stimuli.

The first objective of the study was to determine whether the Klatt synthesis parameters could be used to create signals with different types and degrees of voice quality. The findings from the present study show that the Klatt synthesizer

TABLE IV. Voice quality descriptors used by at least ten judges for particular synthesis parameters. AH—amplitude of aspiration, AV—Amplitude of voicing, DI—Diplophonia, and FL—Flutter.

Stimuli	Descriptors
Female	
AH50, AH55, AH60, AH65, AH70, AH75, AH80	Breathy
DI-40, DI-50	Rough and fry[a]
DI-60, DI-70,	Rough
DI-80, DI-90,	Rough and fry[a]
DI-100	Fry
FL20, FL40, FL60, FL80, FL100	Rough
AV80DI-3	Rough
AV80DI-3.5, AV80DI-4	Rough and fry[a]
Male	
AH65, AH70, AH75, AH80	Breathy
DI-30, DI-40, DI-50, DI-60, DI-70, DI-80, DI-90, DI-100	Fry
FL60, FL80, FL100	Rough
AV80DI-1.5, AV80DI-2, AV80DI-2.5, AV80DI-3, AV80DI-4	Rough
AV80DI-3.5	Fry

[a]None of the two descriptors was statistically more significant than the other, i.e., they did not reach the ''ten judges'' criterion. However, since similar numbers of judges were found in using these two descriptors, both descriptors are therefore reported here.

can be used to create synthesized voice signals with breathy, rough, and fry qualities. However, there are still some limitations with the Klatt synthesizer. First, the synthesized signals might not be exact matches to naturally occurring dysphonic qualities. This may have happened because when the fundamental frequency of the stimulus is not a whole multiple of the sampling rate, artifacts will be created and contribute to perceived roughness.[1] Therefore, uneven roughness could have distributed across the connected speech as each syllable had different fundamental frequency. A second limitation relates to the variation of voice quality across an utterance when connected speech material is used as the stimuli. It is known that voice quality would vary due to consonant articulation, use of different vowels (e.g., tensed versus laxed), or prosody changes (such as glottalization at phrase endings). In the present study, the quality settings were held constant across the whole utterance and these possible variations were not taken into consideration. These probably accounted for some of the ''unnaturalness'' in the synthesized dysphonic stimuli. Third, it is not known whether the Klatt synthesizer is capable of synthesizing all pathological voice qualities found in clinical situations using its current available synthesis parameters. As the Klatt synthesize is originally based on models derived from normal voices and is not designed to readily accommodate pathological qualities, such a question is a valid one. Indeed, more research is needed to develop appropriate models for pathological voice quality. A recent report by Bangayan *et al.* (1997) has explored some of the alternatives and has made two suggestions. The first is to include jitter and shimmer parameters in the Klatt synthesizer, and the second is to modify the DI parameter of the Klatt synthesizer so that fundamental frequency and amplitude could be varied separately. The DI parameter operates by truncating and reducing the amplitude of the closed phase of every second pulse. This is very different from natural signals. Therefore, the DI parameter produced effect which is not just perceived as diplophonia but as rough as well (see Table V). The fourth problem with the Klatt synthesizer, as noted by Hermes (1991), is that when noise is added up to a certain level, the noise is perceived as a separate noise stream rather than as a further increase in the breathiness of the noise signal. Finally, in the present study, it has been shown that the FL does not produce jitter appropriately as Klatt and Klatt (1990) claimed. Although increasing FL does alter the fundamental frequency

TABLE V. Correlation coefficients between values of acoustic parameters and the number of judges that used a particular descriptor. AH—Amplitude of aspiration, AV—amplitude of voicing, DI—diplophonia, FL—flutter, RAP— relative average perturbation, PPQ—pitch perturbation quotient, Shim%—shimmer percent, APQ—amplitude perturbation quotient, NHR—noise to harmonic ratio.

Varied synthesis parameters	Acoustic parameters	Perceptual descriptors	Spearman rho	Two-tailed *p* level
Female AH		Breathy	No significant correlation	
Female DI	APQ	Rough	0.73	0.04
		Fry	No significant correlation	
Female FL		Rough	No significant correlation	
Female AV+DI	RAP	Fry	0.75	0.03
	RAP	Rough	0.80	0.02
Male AH	PPQ	Breathy	0.78	0.02
	Shim%	Breathy	0.83	0.01
	APQ	Breathy	0.78	0.01
Male DI	RAP	Fry	−0.77	0.009
Male FL	Shim%	Rough	0.90	0.04
Male AV+DI	RAP	Fry	0.75	0.008
	NHR	Fry	0.82	0.002

in the time domain, this results in the perception of tremor, not roughness. Provided one takes these limitations into consideration when synthesizing pathological voice stimuli, the Klatt synthesizer is a useful signal synthesizer for researchers who want to study the perception of voice quality.

The second objective of this study was to determine how the acoustic properties affected perceptual voice evaluation. Previous studies have shown that the correlation between acoustic variable and perceptual qualities varies between 0.4 and 0.7 and a particular perceptual quality may correlate with several acoustic measures (Kreiman and Gerratt, 2000). The results from our present study also support these findings. The significant correlation coefficients were moderately high (>0.73; see Table V). Furthermore, many perceptual qualities were also found to correlate with more than one acoustic variable. Nevertheless, we are unable to make a direct comparison between our data and those from the previous studies as the coefficients from the present study were based on the number of judges agreeing on a particular quality rather than on the severity of each quality. It would be more appropriate to investigate the correlation between the acoustic variables and the severity ratings made by the judges on different perceptual qualities. Nevertheless, it is clear from the results that perceptual voice quality is multidimensional in nature. This means that it is determined not by a single acoustic variable, but more likely by a set of acoustic variables.

A pertinent question that many voice clinicians may ask is whether these synthesized signals have any clinical significance. Chan and Yiu (2002) employed the synthesized signals developed in the present study as anchors to investigate whether they could improve the reliability of perceptual voice rating. Their findings showed that the use of synthesized signals as anchors facilitated a better reliability than natural voice anchors (Chan and Yiu, 2002). Therefore, these synthesized signals are of clinical importance as they are the pertinent materials for investigating perceptual voice evaluation. It is hoped that the present study serves to generate more interest in investigating the process of voice quality perception.

ACKNOWLEDGMENTS

This study was supported by a grant from the Hong Kong Research Grant Council (HKU7196/98H). The authors would like to thank Dr. Alex Francis, Dr. Christopher Turner, and two anonymous reviewers for their constructive comments on the initial draft of the manuscript.

[1]This particular issue related to the use of Klatt synthesizer was brought to the attention of the authors by an anonymous reviewer of the manuscript.

Bangayan, P. T., Long, C., Alwan, A. A., Kreiman, J., and Gerratt, B. R. (**1997**). ''Analysis by synthesis of pathological voices using the Klatt synthesizer,'' Speech Commun. **22**, 343–368.

Chan, K. M.-K., and Yiu, E. M. L. (**2002**). ''The effect of anchors and training on the reliability of perceptual voice evaluation,'' J. Speech Lang. Hear. Res. **45**, 111–126.

Childers, D. G., and Ahn, C. (**1995**). ''Modeling the glottal volume-velocity waveform for three voice types,'' J. Acoust. Soc. Am. **97**, 505–519.

Childers, D. G., and Lee, C. K. (**1991**). ''Vocal quality factors: analysis, synthesis, and perception,'' J. Acoust. Soc. Am. **90**, 2394–2410.

Deal, R., and Emanuel, F. (**1978**). ''Some waveform and spectral features of vowel roughness,'' J. Speech Hear. Res. **21**, 250–264.

Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N., and Berke, G. S. (**1993**). ''Comparing internal and external standards in voice quality judgments,'' J. Speech Hear. Res. **36**, 14–20.

Gerratt, B. R., Till, J. A., Rosenbek, J. C., Wertz, R. T., and Boysen, A. E. (**1991**). ''Use and perceived value of perceptual and instrumental measures in dysarthria management,'' in *Dysarthria and Apraxia of Speech: Perspective on Management*, edited by C. A. Moore, K. M. Yorkston, and D. R. Beukelman (Brookes, Baltimore), pp. 77–93.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., and Wedin, L. (**1980**). ''Perceptual and acoustic correlates of abnormal voice qualities,'' Acta Otolaryngol. Suppl. (Stockh) **90**, 441–451.

Hermes, D. J. (**1991**). ''Synthesis of breathy vowels: Some research methods,'' Speech Commun. **10**, 497–502.

Hillenbrand, J. (**1988**). ''Perception of aperiodicities in synthetically generated voices,'' J. Acoust. Soc. Am. **83**, 2361–2371.

Hirano, M., Hibi, S., Yoshida, T., Hirade, Y., Kasuya, H., and Kikuchi, Y. (**1988**). ''Acoustic analysis of pathological voice,'' Acta Otolaryngol. (Stockh) **105**, 432–438.

Karlsson, I. (**1991**). ''Female voices in speech synthesis,'' J. Phonetics **19**, 111–120.

Karlsson, I. (**1992**). ''Modeling voice variations in female speech synthesis,'' Speech Commun. **11**, 491–495.

Klatt, D. H., and Klatt, L. C. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**(2), 820–857.

Kreiman, J., and Gerratt, B. R. (**1996**). ''The perceptual structure of pathologic voice quality,'' J. Acoust. Soc. Am. **100**, 1787–1795.

Kreiman, J., and Gerratt, B. R. (**2000**). ''Measuring vocal quality,'' in *Voice Quality Measurement*, edited by M. Kent and M. Ball (Singular-Thomson Learning, San Diego), pp. 73–102.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (**1993**). ''Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research,'' J. Speech Hear. Res. **36**, 21–40.

Kreiman, J., Gerratt, B. R., and Precoda, K. (**1990**). ''Listener experience and perception of voice quality,'' J. Speech Hear. Res. **33**, 103–115.

Kreiman, J., Gerratt, B. R., Precoda, K., and Berke, G. S. (**1992**). ''Individual differences in voice quality perception,'' J. Speech Hear. Res. **35**, 512–520.

Martin, D. P., and Wolfe, V. I. (**1996**). ''Effects of perceptual training based upon synthesized voice signals,'' Percept. Mot. Skills **83**(3) (part 2), 1291–1298.

Martin, D., Fitch, J., and Wolfe, V. (**1995**). ''Pathological voice type and the acoustic prediction of severity,'' J. Speech Hear. Res. **38**, 765–771.

Murry, T., Brown, W. S. J., and Rothman, H. (**1987**). ''Judgments of voice quality and preference: Acoustic interpretation,'' J. Voice **1**, 252–257.

Price, P. J. (**1989**). ''Male and female voice source characteristics: Inverse filtering results,'' Speech Commun. **8**, 261–277.

Runyon, R. P., Harber, A., Pittenger, D. J., and Coleman, K. A. (**1996**). *Fundamentals of Behavioural Statistics* (McGraw-Hill, New York).

Wolfe, V., Fitch, J., and Martin, D. (**1997**). ''Acoustic measures of dysphonic severity across and within voice types,'' Folia Phoniatr. Logop. **49**, 292–299.

Yiu, E., Worrall, L. E., Longland, J., and Mitchell, C. (**2000**). ''Analysing vocal quality of connected speech using Kay's Computerized Speech Lab: A preliminary finding,'' Clin. Linguist. Phonetics **14**(4), 295–305.

The intelligibility of speech with "holes" in the spectrum

Kalyan Kasturi and Philipos C. Loizou[a)]
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688

Michael Dorman and Tony Spahr
Department of Speech and Hearing Sciences, Arizona State University, Tempe, Arizona 85287

(Received 9 January 2002; revised 11 June 2002; accepted 14 June 2002)

The intelligibility of speech having either a single ''hole'' in various bands or having two ''holes'' in disjoint or adjacent bands in the spectrum was assessed with normal-hearing listeners. In experiment 1, the effect of spectral ''holes'' on vowel and consonant recognition was evaluated using speech processed through six frequency bands, and synthesized as a sum of sine waves. Results showed a modest decrease in vowel and consonant recognition performance when a single hole was introduced in the low- and high-frequency regions of the spectrum, respectively. When two spectral holes were introduced, vowel recognition was sensitive to the location of the holes, while consonant recognition remained constant around 70% correct, even when the middle- and high-frequency speech information was missing. The data from experiment 1 were used in experiment 2 to derive frequency-importance functions based on a least-squares approach. The shapes of the frequency-importance functions were found to be different for consonants and vowels in agreement with the notion that different cues are used by listeners to identify consonants and vowels. For vowels, there was unequal weighting across the various channels, while for consonants the frequency-importance function was relatively flat, suggesting that all bands contributed equally to consonant identification. [DOI: 10.1121/1.1498855]

PACS numbers: 43.71.Es, 43.71.Ky [CWT]

I. INTRODUCTION

It is generally accepted that human listeners rely on cues that exist across several frequency bands to understand speech. The question of how listeners use and combine information across several frequency bands when understanding speech is one that puzzled researchers for many decades. One of the earliest attempts to answer that question was taken by French and Steinberg (1947) with the computation of the articulation index (AI). By systematically low-pass and high-pass filtering the spectrum and measuring speech recognition, French and Steinberg (1947) determined the relative importance of various frequency bands. Although the AI method was found to be very successful in predicting speech intelligibility in many listening conditions, it has one major shortcoming. The AI method does not take into account the fact that listeners may combine and utilize speech information from multiple disjoint bands (e.g., Grant and Braida, 1991).

Although many studies investigated the intelligibility of high-passed-, low-passed- (e.g., French and Steinberg, 1947; Pollack, 1948; Kryter, 1962), and bandpassed-filtered speech (Warren *et al.*, 1995; Stickney and Assmann, 2001), not many studies have investigated the perception of bandstopped-filtered speech (i.e., speech with holes in the spectrum) or speech composed of disjoint frequency bands. Lippmann (1996) investigated the intelligibility of consonants that had a single hole in the middle of the spectrum. High consonant intelligibility (~90% correct) was maintained even after removing speech energy in the middle frequencies (800 to 4 kHz). Shannon *et al.* (2001) assessed the impact of the size and location of spectral holes with cochlear-implant and normal-hearing listeners. For the normal-hearing listeners, holes were created by dropping off 2 to 8 low-, middle-, or high-frequency bands in a 20-noise-band cochlear-implant (CI) simulation. Results showed that holes in the low-frequency region were more damaging than holes in the middle- and high-frequency regions on speech recognition. In the study by Shannon *et al.* a single hole (varying in size) in the low-, middle-, or high-frequency regions of the spectrum was introduced. A few other studies (e.g., Breeuwer and Plomp, 1984; Dorman *et al.*, 1989; Riener *et al.*, 1992) investigated speech recognition of disjoint bands of low- and high-frequency information. Synergy effects were demonstrated in the study by Riener *et al.* (1992) when subjects were presented with spectral information contained in the low- and high-frequency bands. The intelligibility of sentences through single one-third-octave bands centered around either 370 Hz or 6 kHz was roughly 23% when presented alone, but increased to 77% correct when presented simultaneously. The study by Riener *et al.* (1992), as well as those of others, demonstrated that having access to low- and high-frequency information enabled listeners to identify speech with relatively high accuracy. Listeners seemed to ''fill in'' the missing speech information.

The aforementioned studies examined speech recognition either for a single hole varying in frequency location (and size) or for a single hole in the middle of the spectrum. The scope of those studies was therefore limited in the sense that it did not consider how speech is recognized when it is composed of multiple disjoint bands involving low-, middle-, and/or high-frequency information. The present study ad-

a)Electronic mail: loizou@utdallas.edu

TABLE I. The first two formant frequencies (in Hz) of the male and female vowels used in this study.

		had	hod	head	hayed	heard	hid	heed	hoed	hood	hud	who'd
*F*1	Male	627	786	555	438	466	384	331	500	424	629	319
	Female	666	883	693	492	518	486	428	538	494	809	435
*F*2	Male	1910	1341	1851	2196	1377	2039	2311	868	992	1146	938
	Female	2370	1682	1991	2437	1604	2332	2767	998	1102	1391	1384

dressed this question in a systematic fashion considering all possible combinations of missing disjoint bands from the spectrum.

The answer to the question of how listeners use and combine information across frequency bands, whether isolated or disparate, is not only important for understanding speech perception but it is also important for understanding speech perception by cochlear-implant listeners or hearing-impaired listeners in general. Cochlear implants are based on the idea that there are surviving neurons in the vicinity of the electrodes. The lack of hair cells and/or surviving neurons in certain areas of the cochlea essentially creates ''hole(s)'' in the spectrum. The extent of the effect of holes in the spectrum on speech understanding is not well understood. It is not known, for instance, whether the spectral holes can account for some of the variability in performance among CI listeners. It is therefore of interest to first find out which set of hole pattern(s) is most detrimental for speech recognition. The answer to that question would then be useful for determining ways to somehow make up for the lost information.

The aim of this study is to examine the effect of the location and size of spectral holes on vowel and consonant recognition. Understanding this effect will provide insights as to why some CI listeners do not perform well, despite the wealth of information they receive [cochlear-implant listeners receive only a small number (4–6) of channels of frequency information, despite the fact that some implant processors transmit as many as 20 channels of information (e.g., Fishman *et al.*, 1997; Dorman *et al.*, 2000)]. In addition, we could use the data of this study to develop a frequency-importance function that takes into account the fact that listeners could combine information from disparate frequency bands in the spectrum. In experiment 1, speech was processed through six frequency bands, and synthesized as a sum of sine waves with amplitudes equal to the rms energy of each frequency band, and frequencies equal to the center frequencies of the bandpass filters. [Six channels were used as we found in previous studies (e.g., Loizou *et al.*, 1999) that six channels were enough to achieve high levels of speech understanding.] To synthesize speech with a hole in a certain frequency band, we set the corresponding sine wave amplitude to zero. We systematically created holes in each of the six frequency bands (one hole at a time) and examined vowel and consonant recognition. Similarly, speech was synthesized with two holes in the spectrum, by setting the corresponding sine wave amplitudes to zero. All possible combinations were created, including the scenarios where two holes were in adjacent frequency bands (thus making a larger hole) or where the two holes were in disjoint frequency bands. The data from experiment 1 were used in experiment 2 to derive frequency importance functions for vowel and consonant recognition.

II. EXPERIMENT 1: HOLES IN THE SPECTRUM

The intelligibility of speech having either a single hole in various bands or having two holes in disjoint or adjacent bands in the spectrum was assessed with normal-hearing listeners. The extent of the effect of the location, size, and pattern of spectral holes on vowel and consonant recognition was evaluated.

A. Method

1. Subjects

Twenty normal-hearing listeners (20 to 25 years of age) participated in this experiment. All subjects were native speakers of American English. The subjects were paid for their participation. Eleven of the subjects were tested at the University of Texas at Dallas and the remaining nine subjects were tested at Arizona State University.

2. Speech material

Subjects were tested on consonant and vowel recognition. The consonant test used 16 consonants in /aCa/ context taken from the Iowa consonant test (Tyler *et al.*, 1987). All the syllables were produced by a male speaker.

The vowel test included the words: ''heed, hid, hayed, head, had, hod, hud, hood, hoed, who'd, heard'' produced by male and female talkers. A total of 22 vowel tokens was used for testing, 11 produced by 7 male speakers and 11 produced by 6 female speakers [not all speakers produced all 11 vowels]. The stimuli were drawn from a set used by Hillenbrand *et al.* (1995). The first two formant frequencies (as estimated by Hillenbrand *et al.*) of the vowels used for testing are given in Table I.

3. Signal processing

Speech material was first low-pass filtered using a sixth-order elliptical filter with a cutoff frequency of 6000 Hz. Filtered speech was passed through a pre-emphasis filter with a cutoff frequency of 2000 Hz. This was followed by bandpass filtering into six different frequency bands using sixth-order Butterworth filters with center frequencies of 393, 639, 1037, 1685, 2736, and 4444 Hz, respectively. The frequency boundaries of the six bands are given in Table II. The filters were designed to span the frequency range from 300 to 5500 Hz in a logarithmic fashion. The output of each channel was passed through a rectifier followed by a second-order Butterworth low-pass filter with a center frequency of 400 Hz to obtain the envelope of each channel output. Cor-

TABLE II. The 3-dB frequency boundaries of the six bands with the corresponding center frequencies (Hz) of each band.

Band	Lower frequency (Hz)	Upper frequency (Hz)	Center frequency (Hz)
1	300	487	393
2	487	791	639
3	791	1284	1037
4	1284	2085	1685
5	2085	3388	2736
6	3388	5500	4444

responding to each channel a sinusoid was generated with frequency set to the center frequency of the channel and with amplitude set to the root-mean-squared (rms) energy of the channel envelope estimated every 4 ms. The phases of the sinusoids were estimated from the fast Fourier transform (FFT) of the speech segment. The sinusoids of each band were finally summed and the level of the synthesized speech segment was adjusted to have the same rms value as the original speech segment.

To create a hole in frequency band N ($1 \leqslant N \leqslant 6$), we set the amplitude of the sinusoid corresponding to frequency band N to zero. Speech was synthesized using the remaining five channel amplitudes. Similarly, two holes were created by setting the amplitudes of the sinusoids in frequency bands M and N to zero. Speech was synthesized using the remaining four channel amplitudes.

Vowel and consonant stimuli were created for six single-hole conditions and 15 two-hole conditions as shown in Table III. All possible combinations of removing two out of the six frequency bands were considered. For comparative purposes, we also created a baseline condition in which we did not remove any frequency bands. Overall, subjects were tested with a total of 22 conditions.

TABLE III. The 22 test conditions considered in this study. The zeroth condition corresponds to the baseline condition. The channel(s) removed in each condition are indicated with a zero.

Condition	Channel 1	Channel 2	Channel 3	Channel 4	Channel 5	Channel 6
0	1	1	1	1	1	1
1	0	1	1	1	1	1
2	1	0	1	1	1	1
3	1	1	0	1	1	1
4	1	1	1	0	1	1
5	1	1	1	1	0	1
6	1	1	1	1	1	0
7	0	0	1	1	1	1
8	0	1	0	1	1	1
9	0	1	1	0	1	1
10	0	1	1	1	0	1
11	0	1	1	1	1	0
12	1	0	0	1	1	1
13	1	0	1	0	1	1
14	1	0	1	1	0	1
15	1	0	1	1	1	0
16	1	1	0	0	1	1
17	1	1	0	1	0	1
18	1	1	0	1	1	0
19	1	1	1	0	0	1
20	1	1	1	0	1	0
21	1	1	1	1	0	0

4. Procedure

The experiments were performed on a PC equipped with a Creative Labs SoundBlaster 16 soundcard. Stimuli were played to the listeners monaurally through Sennheiser HD 250 Linear II circumaural headphones. The words were displayed on a computer monitor, and a graphical user interface was used that enabled the subjects to indicate their response by clicking a button corresponding to the word played. No feedback was given during the test.

At the beginning of each test the subject was presented with a practice session in which the vowels or consonants were processed through six channels—no holes were introduced (baseline condition). After the practice session, the subjects were tested with the various spectral hole conditions. Two groups of subjects were used, 11 from University of Texas—Dallas and 9 from Arizona State University. The 11 subjects at The University of Texas at Dallas were tested with the 14 test conditions labeled as, 0, 1, 2, 3, 4, 5, 6, 7, 9, 15, 16, 18, 20, and 21 in Table III. The zeroth condition corresponded to the baseline condition in which all six channels were present. The nine subjects at Arizona State University were tested with the fifteen conditions labeled as 0, 1, 2, 3, 4, 5, 6, 8, 10, 11, 12, 13, 14, 17, and 19 in Table III. Note that both groups of subjects were tested with the baseline condition. The order in which the conditions were presented was partially counterbalanced between subjects to avoid order effects. In the vowel and consonant tests, there were nine repetitions of each vowel and each consonant. The vowels and the consonants were completely randomized.

B. Results

The mean percent-correct scores for the single-hole conditions are shown in Fig. 1. A one-way ANOVA with repeated measures showed a significant main effect of the location of the spectral hole [$F(6,60)=9.7, p<0.0005$] on consonant recognition. *Post hoc* tests according to Tukey (at alpha=0.05) showed that the scores obtained with channels 4, 5, or 6 off were significantly lower than the baseline score. The average scores of the baseline condition were not significantly different ($p=0.313$) between the two groups of subjects. The scores obtained with channels 1, 2, or 3 off were not significantly different from the baseline score.

The consonant confusion matrices were analyzed in terms of percent information transmitted as per Miller and Nicely (1955). The feature analysis is shown in Fig. 2. A one-way ANOVA showed a nonsignificant effect [$F(6,60)=14.6, p=0.484$] for the feature "manner" and a nonsignificant effect for the feature "voicing" [$F(6,60)=2.7, p=0.061$]. The feature "place" was significantly [$F(6,60)=15.6, p<0.0005$] affected. *Post hoc* Tukey tests showed that conditions in which channel 4, 5, or 6 were removed were significantly different from the baseline condition.

For the vowel data, a one-way ANOVA showed a significant main effect [$F(6,54)=14.5, p<0.0005$] of the location of the spectral hole on vowel recognition. A *post hoc*

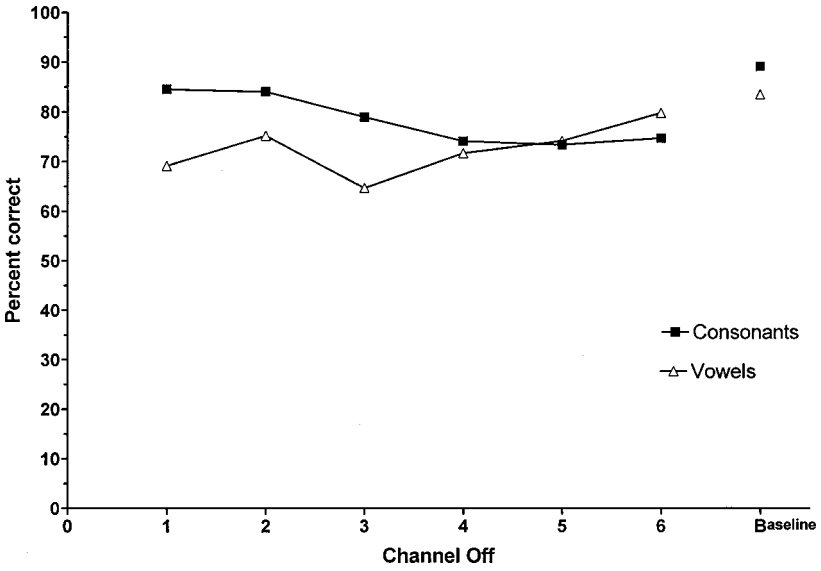


FIG. 1. Mean percent scores for vowel and consonant recognition as a function of the location of the spectral hole. The holes were centered around the channel center frequencies. In the baseline condition, all channels were present.

Tukey test showed that the scores obtained with channels 1, 3, or 4 off were significantly different from the baseline score ($p<0.05$). The score obtained when channel 2 was off was not significantly different from the baseline score. The fact that channels 1, 3, and 4 were found to have a significant effect on vowel recognition was not surprising since those channels cover the $F1-F2$ frequency range.

The mean percent-correct scores for the two-hole conditions are shown in Fig. 3. The mean scores dropped significantly when a second hole was introduced in the spectrum. The baseline score for consonant recognition dropped from 89.06% to an average (across all conditions) of 69.6%. A one-way ANOVA showed a significant main effect on consonant recognition when two holes were introduced in the spectrum $[F(15,120)=6.4, p<0.0005]$. *Post hoc* Tukey tests showed that several channel pair combinations significantly affected consonant recognition: (1,2), (1,4), (1,6) (2,3), (2,6), (3,4), (3,6), (4,5), (4,6), and (5,6). The drop in performance when both channels 1 and 2 were removed was due to the low scores obtained for nasal (/m/,/n/) and labial-stop consonant (/b/,/p/) recognition. Overall, we found that the scores obtained with channel pairs that included channels 4, 5, or 6 were significantly lower than the baseline score ($p<0.05$). This seems to be consistent with the single-hole conditions, and reinforces the message that channels 4, 5, and 6 are very important for consonant recognition.

The consonant confusion matrices were analyzed in terms of percent information transmitted. The feature analysis is shown in Fig. 4. A one-way ANOVA with repeated measures showed a significant effect $[F(15,120)=5.5, p<0.0005]$ for the feature "manner," a significant effect for the feature "voicing" $[F(15,120)=3.5, p<0.0005]$, and a significant effect $[F(15,120)=6.7, p<0.0005]$ for the feature "place." *Post hoc* Tukey tests showed that the manner score obtained with channel pair (1,2) removed was significantly lower ($p<0.0005$) than the baseline score. The voicing scores obtained with channel pairs (1,2) and (1,4) removed were significantly lower ($p<0.005$) than the baseline score. All place scores were significantly ($p<0.005$) lower than the baseline score.

For vowel recognition, a one-way ANOVA showed a significant main effect $[F(15,75)=6.9, p<0.0005]$ when

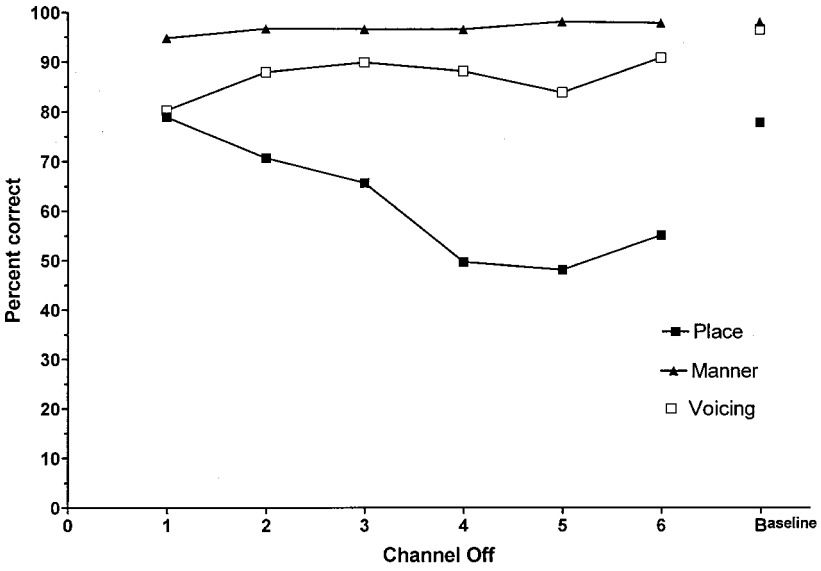


FIG. 2. Percent information transmitted for the features place, manner, and voicing as a function of the location of the spectral hole.

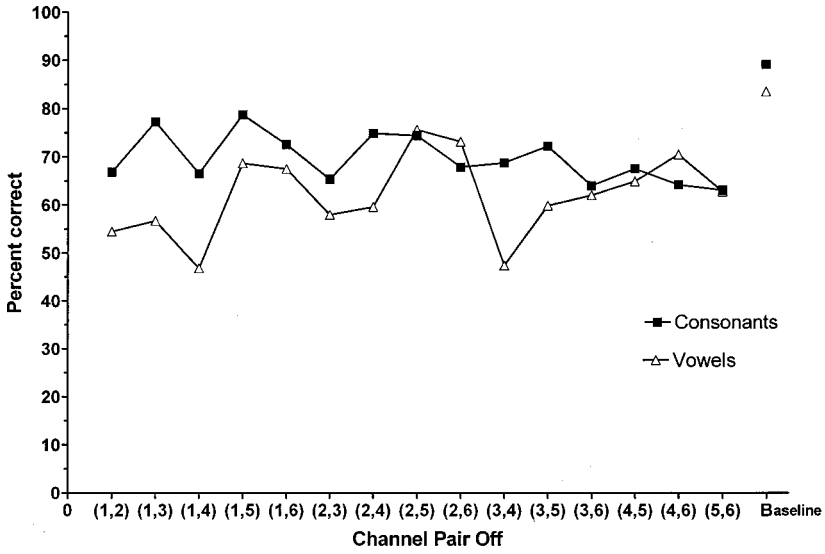


FIG. 3. Mean percent scores for vowel and consonant recognition as a function of the location of the pair of spectral holes. The holes were introduced at frequencies centered at the channel pairs indicated. In condition (1,4), for instance, channels 1 and 4 were removed from the spectrum. In the baseline condition, all channels were present.

two holes were introduced in the spectrum. *Post hoc* Tukey tests showed that several channel pair combinations were significantly affected on vowel recognition: (1,2), (1,3), (1,4),(2,3), (2,4), (3,4), (3,5), and (5,6). The drop in vowel performance when both channels 5 and 6 were removed was due to the low scores obtained for the vowels in ''heed,'' ''hid,'' and ''hayed.'' *Post hoc* Tukey tests showed that the scores obtained with channel pairs that included channel 1, 3, or 4 were significantly lower from the baseline scores ($p < 0.05$), consistent with the outcome in the single-hole conditions. More specifically, the lowest scores on vowel recognition were obtained with channel pairs (1,2), (1,3), (1,4), and (3,4).

C. Discussion

The above results suggest that vowel and consonant recognition suffer when holes are introduced in the spectrum. The degree of degradation in recognition performance as well as effect of the location of the spectral holes was different for vowels and consonants.

1. Effect of location of spectral holes

For vowels, statistical analysis showed a significant drop in performance when either channels 1, 3, or 4, centered at 393, 1037, and 1685 Hz, respectively, were removed. It is safe to assume that channel 1 codes $F1$ information, and channels 3 and 4 code $F2$ information for most vowels (Table I). Channel 3 may also code $F1$ information for some female vowels (i.e., vowels in ''hod'' and ''hud''). Depending on how high the $F2$ frequency is for some speakers, channel 5 (and, indirectly, channel 6) may also be important for the recognition of some vowels. Channel 5 may, for instance, code $F2$ information for some vowels (i.e., heed, hid, hayed) produced by female speakers or children who generally have a high $F2$ frequency. Indeed, close examination of the individual vowel's scores indicated that the identification of the female vowels in ''heed,'' ''hid,'' and ''hayed'' dropped significantly when both channels 5 and 6 were removed.

It is interesting to note that vowel recognition performance was not significantly affected when channel 2 (centered at 639 Hz) was removed. Channel 2 most likely codes

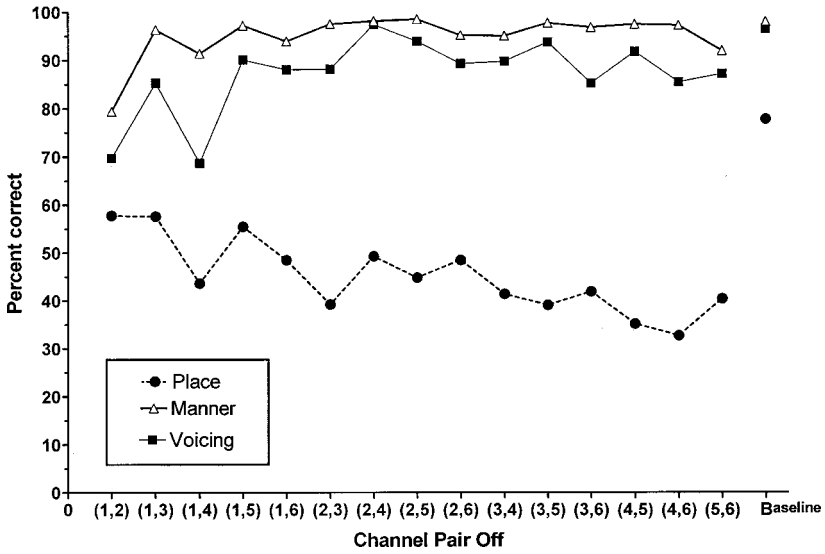


FIG. 4. Percent information transmitted for the features place, manner, and voicing as a function of the location of the pair of frequency bands removed.

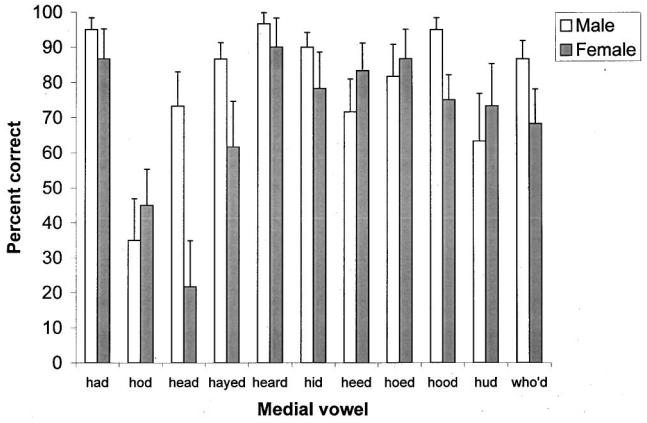


FIG. 5. Mean percent scores on individual vowel recognition for the condition in which channel 2 was removed from the spectrum ($n=20$). The dark and white bars give the scores obtained with vowels produced by female and male speakers, respectively. Error bars indicate standard errors of the mean.

$F1$ information either together with channel 1 or alone. Information about $F1$ is captured by channel 2 alone when the first formant frequency of the vowel falls near the center frequency of channel 2. In that case, a peak in the channel spectrum is observed at channel 2, and consequently removing channel 2 will significantly reduced performance. This is demonstrated in Fig. 5, which shows the listeners' individual vowel performance when channel 2 was removed. Vowels /ɛ/ and /a/ were the only vowels that were significantly affected because the $F1$ frequency of those vowels happened to be near the center frequency of channel 2. For the remaining vowels, however, as evident from Fig. 5, listeners seemed to infer $F1$ information from channel 1 when they did not have access to channel 2 information. This suggests that having a rough estimate of $F1$ is sufficient for the recognition of most vowels. That was not the case with $F2$, since removing either channels 3 or 4 affected vowel recognition.

For consonants, statistical analysis showed a significant drop in performance when either channels 4, 5, or 6 were removed. This outcome is consistent with the conventional view that high-frequency cues are important for recognition of place. So, the drop in consonant recognition performance was primarily due to a reduction in information transmitted for place (Fig. 2).

Removing any of the low-frequency channels (1–4) affected vowel recognition, and removing any of the high-frequency channels (4–6) affected consonant recognition. Interestingly enough, channel 4, which had a center frequency of 1685 Hz, was found to be important for both vowel and consonant recognition. The frequency (1685 Hz) corresponding to channel 4 is close to the well-known crossover frequency[1] estimated in articulation index studies. Depending on the speech material used, the crossover frequency was found in articulation index studies to be in the range of 1550 to 1900 Hz (Studebaker *et al.*, 1987; Hirsh *et al.*, 1954; French and Steinberg, 1947).

Overall, with the exception of channel 2 which did not significantly affect either vowel or consonant recognition, removing single channels caused a modest, but significant reduction in performance in vowel and consonant recognition. Consonant recognition was less affected than vowel recognition. It should be noted that the drop in performance, although statistically significant, was not dramatic for either consonant or vowel recognition. Even in the worst-case conditions, vowel and consonant recognition remained about 70% correct. So, relatively high vowel and consonant recognition performance can be maintained even with a single hole in the spectrum. This outcome is consistent with the data reported by Shannon *et al.* (2001) with cochlear-implant listeners. Shannon *et al.* artificially created single holes by turning off a number of (apical, middle, or basal) electrodes in CI listeners who were fitted with the 22-electrode Nucleus device. Holes were created that were 2–8 electrodes wide corresponding to 1.5–6.0-mm width. High vowel and consonant recognition was maintained even when as many as four electrodes were turned off either in the low-, middle-, or high-frequency regions.

2. Effect of size and pattern of spectral holes

In 5 out of the 15 conditions tested, the size of the hole or equivalently, the width of the notch in the spectrum, doubled, since in these conditions [i.e., channel pairs (1,2), (2,3), (3,4), (4,5), and (5,6)] the channels that were removed were adjacent to each other. This caused a large drop in vowel recognition performance, and only a moderate drop in performance for consonant recognition. Vowel recognition dropped in some cases to as low as 47% correct. The lowest performance occurred when $F1$ information was missing [e.g., pair (1,2)], when $F2$ information was missing [pairs (2,3), (3,4)], or when both $F1$ and $F2$ information was missing [pair (1,4)].

Consonant recognition was only mildly affected by the location of the pairs of frequency bands removed. The decrease in consonant identification was due primarily to the loss of place information (Fig. 4). The manner and voicing features were significantly affected only when information about $F1$ was missing. Overall, consonant identification remained robust and hovered around 70% for most conditions. Even when the middle- and high-frequency bands were absent, consonant recognition remained around 70% correct. This outcome is consistent with the data reported by Lippmann (1996), who evaluated consonant recognition by presenting a low-pass band below 800 Hz and a high-pass frequency band with cutoff frequency varying from 3.15 to 8 kHz. He observed a high score of 91% correct when the high-pass cutoff frequency was 3.15 kHz. This corresponded to the case where channels 4 and 5 were removed in our study. The score for that condition was 67.5% correct. The difference in scores between our study and Lippmann's can be attributed to the fact that our listeners only had access to four channels (two channels were removed) of frequency information. Similar findings were reported by Dorman *et al.* (1989) with CI listeners fitted with a four-channel processor. No significant difference was found between the consonant identification score obtained with only channels 1 (low frequency) and 4 (high frequency) activated and the score obtained with all four channels activated. Our study extended Dorman's and Lippmann's findings to show that high consonant recognition can be maintained even in the absence of

not only middle frequencies but also low-high, low-middle, low-high, and middle-high frequency information.

Overall, we can say that vowel recognition seems to be sensitive to the size and pattern of holes in the spectrum. This was not surprising, since it is known that listeners rely primarily on spectral cues to identify vowels. In contrast, listeners make use of both temporal-envelope cues and spectral cues to identify consonants. In the absence of sufficient spectral cues, listeners probably rely more on temporal cues to identify consonants. As shown in this experiment (Fig. 4), these temporal cues did not seem to be affected by the frequency location of the pair of bands removed [except when channels (1,2) were removed]. We believe that is the reason that consonant recognition remained relatively high (~70% correct) even when two holes were introduced in the spectrum. The above results have certain implications for cochlear implants. The finding that the location and pattern of holes affects mostly vowel recognition suggests that in cochlear implants, neuron survival (responsible for the holes in the spectrum) ought to account for some of the variability in vowel recognition performance among CI listeners.

III. EXPERIMENT 2: FREQUENCY-IMPORTANCE FUNCTIONS

Several investigators have used the AI method to determine frequency-importance functions. The AI method uses a quantity between 0 and 1 to represent the proportion of speech information available in a specific frequency band to the listener. This information is then multiplied by a frequency-importance or ''weighting'' function, which is obtained using a rather time-consuming process of low-pass and high-pass filtering speech. The AI method assumes that the information contained in each band is independent of the information contained in other bands and does not take into account the fact that listeners may combine speech information from multiple disjoint bands. This was first demonstrated by Kryter (1962), who evaluated recognition of passband speech, and showed that the AI could not adequately predict intelligibility of passband speech. Similar findings were also reported by Grant and Braida (1991). Several methods were proposed in the literature to circumvent this shortcoming, including the correlation method by Doherty and Turner (1996) and a recent method based on statistical decision theory by Musch and Buus (2001). In this experiment, we use the data from experiment 1 to derive a frequency-importance function based on a least-squares approach. Unlike the AI method, the proposed least-squares approach makes use of the listener's scores on perception of vowels and consonants composed of disjoint frequency bands.

A. Least-squares approach

Our approach to obtain the importance of each frequency band follows the method proposed by Ahumada and Lovell (1970). We used the results from experiment 1 to predict the importance or perceptual ''weight'' of each channel.

We calculated the weight w_i of each channel by predicting the responses of the subject as a linear combination of the strength of each channel, i.e.,

$$R_k = \sum_{i=1}^{6} w_i E_{ik}, \tag{1}$$

where R_k is the mean percent-correct score for condition k and E_{ik} is the strength of the ith channel corresponding to condition k. The strength of each channel is a binary value that can be either 0 or 1 depending on whether the channel is off or on, respectively. The value of k ranges from 1 to 22 spanning all channel combinations (Table III). Forming the prediction error e_k as

$$e_k = R_k - \sum_{i=1}^{6} w_i E_{ik}, \tag{2}$$

we can estimate the channel weights by minimizing the sum of all the squared errors with respect to w_i. Alternatively, Eq. (1) can be written in matrix form as

$$R = EW, \tag{3}$$

where R is a 22-dimensional vector containing the mean percent-correct scores for conditions 1 to 22, E is the data matrix (22×6) consisting of the strengths of each channel (Table III), and $W=[w_1, w_2, \ldots, w_6]$ is a 6-dimensional vector consisting of the desired channel weights.

The above set of equations represents an overdeterministic system of equations since we have 6 unknowns (the channel weights) and 22 equations (one for each condition). We calculated the weights W by solving the matrix equation given by (3) using a least-squares approach

$$W = (E^T E)^{-1} E^T R. \tag{4}$$

After obtaining the solution from Eq. (4), we normalized the weights so that the sum of all the weights was equal to 1.

B. Results and discussion

The relative weights of the various channels are shown in Fig. 6 for the vowel and consonant stimuli. As can be seen, the shape of the weighting function was different for vowels and consonants. For vowels, there was unequal weighting across the various channels, suggesting that each channel contributed differently in understanding these vowel tokens. Channels, 1, 3, and 4, centered at 393, 1037, and 1685 Hz, respectively, received the largest weight. This outcome was consistent with the listener's reduction in performance in experiment 1 when those channels were removed. Also, consistent with our data from experiment 1, channels 2, 5, and 6 received the lowest weight.

The weighting function for the consonants was relatively flat. This suggests that for consonant recognition all channels are equally important. This outcome is consistent with the data reported recently by Mehr *et al.* (2000). Mehr *et al.* estimated the frequency-importance function of nonsense syllables using the correlational method. Speech was divided into six frequency bands, and a randomly chosen level of filtered noise was added to each channel on each trial. Channels in which the signal-to-noise ratio was more highly cor-

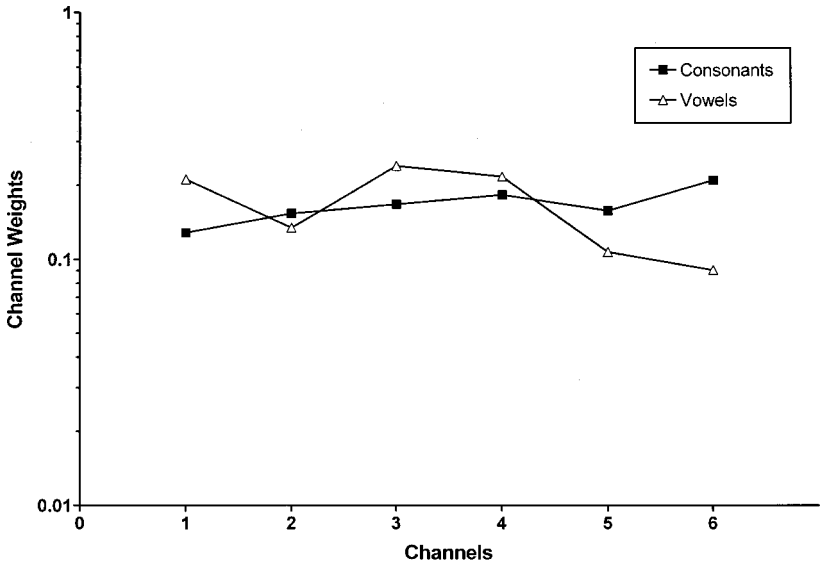


FIG. 6. Frequency-importance functions for vowels and consonants.

related with performance had a larger weight, and channels with smaller correlations had lower weights. Their results showed a flat weighting function for normal-hearing listeners. Unequal weighting functions, accompanied with a large variability among subjects, was noted for the CI listeners in their study.

The individual listener's weighting functions are given in Fig. 7 for vowel and consonant recognition. Weighting functions are given for 6 of the 20 subjects, 2 subjects with the highest vowel scores (Fig. 7, panels a and b), 2 with the middle vowel scores (panels c and d), and 2 with the lowest vowel scores (panels e and f). Most listeners had a relatively flat weighting function for consonants with a small variability. There was a larger variability among subjects in the

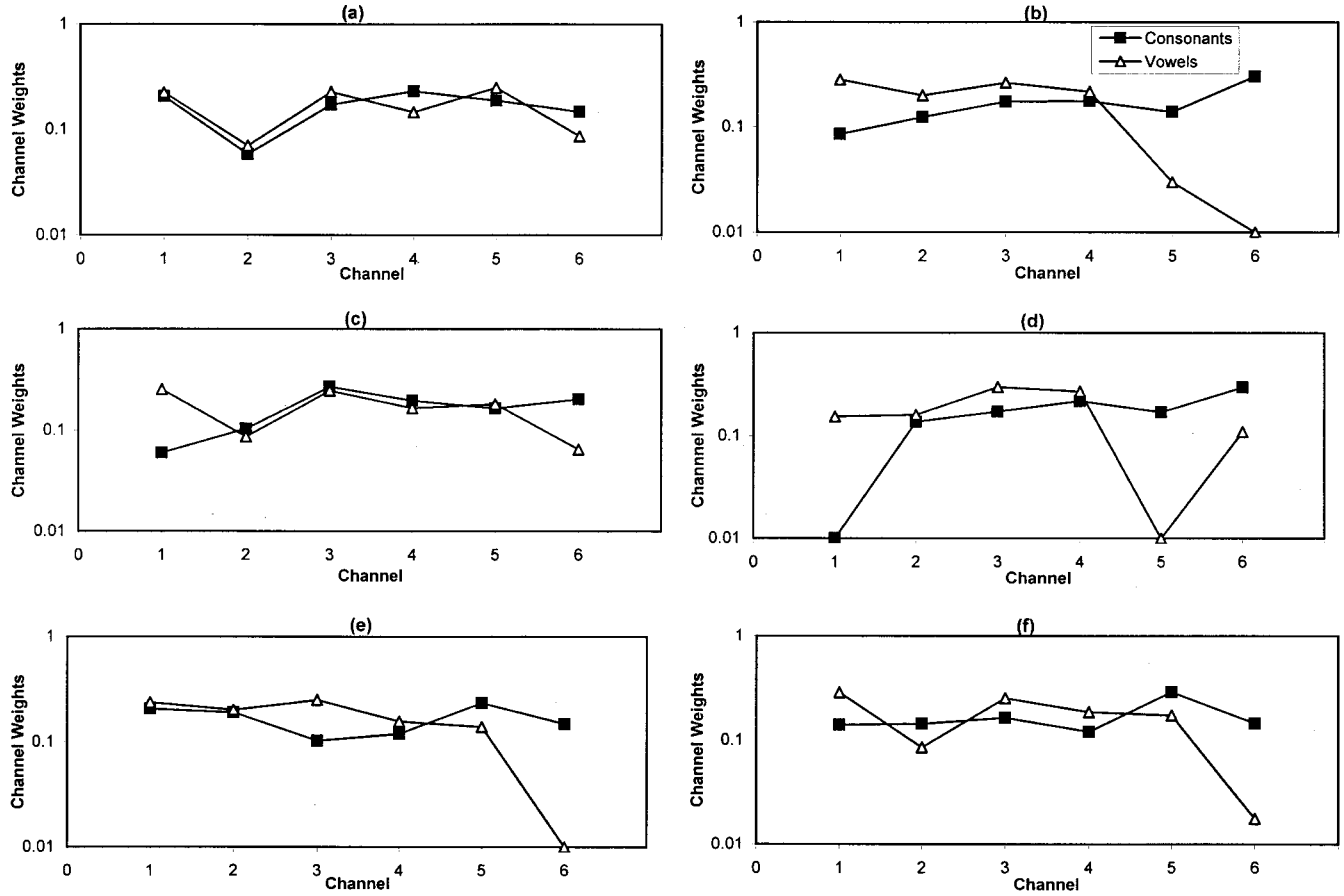


FIG. 7. Individual listener's frequency-importance functions for vowel and consonant recognition. Panels (a) and (b) show the frequency-importance functions for two subjects with the highest vowel scores, panels (c) and (d) show the functions for two subjects with middle scores, and panels (e) and (f) show the functions for two subjects with the lowest vowel scores.

shape of the weighting functions for vowels, suggesting that subjects used different listening strategies for vowel recognition.

The fact that the weighting functions for vowels and consonants were different suggests that subjects were using different listening strategies to identify vowels and consonants. For vowel identification, listeners rely primarily on spectral cues and therefore place more emphasis or more ''weight'' on the channels that code $F1$ and $F2$ information. For consonant identification, listeners rely on both temporal-envelope and spectral cues, which are distributed across all channels. Hence, all frequency bands contributed equally to consonant identification, at least for the filter spacing used in this study. The data from experiment 1 (Fig. 3) are consistent with this conclusion. The fact that consonant recognition remained relatively constant, around 70% correct, regardless of which pairs of channels were removed, clearly demonstrated that all channels contributed equally to consonant recognition. Had the listeners placed more emphasis on certain channels or pairs of channels, we would have seen a dramatic decrease in performance at those channel(s), as we did with the vowels. We suspect that, in general, the frequency-importance function must be dependent, among other factors, on the speech material and the frequency spacing used. Studebaker *et al.* (1987), for instance, showed that the shape of articulation index function and the crossover frequency depended on the speech material.

We did not vary the frequency spacing in this study, but rather used the logarithmic spacing typically implemented in current cochlear-implant processors (Loizou, 1998). According to the data obtained in this experiment, logarithmic spacing provided an equal amount of speech information in each frequency band for consonant identification. This outcome has important implications for cochlear implants. Logarithmic spacing would be desirable assuming that CI listeners are able to extract information from *all* their electrodes. As shown by many investigators (e.g., Fishman *et al.*, 1997; Dorman *et al.*, 2000; Zwolan *et al.*, 1997), that is not the case. This suggests that the frequency spacing should be customized for each CI subject in such a way that their resulting frequency-importance function has larger weights on the functional electrodes and smaller weights on the not-so-functional electrodes.

Despite the differences between the least-squares approach used in this study and the correlational method used by Mehr *et al.* (2000), we obtained a similar (almost identical) weighting function for nonsense syllables. The testing process involved in deriving the weighting functions is time consuming, and therefore both methods are impractical for clinical applications. Another drawback of the correlational method is that it is dependent on the number of trials used for testing. As many as 1200 trials were required in some cases to get significant raw correlations (Mehr *et al.*, 2000; Turner *et al.*, 1998). Our method is not largely dependent on the number of trials, but requires an adequate number of conditions. In our study, we needed to run a total of 22 conditions, which is considerably less than the 135 conditions needed for articulation index studies (e.g., Studebaker *et al.*, 1987) to estimate the frequency-importance function. In brief, the least-squares approach proposed in this study is another viable approach for obtaining frequency-importance functions.

IV. SUMMARY AND CONCLUSIONS

(i) When a single hole was introduced in the spectrum, vowel and consonant recognition decreased. The degree of degradation in performance depended on the location of the hole or, equivalently, the frequency band removed. For vowels, there was a significant drop in performance when either of the frequency bands, 1, 3, and 4 centered around 393, 1037, and 1685 Hz, respectively, were removed. For consonants, there was a modest, yet significant, drop in performance when either of the frequency bands 4, 5, and 6, centered around 1685, 2736, and 4444 Hz, respectively, were removed.

(ii) Vowel recognition was affected the most, with the lowest performance (60% correct) obtained when channel 3, responsible for coding $F2$ information, was removed. Consonant recognition remained relatively high at around 70% correct even when high-frequency channels were removed. Feature analysis indicated that the drop in consonant performance was primarily due to loss of place information. The manner and voicing features were not affected by the location of the hole in the spectrum.

(iii) When two holes were introduced in the spectrum, vowel recognition decreased even further, and consonant recognition remained constant around 70% correct (the same as in the single-hole condition).

(iv) Vowel recognition performance was dependent on the frequency location of the pairs of bands removed. In particular, removing pairs of bands that contained $F1$ and/or $F2$ information caused a significant drop in performance.

(v) In contrast, consonant recognition was only mildly affected by the location of the pair of frequency bands removed. Consonant recognition remained robust at 70% correct, even when the middle- and high-frequency speech information was missing. This outcome is consistent with Lippmann's (1996) findings that accurate consonant recognition can be maintained even when the middle frequencies in the spectrum are absent. Our study extended Lippmann's findings to show that high consonant recognition can be maintained even in the absence of disjoint frequency bands involving low-, high-, and/or middle-frequency information.

(vi) The shapes of the frequency-importance functions, derived in experiment 2 using a least-squares approach, were different for vowels and consonants. This is in agreement with the notion that different cues are used by listeners to identify consonants and vowels.

(vii) For vowels, there was unequal weighting across the various channels. Channels 1, 3, and 4 received the largest weight. The frequency-importance function for consonants was relatively flat, suggesting that all

channels contributed equally to consonant identification, at least for the logarithmic filter spacing used in this study. This has important implications for cochlear implants. For CI listeners who are not able to extract useful information from *all* their electrodes, the logarithmic filter spacing might not be the optimal filter spacing.

ACKNOWLEDGMENTS

This research was supported by Grant No. R01 DC03421 from the National Institute of Deafness and other Communication Disorders, NIH. This project was the basis for the Master's thesis of the first author (K.K.) in the Department of Electrical Engineering at the University of Texas—Dallas.

[1]The crossover frequency is the frequency which divides the frequency spectrum into two parts contributing equally to intelligibility. Estimated in articulation index studies, it is the frequency at which the performance with high-passed speech and the performance with low-passed speech is the same.

Ahumad, Jr., A., and Lovell, J. (**1990**). ''Stimulus features in signal detection,'' J. Acoust. Soc. Am. **49**, 1751–1756.

Breeuwer, M., and Plomp, R. (**1984**). ''Speechreading supplemented with frequency-selective sound-pressure information,'' J. Acoust. Soc. Am. **76**, 686–691.

Doherty, K. A., and Turner, C. W. (**1996**). ''Use of a correlational method to estimate a listener's weighting function for speech,'' J. Acoust. Soc. Am. **100**, 3769–3773.

Dorman, M., Dankowski, K., McCandless, G., and Smith, L. (**1989**). ''Consonant recognition as a function of the number of channels of stimulation by patients who use the Symbion cochlear implant,'' Ear Hear. **10**(5), 288–291.

Dorman, M., Loizou, P., Fitzke, J., and Tu, Z. (**2000**). ''The recognition of NU-6 words by cochlear implant patients and by normal-hearing subjects listening to NU-6 words processed in the manner of CIS and SPEAK strategies,'' Ann. Otol. Rhinol. Laryngol. Suppl. **109**(12), Suppl. 185, 64–66.

Fishman, K. E., Shannon R. V., and Slattery, W. H. (**1997**). ''Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,'' J. Speech Hear. Res. **40**(5), 1201–1205.

French, N. R., and Steinberg, J. C. (**1947**). ''Factors governing the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **18**, 90–119.

Grant, K., and Braida, L. (**1991**). ''Evaluating the articulation index for auditory-visual input,'' J. Acoust. Soc. Am. **89**, 2952–2960.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Hirsh, I., Reynolds, E., and Joseph, M. (**1954**). ''Intelligibility of different speech materials,'' J. Acoust. Soc. Am. **26**, 530–538.

Kryter, K. (**1962**). ''Validation of the articulation index,'' J. Acoust. Soc. Am. **34**, 1698–1702.

Lippmann, R. P. (**1996**). ''Accurate consonant perception without mid-frequency speech energy,'' IEEE Trans. Speech Audio Process. **4**, No. 1, 66–69.

Loizou, P. (**1998**). ''Mimicking the human ear: An overview of signal processing techniques for converting sound to electrical signals in cochlear implants,'' IEEE Signal Process. Mag. **15**(5), 101–130.

Loizou, P., Dorman, M., and Tu, Z. (**1999**). ''On the number of channels needed to understand speech,'' J. Acoust. Soc. Am. **106**, 2097–2103.

Mehr, M. A., Turner, C. W., and Parkinson, A. (**2000**). ''Channel weights for speech recognition in cochlear implant users,'' J. Acoust. Soc. Am. **109**, 359–366.

Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.

Musch, H., and Buus, S. (**2001**). ''Using statistical decision theory to predict intelligibility. I. Model structure,'' J. Acoust. Soc. Am. **109**, 2896–2909.

Pollack, I. (**1948**). ''Effects of high-pass and low-pass filtering on the intelligibility of speech in noise,'' J. Acoust. Soc. Am. **20**, 259–266.

Riener, K., Warren, R., and Bashford, J. (**1992**). ''Novel findings concerning intelligibility of bandpass speech,'' J. Acoust. Soc. Am. **91**, 2339.

Shannon, R., Galvin, J., and Baskent, D. (**2001**). ''Holes in hearing,'' J. Assoc. Res. Otolaryngol. **3**, 185–199.

Stickney, G., and Assmann, P. (**2001**). ''Acoustic and linguistic factors in the perception on bandpass-filtered speech,'' J. Acoust. Soc. Am. **109**, 1157–1165.

Studebaker, G., Pavlovic, C., and Sherbecoe, R. (**1987**). ''A frequency importance function for continuous discourse,'' J. Acoust. Soc. Am. **81**, 1130–1138.

Turner, C., Kwon, B., Tanaka, C., Knapp, J., and Doherty, K. (**1998**). ''Frequency importance functions for broadband speech as estimated by the correlational method,'' J. Acoust. Soc. Am. **104**, 1580–1585.

Tyler, R., Preece, J., and Lowder, M. (**1987**). The Iowa audiovisual speech perception laser videodisc. *Laser Videodisc and Laboratory Report*, Dept. of Otolaryngology, Head and Neck Surgery, University of Iowa Hospital and Clinics, Iowa City.

Warren, R. M., Riener, K. R., Bashford, Jr., J. A., and Brubaker, B. S. (**1995**). ''Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits,'' Percept. Psychophys. **57**, 175–182.

Zwolan, T., Collins, L., and Wakefield, G. (**1997**). ''Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects,'' J. Acoust. Soc. Am. **102**, 3673–3685.

Factors underlying the speech-recognition performance of elderly hearing-aid wearers

Larry E. Humes[a)]
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 22 March 2002; revised 14 June 2002; accepted 18 June 2002)

This paper reports the aided and unaided speech-recognition scores from a group of 171 elderly hearing-aid wearers. All hearing-aid wearers were fit with identical instruments (linear Class-D amplifiers with output-limiting compression) and evaluated with a standard protocol. In addition to including multiple measures of speech recognition, an extensive set of physiological and perceptual measures of auditory function, as well as general measures of cognitive function, were completed prior to the hearing-aid fitting. Comparison of the results from this study to available norms suggested that this group of participants was fairly typical or representative for their hearing loss and age. Approaches to the prediction of general speech-recognition performance that were examined included methods based on an acoustical index, the Speech Intelligibility Index (SII), and others based on linear-regression statistical analysis. The latter approach proved to be the most successful, accounting for about two-thirds of the variance in speech-recognition performance, with the primary predictive factors being measures of hearing loss and cognitive function. [DOI: 10.1121/1.1499132]

PACS numbers: 43.71.Ky, 43.66.Ts, 43.71.Lz, 43.66.Sr [CWT]

I. INTRODUCTION

Attempts to predict the speech-recognition performance of normal-hearing and hearing-impaired listeners for a variety of acoustical conditions have had a long history. One basic approach that has been pursued actively over the past half century, initially for normal-hearing individuals and more recently for hearing-impaired persons, is based on the original Articulation Index (AI) model of Fletcher and colleagues (French and Steinberg, 1947; Fletcher and Galt, 1950; Fletcher, 1953). The AI model has evolved over the past 50 years, but the basic notion underlying the model remains the same. Specifically, there is a well-defined, monotonic relationship between a weighted sum of band-specific speech-to-noise ratios, the AI, and average speech-recognition performance measured for those same acoustical conditions. Over the years, there have been variations in approaches regarding the number and width of the analysis bands (e.g., 1/3-octave, octave), the importance function or weighting factors assigned to each band, the range of speech-to-noise ratios to be considered in each band (e.g., +15 to −15 dB vs +12 to −18 dB), and the method used to establish the speech-to-noise ratio in each band. The effects of many of these factors on the resulting AI calculation have been reviewed previously by Humes *et al.* (1986). Given the large number of variations in the basic AI model that had emerged in the 1980s and 1990s, the American standard for the calculation of this index was revised in 1997 (ANSI S3.5-1997). In this national standard, the acoustical index was also renamed the Speech Intelligibility Index (SII).

When the SII is applied to the aided and unaided speech-recognition performance of elderly hearing-impaired listeners, several correction factors, either incorporated in the standard or available in the research literature, have been recommended. If high speech levels are used, either aided or unaided, then a level-distortion factor incorporated in the standard will reduce the resulting SII. It should be noted, however, that this factor would also apply to SII calculations for normal listeners evaluated at the same high speech levels. While such a correction factor may seldom be needed for normal listeners under ''typical'' listening conditions, it will frequently be needed for hearing-impaired listeners wearing hearing aids under those same ''typical'' listening conditions. In addition to the level-distortion adjustment to the SII, correction factors associated with hearing loss desensitization (e.g., Pavlovic, Studebaker, and Sherbecoe, 1986; Studebaker, Gray, and Branch, 1999) and aging (Studebaker *et al.*, 1997; Sherbecoe and Studebaker, 2002b) have also been advocated for applications of the SII to elderly hearing-impaired listeners. For the most part, these correction factors are derived from group data and attempt to bring the group data for impaired listeners more in line with the performance-vs-SII transfer function for average normal listeners.

A statistical linear-regression model of aided and unaided speech-recognition performance represents an entirely different approach to the prediction of speech-recognition performance in elderly hearing-aid wearers. Rather than attempting to describe the performance of the ''typical'' listener from consideration of the acoustical conditions, the regression model attempts to predict the performance of individual listeners from available predictor variables. Such variables might include those inherent in SII calculations, such as the hearing thresholds of the listeners and the gain of the hearing aid, as well as those included in various SII correction factors, such as the age of the listener. The regression model, however, can also include additional factors that might contribute to performance, such as measures of audi-

a)Electronic mail: humes@indiana.edu

tory perceptual processing or cognitive function. Although there have been a large number of studies pursuing a regression approach to this problem over the past 10–15 years, especially regarding *unaided* speech-recognition performance, perhaps the series of studies by van Rooij and colleagues, Humes and colleagues, and Divenyi and Haupt are most representative (Divenyi and Haupt, 1997a, 1997b, 1997c; Humes and Roberts, 1990; Humes and Christopherson, 1991; Humes *et al.*, 1994; van Rooij, Plomp, and Orlebeke, 1989; van Rooij and Plomp, 1990, 1992). In all of these studies, good regression fits were obtained with large proportions of the variance (typically 67%–90%) in unaided speech-recognition performance accounted for by various predictor variables. Although all studies identified cognitive or auditory-processing factors that made secondary contributions to the best-fitting regression model, audibility or hearing loss was universally found to be the single best predictor of *unaided* speech-recognition performance.

Perhaps the predominant role of audibility in such regression models is not surprising given that one is predicting *unaided* speech-recognition performance and, in most studies, speech has been presented at or near conversational levels. A notable exception is the study by Humes *et al.* (1994) in which speech was also presented at high sound levels (90 dB SPL) and, in one condition, was also spectrally shaped to mimic the frequency-gain characteristics generated by a common clinical hearing-aid-fitting protocol. Although audibility still accounted for more variance in speech-recognition performance than any other factor, this appeared to be less the case for high speech levels in noise backgrounds.

Nonetheless, it is very likely that audibility will also play a role in *aided* speech-recognition, at least when listeners are fit with linear hearing aids having frequency-gain characteristics matched to the most common prescriptive formulas employed clinically [see Humes (1991, 1996) and Humes and Halling (1993) for reviews]. That is, prescriptive approaches for such hearing aids are generally not designed to maximize the audibility of speech, but incorporate a tradeoff between the optimization of speech audibility and the maintenance of a pleasant sound quality so as not to lead to the rejection of the hearing aids. In practice, this typically means a reduction in the high-frequency gain of linear hearing aids relative to that which would optimize speech understanding (Humes, 1986, 1991; Humes and Halling, 1993). As a result, the high frequencies of speech are often still inaudible for many wearers while using hearing aids.

The present study evaluates both acoustical SII-based and statistical regression-based approaches to the prediction of unaided and aided speech-recognition performance in the same group of elderly hearing-aid wearers. In addition to obtaining multiple measures of speech-recognition performance in both aided and unaided listening conditions, all participants completed an extensive battery of audiological, auditory-processing, and cognitive measures prior to being fit with hearing aids. These measures were used in subsequent regression analyses, but were also examined in detail to establish that the present group of elderly hearing-aid wearers were ''typical'' or representative of elderly hearing-impaired listeners in general. Since the focus was on individual differences among wearers, the devices used and the study protocol were identical for all participants. Additional procedural details are provided in Sec. II.

II. METHOD

A. Participants

The participants in this study were recruited for a large-scale, longitudinal study of hearing-aid outcome measures via newspaper ads, flyers posted in the community, printed announcements in church/synagogue bulletins, and word of mouth. All participants enrolled in the study met the following selection criteria: (1) age between 60 and 89 years; (2) hearing loss that was flat or gently sloping (from 250 to 4000 Hz, no inter-octave change in hearing thresholds of more than 20 dB); (3) hearing loss that was of sensorineural origin (normal tympanometry and air-bone gaps no greater than 10 dB at three or more frequencies); (4) hearing loss that was bilaterally symmetrical (interaural difference within 30 dB at all octave and half-octave intervals from 250 to 4000 Hz); (5) pure-tone thresholds within the following ranges at frequencies of 250, 500, 1000, 1500, 2000, 3000, 4000, and 6000 Hz, respectively: 5-85, 5-85, 10-90, 20-95, 25-95, 30-120, 30-120, and 30-120 dB HL (ANSI, 1996); (6) no known medical or surgically treatable ear-related condition; (7) no known fluctuating or rapidly progressing hearing loss; (8) no cognitive, medical, or language-based conditions which may have limited the participant's ability to complete the procedures used in the longitudinal study of outcome measures; (9) no use of medications that could affect hearing or cognition; and (10) completion of a signed medical clearance form, or waiver of such by the participant, and a signed informed consent form.

During the first few years of this study, over 400 individuals had been recruited as potential participants and 205 met the selection criteria and were enrolled in the large-scale, longitudinal study. Of this group, 198 represented all the participants to have completed a large battery of tests prior to being fit with amplification. A total of 171 of these 198 individuals remained in the study at least to the one-month postfit interval, at which time the aided speech-recognition measures were completed. These 171 individuals represent the participants in this study. They ranged in age from 60 to 87 years ($M = 73.0$ years) and 54 (31.6%) were female. A total of 104 individuals (60.8%) were new hearing-aid users. Of the 67 experienced hearing aid users, 36 (53.7%) wore binaural amplification and the rest wore monaural amplification.

B. Prefit test materials and procedures

1. Audiological evaluation

Following completion of a detailed case history, a comprehensive audiological evaluation was conducted for all participants in this study. All audiologic measurements were obtained using ER-3A insert earphones and all equipment was calibrated in accordance with ANSI S3.6-1996. This evaluation included: (1) air-conduction pure-tone audiometry at 250, 500, 1000, 2000, 3000, 4000, 6000 and 8000 Hz, with additional threshold measurements at 750 and 1500 Hz

whenever thresholds at the adjacent octave test frequencies differed by 20 dB or more; (2) bone-conduction pure-tone audiometry at octave intervals from 250 to 6000 Hz; (3) immittance measurements, including tympanometry, acoustic reflex thresholds, and acoustic reflex decay; (4) speech-recognition threshold (SRT) for CID W-1 spondaic words presented via monitored live voice; (5) 50-item word-recognition scores (Auditec recordings of NU-6 materials; Tillman and Carhart, 1966) at either 40 dB SL or maximum audiometer output, whichever corresponded to a lower sound pressure level, in quiet and in white noise at a +12 dB signal-to-noise ratio (SNR); and (6) LDL measurements using the scaling categories and instructions described by Hawkins *et al.* (1987), an ascending approach with 5 dB step size, and pure-tone frequencies of 500, 1000, 2000, 3000, and 4000 Hz. For the air-conduction pure-tone thresholds, measures of average hearing loss were also calculated. These included the pure-tone average (PTA) based on thresholds at 500, 1000, and 2000 Hz and the high-frequency pure-tone average (HFPTA) based on thresholds at 1000, 2000, and 4000 Hz.

In addition to these audiological measures, those participants who, based on the preceding audiological evaluation, met the selection criteria for the study, were invited to enroll in the longitudinal study of hearing-aid outcome. For the eligible participants who elected to enroll in the longitudinal study, a series of additional measures were obtained across four more test sessions. As much as possible, similar types of measures were grouped into specific test sessions. The order of testing was identical for all of these participants and proceeded as follows.

2. Additional audiological testing

Tests included in this session were: (1) distortion-product otoacoustic emissions (DPOAEs) collected with the ILO92 system, L1=L2=70 dB SPL, f2/f1=1.224, and ''DP-Gram'' operation mode with f1 sweeps from 818 to 5200 Hz with the SNR in dB at the 2f1−f2 frequency recorded and a low-, mid-, and high-frequency band average computed (f2 ranges of 1001–1587, 2002–3174, and 4004–6348 Hz, respectively); (2) auditory brainstem responses (ABRs) measured for each ear, using a Bio-Logic Model 54 system, for 2000 rarefaction click stimuli presented at a level of 90 dB nHL and at a rate of 11.1 clicks per second, with wave-V latency from two repeatable responses recorded; and (3) performance-intensity functions for each ear for NU-6 words with 25 monosyllabic words presented at several levels, beginning 40 dB above SRT and progressing, in 10 dB steps to a maximum of 100 dB HL, with the maximum score (PB-max) and the score at the maximum possible presentation level (PB-maxHL) recorded. A measure of PI-PB rollover was calculated by subtracting PB-maxHL from PB-max.

3. Auditory processing

Auditory-processing capability at high sound levels represented another area assessed in each participant prior to fitting of the hearing aid. There were two primary tools used to assess auditory-processing capabilities and in both cases the stimulus presentation level was 90 dB SPL. The first was the Test of Basic Auditory Capabilities (TBAC) developed by Watson and colleagues (Watson, 1987). This battery of auditory-discrimination tests has been demonstrated to be reasonably reliable in a small sample of elderly hearing-impaired listeners tested at moderate presentation levels (Christopherson and Humes, 1992) and has been used successfully with elderly hearing-impaired listeners (Humes and Christopherson, 1991; Christopherson and Humes, 1992). The studies with hearing-impaired listeners have found that the TBAC provides measures of auditory-discrimination ability that are unaffected by the presence of peripheral hearing loss when administered at sensation levels greater than 30 dB. In the present project, the TBAC was administered at a level of 90 dB SPL. Based on prior principal-component analyses of the application of the TBAC to hearing-impaired listeners (Humes *et al.*, 1994), the following three tests were administered sequentially: (1) duration discrimination for a 1000 Hz pure tone; (2) temporal order for tones, which is a temporal-order discrimination task for a four-tone sequence spectrally centered at 1000 Hz; and (3) syllable sequence test, which measures temporal-order discrimination performance for consonant-vowel syllables (/fa,ta,ka,pa/) as a function of syllable duration. All TBAC tests were administered diotically.

In addition, three tests of auditory perceptual processing were selected from the Veterans Administration Compact Disk for Auditory Perceptual Assessment (VACD; Noffsinger, Wilson and Musiek, 1994). This test had been evaluated previously for use with a similar group of participants (Humes, Coughlin, and Talley, 1996). Based on the findings of Humes *et al.* (1996), the following three measures were selected for use in this study: (1) dichotic consonant-vowel syllable identification for syllables delivered with 90 ms interaural onset disparity (Noffsinger, Martinez, and Wilson, 1994) with the score based on 30 dichotic presentations (60 syllables); (2) the pitch-pattern test, which assesses temporal-order identification for three-tone sequences (Musiek, 1994), with the score based on 60 trials to the left ear; and (3) recognition of NU-6 monosyllabic words that have been 45% time-compressed (Wilson *et al.*, 1994), with each score based on 50 words presented to the right ear.

For all the measures of auditory processing, stimuli were presented from either a digital audio tape (TBAC) or CD (VACD), through an attenuator and amplifier, to ER-3A insert earphones. All sound pressure levels specified for stimulus presentation level are referenced to an HA-2 2 cm^3 coupler (Frank and Richards, 1991).

4. Cognitive assessment

Participants in this study completed the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981). Eleven standard scale scores and three measures of intelligence emerged from this testing. In addition, since the WAIS-R does not provide age-related norms beyond an age of 74 years and absolute measures of cognitive function were desired, rather than relative measures within age-matched peer groups, raw scores on the 11 subscales of the WAIS-R were also recorded. Testing was performed by a graduate

student in educational psychology with experience in test administration and scoring. In addition, a hardwired assistive listening device was made available for use during WAIS testing in the event any of the participants had trouble understanding the instructions and explanations provided for various measures by the examiner.

The foregoing audiological, auditory-processing, and cognitive measures were completed over a period of two to four weeks prior to fitting the participant with amplification. Each session required 90–120 min for completion.

C. Hearing aids and fitting procedures

Once the prefit testing was completed, the participants returned for the initial fitting of their hearing aids. Procedural details presented here regarding this portion of the study can also be found in Humes *et al.* (2001), a study that focused on the analysis of a battery of hearing-aid outcome measures for the same study sample. Based on the previously obtained audiologic information and using the Hearing Aid Selection Program (HASP, Version 2.07) fitting software, produced and distributed by the National Acoustics Laboratories (NAL), NAL-R targets, including corrections to targets for severe or profound hearing loss (NAL-RP; Byrne *et al.*, 1990), were generated for each ear and the corresponding circuit was selected and ordered. The HASP software returns a variety of targets, including real-ear insertion gain and full-on coupler gain. The latter was adjusted downward by the 10 dB reserve gain incorporated into the HASP software to create recommended coupler-gain values.

All hearing aids made use of linear circuits with output-limiting compression and Class D amplifiers. All instruments were full-shell, in-the-ear (ITE) devices and included a telecoil switch on one instrument (determined by the wearer's preference). Active tone (low-cut only) and output-limiting controls, adjustable select-a-vent venting, and wax guards were included on all devices. The volume-control wheels were marked by the manufacturer with a small white dot at the perimeter to provide a visual reference for its position and adjustment. All hearing aids were made by the same manufacturer.

Using real-ear insertion gain targets for the NAL-R prescription formula incorporated in the HASP fitting software, the clinician adjusted the settings of the controls and vent to achieve the closest match possible to target gain for a 60 dB SPL swept pure-tone signal using either Frye 6500 or Audioscan real-ear measurement equipment. Matching criteria were + or − 10 dB from 250 to 2000 Hz and + or − 15 dB at 3000 and 4000 Hz. If a match could not be obtained using these fairly broad matching criteria, the participant was not permitted to continue in the study. In fact, no participants were eliminated from the study for this reason and the quality of the matches to target gain were considerably better than the broad exclusion criteria cited previously, as will be seen in the following. Previous reports from this project (Humes *et al.*, 2000; Humes *et al.*, in press) on 55–134 participants followed for a full year demonstrated that: (1) users consistently preferred a gain setting that was typically 6–9 dB below that prescribed; (2) the experimenters were able to return the volume control for all aided measurements to a consistent position that matched closely the prescribed target gain for subsequent aided testing; and (3) problems, such as dead batteries, cerumen blockage of the receiver port, etc., were relatively uncommon. For more details regarding the gain measurements and adjustment procedures, the reader is referred to Humes *et al.* (2000).

Following the delivery of the hearing aids and their initial fitting by the clinician, a hearing-aid orientation was conducted. Participants were instructed to use their hearing aids at least four hours per day and to begin use in easier listening conditions (quiet, one-on-one conversation, etc.) when possible.

D. Speech-recognition measures

Each participant returned two weeks later for a follow-up session. At the beginning of this session, gain measurements were again made in the coupler to evaluate the instruments and the aids were removed, inspected, and subsequently adjusted as needed to restore their function to that recorded in the initial session. The participant was also instructed to increase his/her minimum daily hearing aid usage to at least six hours.

It was also during this two-week follow-up session that all unaided measures of speech recognition were completed. A total of 12 unaided speech-recognition scores were obtained during this two-week follow-up session. There were four basic test conditions, but in each condition scores were obtained in the sound field from the right ear, the left ear, and then binaurally. For monaural testing, the nontest ear was occluded with a foam earplug. (Monaural unaided testing was included to permit examination of alternative definitions of objective hearing-aid benefit in subsequent analyses not reported here.) The four test conditions were as follows: (1) CUNY Nonsense Syllable Test (NST; Levitt and Resnick, 1978) presented at an overall level of 65 dB SPL and a +8 dB (SNR) using the recorded multitalker babble from the SPIN test (Kalikow, Stevens, and Elliott, 1977) as the competition; (2) Connected Speech Test (CST; Cox *et al.*, 1988) presented at an overall level of 50 dB SPL in quiet; (3) CST presented at an overall level of 65 dB SPL and a SNR of +8 dB using the recorded multitalker babble provided with the CST; and (4) CST presented at an overall level of 80 dB SPL and a SNR of 0 dB. The particular combinations of speech level and SNR were selected to cover a range of anticipated ''real world'' listening conditions as suggested recently by Walden (1997) with the specific combinations of speech level and SNR for this study based on the data of Pearsons, Bennett, and Fidell (1977). All speech materials were commercially available recorded versions. For the NST, the full 102-item, 11-subtest version was used and, for the CST, each score was based on two consecutive passages with each passage containing 25 key words for scoring. Different forms of the NST and different passages of the CST were used for each condition.

The speech signal for all speech-recognition measurements was presented from a loudspeaker (Radio Shack Optimus 7) located 1 m in front of the participant at 0° azimuth and elevation whereas the babble competition was delivered from an identical loudspeaker located 1 m behind the partici-

pant at 180° azimuth and 0° elevation. All testing was completed in double-walled sound booths complying with ambient noise levels specified by ANSI S3.1-1991 (ears uncovered testing). All presentation levels were established using the method of substitution with the measurement microphone positioned at a location corresponding to the center of the listener's head. For the CST, the calibration noise supplied with the materials was used to specify the sound levels. For the NST, a 400 Hz FM signal was generated that had the same rms level as the 400 Hz pure-tone calibration signal supplied with the materials and this stimulus was used to set the level of both the speech and babble competition in the sound field. Recordings were made of the FM calibration stimulus on both channels of a digital audio tape (DAT), with the NST materials on one channel and the SPIN multi-talker babble on the other channel. Differences between the rms levels of the FM calibration signal and the rms levels of the NST syllables and SPIN babble were taken into consideration for subsequent SII calculations.

For the NST, the participant marked the syllable heard on a large-font answer sheet containing seven to nine alternatives that differed from the stimulus by only one phoneme. For the CST, before testing began, the participant was provided with the passage topic and was encouraged to guess if uncertain about what was heard. After each sentence of the passage, there was a pause in the presentation of the speech signal and the participant orally repeated what had been heard. Using an orthographic representation of the passage, the experimenter proceeded to score the participant's response using the highlighted keywords.

Approximately two weeks later, the participant returned for the one-month follow-up visit. The hearing aids were again examined, evaluated in the testbox, and adjusted as needed to return their function to the target levels from the initial fitting session. Next, aided speech-recognition measures were obtained for the four test conditions in a manner identical to that described for unaided testing, but only for the binaural listening condition. Thus, one NST and three CST scores were obtained in this session with the participant wearing both hearing aids, each adjusted in the testbox to match the electroacoustic performance recorded in the initial fitting session. Test forms (NST) and passages (CST) not used previously were employed in this session.

III. RESULTS

Mean and individual data are presented in this section prior to evaluation of the SII and regression predictive schemes. These data are presented here both to provide a more thorough description of the characteristics of the study sample and to demonstrate that this study sample is representative of the elderly hearing-impaired population in general through comparisons to available ''norms'' for this population.

A. Audiological measures

Median air-conduction thresholds for each ear are presented in Fig. 1 for the 171 elderly hearing-aid wearers grouped by gender and age decade. There is a progressive decrease in hearing with age, especially in the high frequencies, with the possible exception of the left-ear hearing thresholds of the male participants (open symbols, top panel). As expected (e.g., Moscicki *et al.*, 1985), female participants had better hearing above 1000 Hz than their male counterparts. Although these relative age-related and gender-related trends in hearing thresholds are consistent with available data on presbycusis, it should be noted that the severity of hearing loss is greater than that of ''typical'' elderly individuals, especially for the 60 and 70 year olds. This is due to the exclusion of those in each age group with milder hearing loss from this study of elderly hearing-aid wearers.

Table I presents the means and standard deviations for the same group of participants on several additional audiological measures obtained. The pattern of results in the group data in Table I for SRT, word-recognition scores (NU-6) in quiet, ABR wave-V latency, DPOAE SNR, and PI-PB functions is consistent with the presence of a high-frequency sensorineural hearing loss bilaterally (as depicted previously in Fig. 1). That is, the group data suggest no significant declines in word-recognition scores when the presentation level is increased beyond that yielding the maximum score. The average decline in word-recognition score with increase in presentation level is approximately 4%–5% in each ear. Thus, on a group basis, there is little evidence of significant ''roll over'' in the PI-PB function which is consistent with cochlear site of lesion. In addition, mean DPOAE signal-to-noise ratios (SNRs) for low (1000–1500 Hz), mid (2000–3100 Hz), and high (4000–6300 Hz) frequencies are all consistent with the absence of functional outer hair cells in each ear. Finally, average wave-V latencies in each ear are typical of those measured in elderly listeners with comparable amounts of hearing loss and cochlear site of lesion (Jerger and Johnson, 1988).

Figure 2 provides a more detailed look at the DPOAE SNR values obtained in this study. The individual SNR values are plotted as a function of average hearing loss, either PTA (500, 1000, and 2000 Hz; top two panels) or HFPTA (1000, 2000, and 4000 Hz; lower four panels). The top two panels depict SNRs for the lower-frequency DPOAEs, middle panels for the mid-frequency DPOAEs, and bottom panels for the high-frequency DPOAEs, with the left-hand panels providing data for the left ear and right-hand panels for the right ear. Using a SNR criterion of 6 dB to indicate the presence of a DPOAE (represented by the horizontal dashed line in each panel), with a few exceptions, the presence of DPOAEs is only observed for milder amounts of hearing loss. Thus, the *individual data* for DPOAEs are also consistent with a cochlear site of lesion for the overwhelming majority of participants.

In a similar manner, Fig. 3 provides a more detailed examination of the ABR wave-V latencies obtained from the participants in this study. In each panel of Fig. 3, wave-V latency is plotted as a function of HFPTA, with the panels on the left depicting data for the left ear and those on the right providing data from the right ear. Data have been plotted separately for females (top panels) and males (bottom panels). The open circles connected by solid lines in each panel depict ''norms'' provided by Jerger and Johnson (1988) from

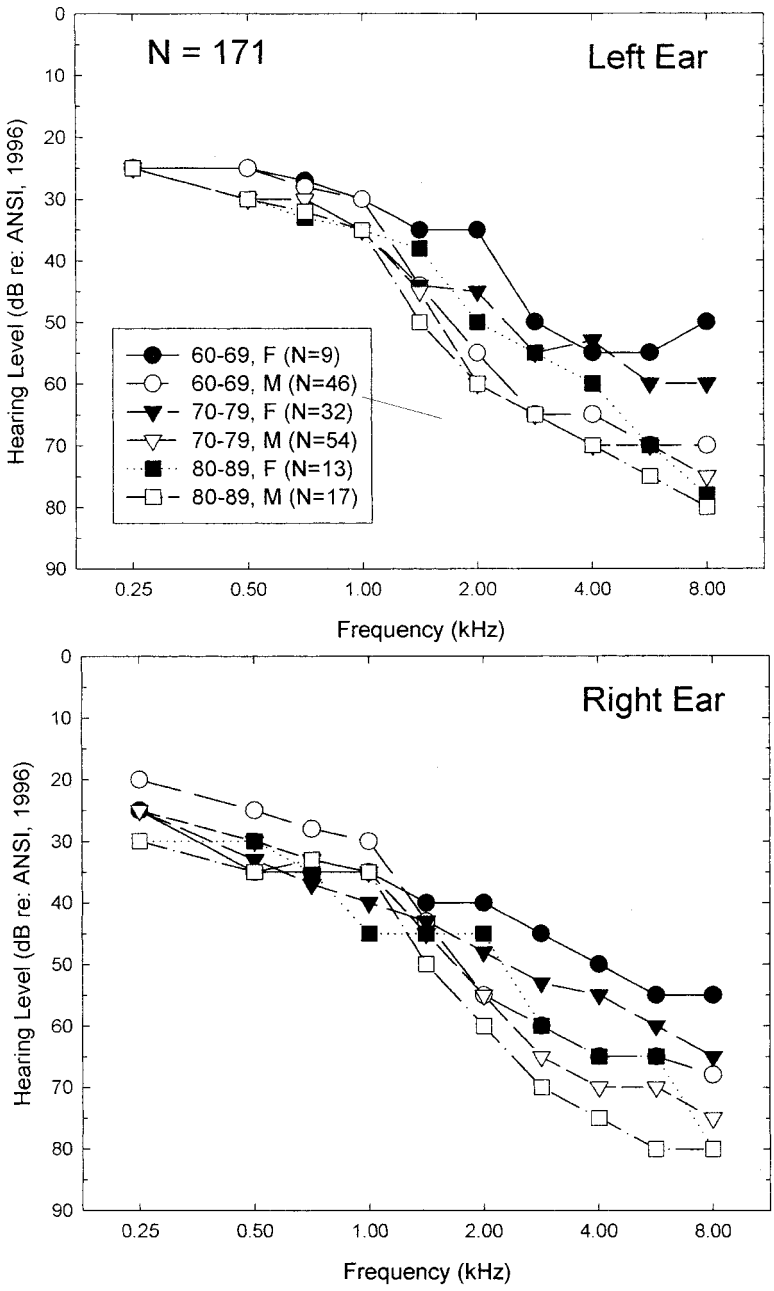


FIG. 1. Mean air-conduction pure-tone thresholds for left (top panel) and right (bottom panel) ears of 171 elderly participants in this study partitioned by age (circles: 60–69 years; triangles: 70–79 years; squares: 80–89 years) and gender (female: closed symbols; male: open symbols).

large groups of elderly males and females obtained for stimulus and recording conditions very similar to those used here. These normative data, moreover, were obtained from patients with cochlear site of lesion. For the elderly females in the present study (top two panels), there is good agreement with the average data from Jerger and Johnson (1988). Although there is similar agreement for the overwhelming majority of data gathered from the elderly males in this study (lower two panels), there are clearly several males for whom prolonged wave-V latencies are apparent, especially in the left ear (lower left-hand panel). Most of these prolonged latencies, however, occur for participants with severe amounts of high-frequency hearing loss (HFPTA >50 dB HL); cases for whom Jerger and Johnson (1988) used a stimulus level that was at least 10 dB higher than that used in the present study. Even for the data shown for females in the top panels of Fig. 7, once the average high-frequency hearing loss exceeds 50 dB HL, the majority of observed wave-V latencies are above the reported norms. This again is likely due to the use of a higher stimulus level by Jerger and Johnson (1988) in deriving these norms for individuals with more severe hearing loss (HFPTA >50 dB HL). In general, then, the in-

TABLE I. Means and standard deviations (in parentheses) for various audiological measures completed for each ear for the 171 elderly hearing-impaired listeners.

Variable	*M* (SD) for left ear	*M* (SD) for right ear
SRT (dB HL)	33.4(13.2)	33.2(13.1)
NU-6 Score in Quiet (%)	85.5(16.1)	85.9(17.6)
ABR Wave-V Latency (ms)	6.06(0.40)	6.03(0.32)
DPOAE-lf SNR (dB)	0.0(6.8)	−0.2(8.3)
DPOAE-mf SNR (dB)	−3.2(7.6)	−3.3(8.1)
DPOAE-hf SNR (dB)	−5.2(8.4)	−5.4(8.4)
PB-max (%)	92.9(15.3)	92.6(17.0)
PB-maxHL (%)	89.1(15.9)	88.4(16.2)
Rollover (%)	4.1(7.5)	4.7(6.5)

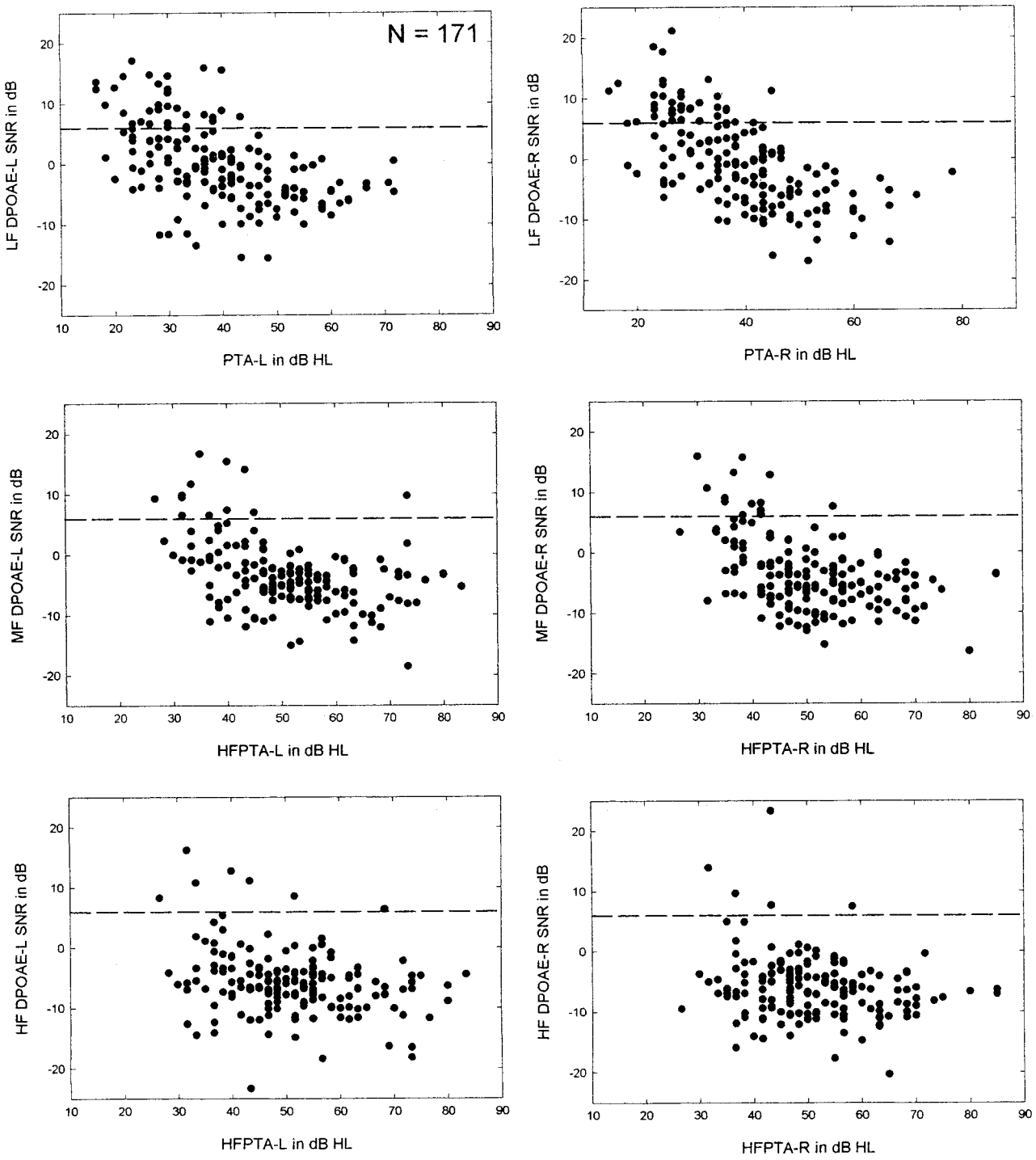


FIG. 2. Scatterplots of DPOAE SNR values as a function of various pure-tone averages for 171 elderly participants. Left panels depict data from left ear whereas right panels depict data from right ear. Top panels provide results for low-frequency DPOAEs, middle panels data for mid-frequency DPOAEs, and bottom panels results for high-frequency DPOAEs. The horizontal dashed line in each panel represents a 6 dB DPOAE SNR.

dividual data for ABR wave-V latency are also overwhelmingly consistent with a cochlear site of lesion, with the possible exception of a small number of males (<10) who may have wave-V latencies longer than expected given their gender and severity of hearing loss.

All told, there is little evidence in the audiological data from this study that supports the presence of so-called ''neural presbycusis'' (Schuknecht, 1964). Similarly, there is little evidence in this population for ''auditory neuropathy'' (Starr *et al.*, 1996). Only three cases, for example, were found of normal outer hair cell function (DPOAE SNR$\geqslant$6 dB) in any of the three frequency regions with presence of prolonged wave-V latency ($\geqslant$6.7 ms) in the same ear. All three cases were males, with two involving the left ear and one involving the right ear. Moreover, in two of the three cases, the HFPTA was greater than 60 dB HL. Further, in no cases were the DPOAE SNR values above 6 dB in more than one frequency region (that is, two of the three DPOAE measures in each ear were consistent with outer-hair-cell lesions). In summary, it appears that the cochlea, and the outer hair cells

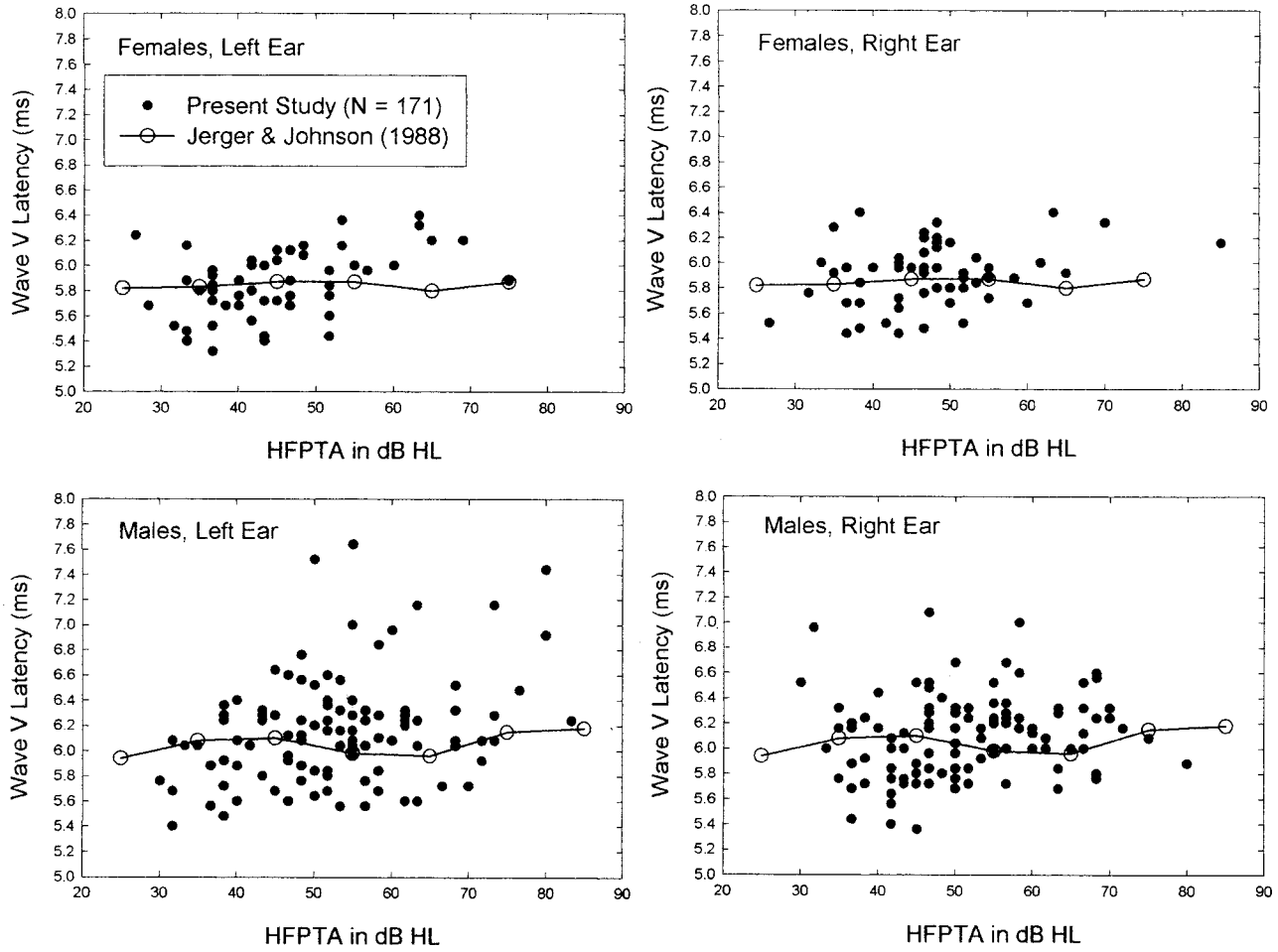


FIG. 3. Scatterplots of ABR wave-V latencies as a function of high-frequency pure-tone average (HFPTA) for elderly females (top two panels) and males (bottom two panels) and left and right ears (left and right panels, respectively). Normative data for reasonably similar stimulus and recording conditions, as well as similar study samples, from Jerger and Johnson (1988) are also provided in each panel (open circles connected by solid line).

in particular, are the primary site of lesion underlying the observed sensorineural hearing loss in the elderly participants in this study.

B. Auditory-processing measures

Figure 4 depicts the means (black bars) and standard errors for the six measures of auditory processing used in this study, three measures from the VACD and three from the TBAC. Performance on each of the six measures is expressed in terms of percent-correct scores. The gray vertical bars adjacent to the means for the three VACD measures are mean values from another group of elderly hearing-impaired listeners tested by Humes, Coughlin, and Talley (1996) at the same presentation level as the current study. There is close agreement between these two sets of data indicating that the present group of subjects is typical of other similar groups evaluated previously. For the TBAC, appropriate comparison data available for the same 90 dB SPL presentation condition were not available. Humes and Christopherson (1991) reported mean data for a group of 23 elderly hearing-impaired subjects for a presentation level of 75 dB SPL. The scores from the participants in the present study for the tonal duration-discrimination and temporal-order discrimination tasks, however, are not likely to be very different for presentation levels of 75 and 90 dB SPL (Christopherson and Humes, 1992). Humes and Christopherson (1991) reported mean scores of approximately 71% and 65% for the duration-discrimination and temporal-order discrimination tasks, respectively. These values are in good agreement with those shown for the same measures in the right-hand portion of Fig. 4. The syllable-sequence task of the TBAC, on the other hand, *is* impacted by presentation level and the audibility of the speech syllables used in the task (Christopherson and Humes, 1992) and there are no appropriate ''norms'' available for the 90 dB condition used in this study. In general, though, elderly hearing-impaired listeners consistently have scored much lower (>25% lower) on the syllable-sequence task than on the other two TBAC measures and that pattern is also observed in Fig. 4. In summary, the auditory-processing performance of the participants in this study is consistent with that observed previously in similar studies of elderly hearing-impaired listeners.

C. Cognitive measures

Figure 5 depicts the means and standard errors obtained for the WAIS-R for the six verbal scales, the five performance scales, and the three IQ measures derived from this cognitive measure. For the scaled scores, shown in the right

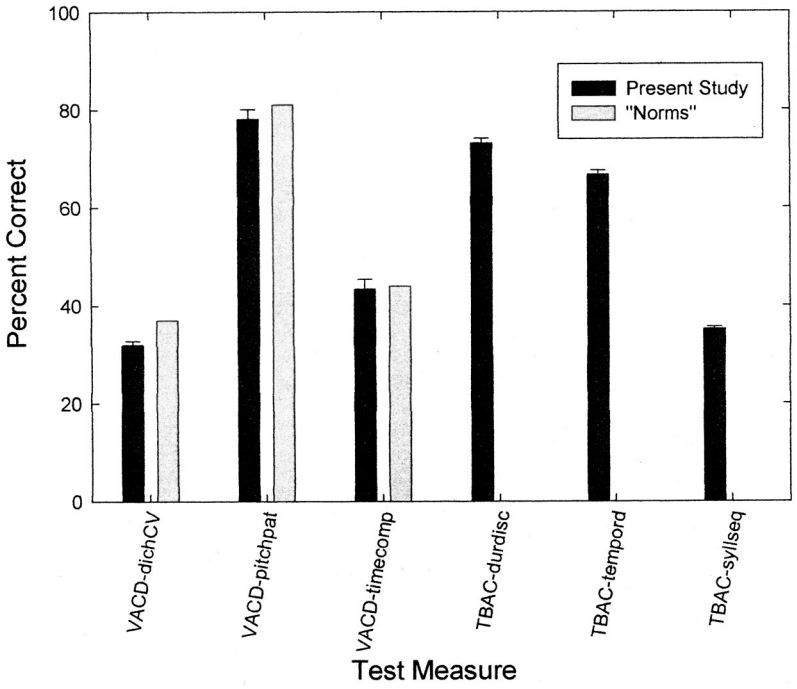


FIG. 4. Mean scores from 171 elderly participants for six measures of auditory processing, three from the VACD and three from the TBAC. Thin vertical lines represent one standard error above the means. For the VACD, normative data from similar listeners under similar test conditions were available for comparison and these mean values are plotted as gray bars. See the text for further details.

panel, age-related norms are available for every five-year span in age from 60 to 74 years and these norms are used to scale the raw scores so that the average age-corrected scale score will be 10 for every age group. In a similar fashion, IQ is scaled for each age group so that average IQ will be 100. From the means depicted in Fig. 5, it is apparent that the elderly participants in the present study were average or slightly above average overall for the verbal scaled scores, a little below average for the performance scaled scores, and above average in IQ. The norms used to scale the scores for each age group in the WAIS-R, however, are designed to determine how a particular individual's performance compares to those typical ''for their age.'' In general, the mean cognitive function of the present study sample indicates that cognitive function was representative of adults of the same age.

Such age-related norms are not of particular interest in the present study in that it was desirable to quantify age-based changes in cognitive function that might have an impact on speech-recognition performance. For example, the raw score on a particular WAIS scale might decline from age 60 to age 70, but still reflect average performance ''for their

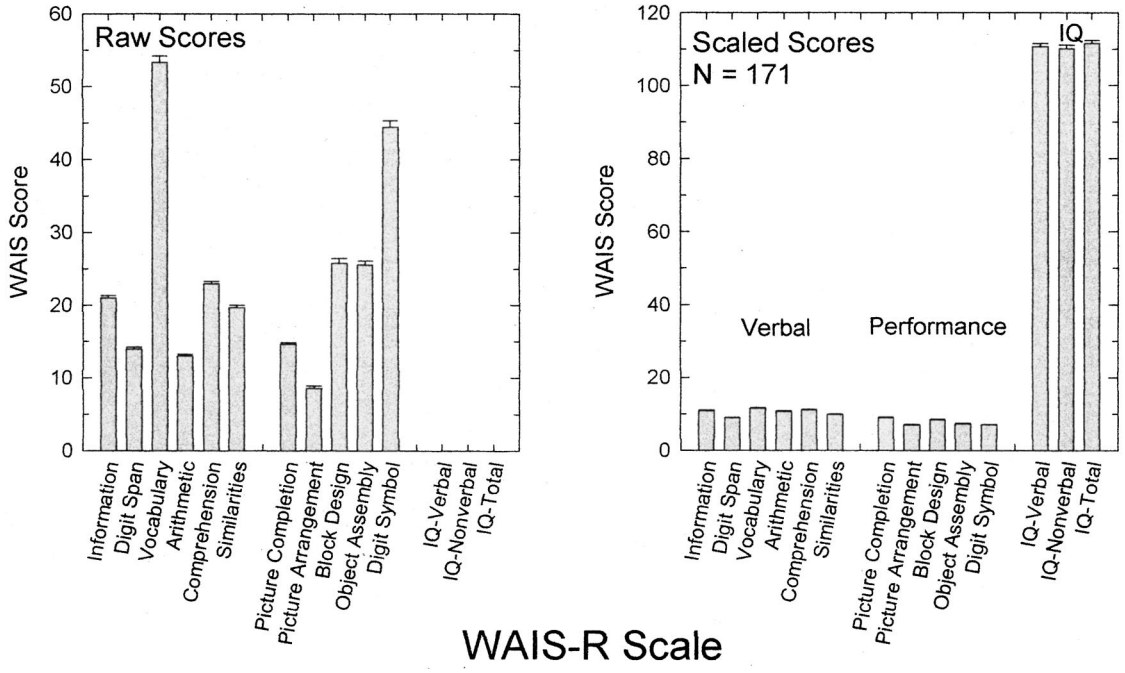


FIG. 5. Mean raw (left panel) and scaled (right panel) WAIS-R scores for 171 participants in this study. Thin vertical lines represent one standard error above the mean. IQ is only available as a scaled score. The scores on the six verbal scales of the WAIS-R are separated along the abscissa in each panel from the five performance-based scales (and IQ, in the right panel).

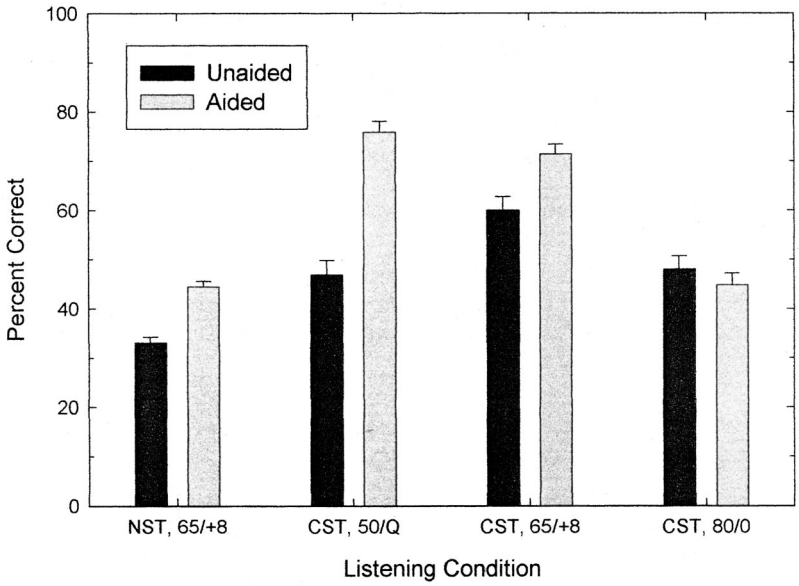


FIG. 6. Mean speech-recognition scores for each of the four test conditions in this study and for unaided (black bars) and aided (gray bars) listening conditions. Thin vertical lines at the top of each bar represent one standard error above the mean. NST=Nonsense Syllable Test and CST=Connected Speech Test.

age.'' The *scaled* score, as a result, would be 10 in each case and would obscure the age-related decline in cognitive function apparent in the *raw* scores. Thus, the raw scores for the WAIS-R are of greater interest for regression analyses in the present context. The means and standard errors for the raw WAIS scores are shown for each subscale in the left-hand portion of Fig. 5. Unless noted otherwise, all subsequent presentation of data for the WAIS-R will make use of raw scores for the 11 scales and scaled scores for the three measures of IQ.

D. Speech-recognition measures

Figure 6 presents the means and standard errors for unaided (black bars) and aided (gray bars) speech-recognition performance in quiet. The percent-correct scores for each participant and listening condition were converted to rationalized arcsine units (rau; Studebaker, 1985) to stabilize the error variance prior to statistical analysis. A paired-sample t-test for the transformed NST scores indicated that aided performance was significantly greater ($p<0.001$) than unaided performance. A repeated-measures General Linear Model (GLM) analysis of the transformed aided and unaided CST scores across the three conditions indicated significant ($p<0.001$) main effects of hearing aid and listening condition, as well as a significant interaction. Post-hoc paired-sample t-tests indicated that aided performance exceeded unaided performance for the CST at 50 dB in quiet and at 65 dB in a +8 dB SNR, but there were no differences between aided and unaided listening for the CST presented at 80 dB and a 0 dB SNR. In general, these findings for aided and unaided speech-recognition performance and the pattern of results across listening conditions are consistent with prior observations from similar groups.

IV. SII CALCULATIONS

Having demonstrated that the study sample was representative of ''typical'' elderly hearing-impaired listeners with regard to audiological, auditory-processing, cognitive and speech-recognition function, the ability of the SII to account for the speech-recognition performance of the study sample is examined in this section. Table II lists the one-third-octave-band sound pressure levels used in the calculation of SII values for each listener. Band-specific weights (W_i), rms speech levels (Lsp_i), and rms babble levels (Ln_i) for the

TABLE II. One-third-octave-band frequencies, sound pressure levels and weighting factors used in the calculation of SII values for the NST and CST. fc_i=one-third-octave-band center frequency; T_i=the binaural reference equivalent threshold sound pressure level in the free field for each band, from ANSI S3.6-1996; Lsp_i=rms speech levels in dB SPL for each band; Ln_i=rms noise or babble levels in dB SPL for each band; and W_i=weighting factor (importance) for each band.

Frequency fc_i	RETSPL T_i	NST Lsp_i	NST Ln_i	NST W_i	CST Lsp_i	CST Ln_i	CST W_i
250	11.0	56.2	67.1	0.0153	58.14	53.08	0.0255
500	4.0	52.7	65.1	0.0422	55.65	56.87	0.0694
750/800	2.0	54.2	59.6	0.0584	51.76	51.32	0.0374
1000	2.0	53.2	55.1	0.0667	50.80	46.05	0.0441
1500/1600	0.5	37.2	48.6	0.0893	46.94	45.89	0.1035
2000	−1.5	36.2	45.1	0.1104	43.54	44.09	0.1023
3000/3150	−6.0	37.2	35.1	0.0981	38.43	41.35	0.0738
4000	−6.5	34.2	32.1	0.0867	37.62	45.02	0.0596
6000/6300	2.5	23.2	24.1	0.0551	37.73	41.03	0.0365

CST were taken from Sherbecoe and Studebaker (2002a, 2002b). For the NST, the weights for nonsense syllables from the ANSI SII standard (ANSI S3.5-1997) were used. The one-third-octave-band speech and noise levels for the NST were directly measured by digitizing the speech and babble stimuli at a sampling rate of 22 kHz (low-pass filtered at 10 kHz) with 16 bit resolution. For the nonsense syllables, the carrier phrase (''You will mark __, please'') was digitally deleted for each token and the remaining CV- or VC-syllable stimulus files were saved to disk. The entire set of nonsense syllables was then concatenated to form one long stimulus file. The rms level of the concatenated set of syllables was then measured using a 100 ms analysis window. Successive 2048-point fast Fourier transform spectra, with Blackmann–Harris windowing, were averaged to generate the long-term rms spectra for the speech and the babble. The amplitudes of the components within each band were then summed on a power basis to produce the one-third-octave-band levels for the NST speech and SPIN babble appearing in Table II. (The FM calibration signal used to set the stimulus levels in the sound field was analyzed in an identical manner to provide a reference level in dB SPL for the analyses of the speech and babble stimuli.) For the NST and CST, and for both the speech and the babble stimuli of each test, levels appearing in Table II are for an overall rms level of 65 dB SPL as measured with the method of substitution at a location corresponding to the center of the listener's head. Finally, the reference equivalent-threshold sound pressure levels (RETSPLs) appearing in Table II are the binaural free-field thresholds for normal listeners from ANSI S3.6-1996.

As is apparent in Table II, not all one-third-octave-band center frequencies in the ANSI standard were included in these SII calculations. Unaided calculations made use of the bands listed in Table II from 250 through 6000 (6300) Hz. The gain at 6000 Hz was not recorded for the listeners in this study and this frequency was omitted from calculation of *aided* SII values. Based on prior analyses by Humes *et al.* (1986) and given the nature of the participants' hearing loss and the frequency-gain characteristics of the hearing aids in this study, it was felt that the center frequencies appearing in Table II provided sufficient resolution in the frequency domain for valid SII calculations in hearing-impaired listeners. Since not all of the bands included in the SII standard were used, the sum of the band-specific weights could not sum to 1.0 under optimal conditions, as expected. This was addressed in these analyses for both the NST and CST by summing the weights provided in Table II (separately for each test), calculating the SII for the bands indicated in Table II, and then dividing the calculated SII by the summed weights. For example, for the calculation of SII values for the NST and aided listening conditions (250–4000 Hz), the weights in Table II sum to 0.5671. If the SNR in each band is 15 dB, the resulting SII for these analysis would be 0.5671, but should be 1.0. Dividing the calculated SII value by 0.5671 results in the correct SII value of 1.0.

For the calculation of the SII, the range of band-specific SNR values recommended in ANSI S3.5-1997 (+15 to −15 dB) was used for all calculations. In addition, the level-distortion factor incorporated in the standard was used in all

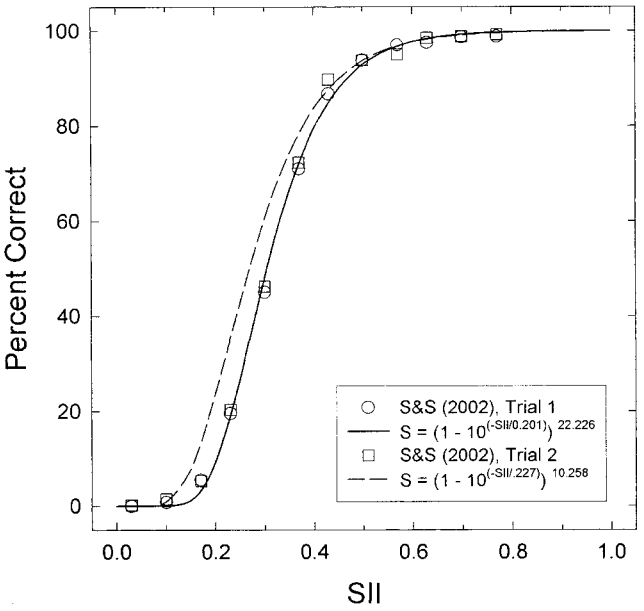


FIG. 7. Mean CST percent-correct scores from Sherbecoe and Studebaker (2002a) for 12 young normal-hearing listeners under wide-band listening conditions. Listeners were tested on two occasions, represented by the circles (Trial 1) and squares (Trial 2). The dashed line represents the best-fitting transfer function obtained by Sherbecoe and Studebaker (2002a) for this wide-band condition, as well as several low-pass and high-pass filtered conditions. The solid line represents the best-fitting transfer function fit to the wide-band data only and using the SII values as calculated in this study.

calculations. The best hearing threshold at each frequency in either ear was selected for each listener and used in the SII calculations. Aided speech and noise spectra were amplified by an amount determined from application of the NAL-RP prescriptive formula to the thresholds selected at each frequency for each listener from the preceding step.

Finally, two different sets of SII calculations were performed, one with hearing-loss desensitization (HLD) included and one without it. For the inclusion of HLD corrections, a formula similar to that proposed by Sherbecoe and Studebaker (2002b) was employed. Specifically, $HLD = 1 - (HFPTA_{bin}/108.3072)^3$, where $HFPTA_{bin}$ is the average hearing loss in both ears at 1000, 2000, and 4000 Hz. This correction factor is applied by multiplying the SII value by HLD. SII values with the HLD correction applied are identified here as SII_{HLD}.

To predict speech-recognition scores from SII values, transfer functions from young normal-hearing listeners linking these two variables are also needed. For the CST, Sherbecoe and Studebaker (2002a) provided a transfer function for the data obtained in their study. This transfer function appears as the dashed line in Fig. 7, together with the mean data from their study (circles, squares) for a wideband (100–11 000 Hz) listening condition and a group of 12 normal-hearing listeners tested on two different occasions (Trials 1 and 2). The solid line in Fig. 7 represents the best-fitting transfer function obtained in the present study for the wide-band condition depicted, with SII values computed as described in this paper. In both cases the equations for the transfer functions are of the form first described by Fletcher and Galt (1950), but each differs in the specific values of the two free parameters. The specific equations for each transfer

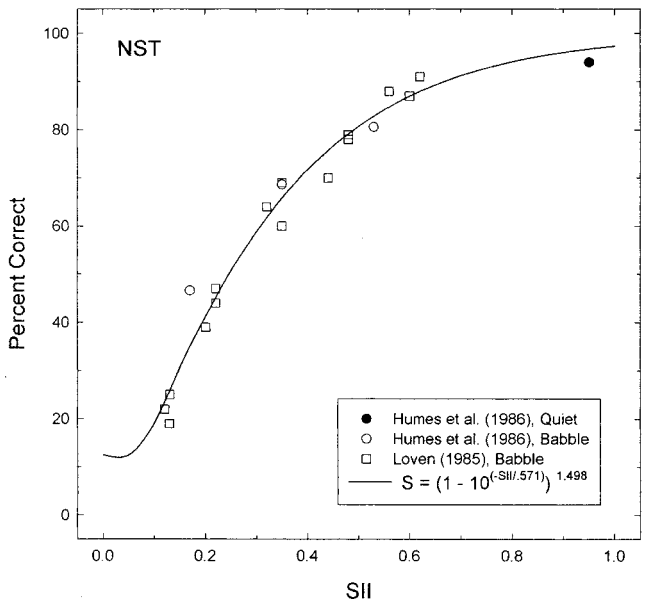


FIG. 8. Mean NST percent-correct scores from Humes *et al.* (1986) and Loven (1985), represented by circles and squares, respectively. Except for the closed circle, representing data obtained in quiet, all other conditions made use of the same multitalker babble from the SPIN test as used in this study. The solid line represents the best-fitting transfer function for these data, with adjustment made to the performance floor (12.5% instead of 0%) for SII values below 0.1.

function are provided in Fig. 7. Since there were slight differences between the SII calculation procedures of Sherbecoe and Studebaker (2002a) and those used in this study and the solid transfer function provides a better fit of these reference data to our SII values than the dashed transfer function, the solid transfer function derived here was used to generate predicted CST scores. [Unlike the SII calculation procedure in this study, Sherbecoe and Studebaker (2002a) used the full complement of one-third-octave bands provided in the ANSI standard, their own directly measured band-specific speech-peak values, and did not incorporate the band-specific level-distortion correction factor from the standard in their SII calculations.]

Figure 8 provides a comparable performance-vs-SII transfer function for the NST. The data in this study were obtained, with the exception of one data point obtained in quiet (closed circle), with the SPIN babble as the background competition. The best-fitting transfer function was: $S=(1-10^{(-\mathrm{SII}/0.571)})^{1.498}$. This function is presented by the solid line for NST scores greater than 12.5%. Since there are an average of eight alternative responses for each stimulus in this closed-set task, chance performance of 12.5% represents the performance floor, rather than 0%. Once the transfer function was derived, those predicted scores less than 12.5% were set to 12.5% and this portion of the solid function in Fig. 8 represents a cubic-spline fit of the transfer function combined with the 12.5% performance floor.

Having established appropriate reference performance-vs-SII transfer functions for young normal-hearing listeners using the SII calculation scheme in this study, individual data from the elderly hearing-impaired listeners could now be examined relative to these normative transfer functions. Figure 9 provides a series of scatterplots depicting speech-recognition performance of the elderly listeners as a function of SII. The top left panel shows the results for the NST whereas the other three panels depict results for the CST, each for a different listening condition. In each panel, the performance-vs-SII transfer function for each test is provided by the solid line and the dashed lines above and below the transfer functions represent 95% critical-difference bounds from Thornton and Raffin (1978) for tests of 100 items (NST) or 50 items (CST). The closed circles in each panel represent performance for the unaided listening conditions whereas the open circles represent performance in the aided conditions. Several general statements can be made about the scatter plots in Fig. 9. First, regardless of test or listening condition, more individual data points fall within the 95% critical-difference boundaries for unaided listening conditions than for aided listening conditions. Second, except for unaided CST scores obtained at 50 dB in quiet (top right panel), actual speech-recognition performance of the elderly hearing-impaired listeners tends to be lower than expected based on the SII. Third, the range of SII values is much more restricted for the CST at 80 dB and +0 dB SNR (bottom right panel) compared to the other conditions. This simply reflects the fact that the high speech and noise levels used in this condition determine the audibility of speech much more so than individual variations in threshold.

Figure 10 provides scatterplots equivalent to those in Fig. 9, but for the SII corrected by the hearing-loss desensitization factor (HLD). Clearly, by comparing the scatterplots in Figs. 9 and 10, the HLD correction shifts many of the data points to lower SII values and results in more of the speech-recognition scores falling within the 95% critical-difference bounds than for the uncorrected SII. Still, even with this correction, there are many data points below the 95% critical-difference boundaries of each transfer function, especially for the aided listening conditions.

As noted previously, age corrections for SII predictions have also been advocated (Studebaker *et al.*, 1997; Sherbecoe and Studebaker, 2002b). Rather than incorporate age corrections into the SII values, however, this issue will be addressed in the subsequent section on the linear-regression predictions.

V. LINEAR-REGRESSION PREDICTIONS

Based on prior experience with a similar approach to the prediction of unaided speech-recognition performance in a group of 50 elderly hearing-impaired listeners (Humes *et al.*, 1994), it was anticipated that the analysis of the association between prefit predictor variables and subsequent speech-recognition measures would make use of canonical correlations, a statistical tool that examines associations between two *sets* of variables. In this case, it was the association between the set of prefit predictor variables and the set of speech-recognition measures that was of potential interest. However, inspection of the correlation matrix for the eight measures of speech-recognition found the correlations to range from 0.43 to 0.78, with most in the range of 0.55–0.65. Subsequent principal-component factor analysis (Gorsuch, 1983) indicated that a single speech-recognition factor, ac-

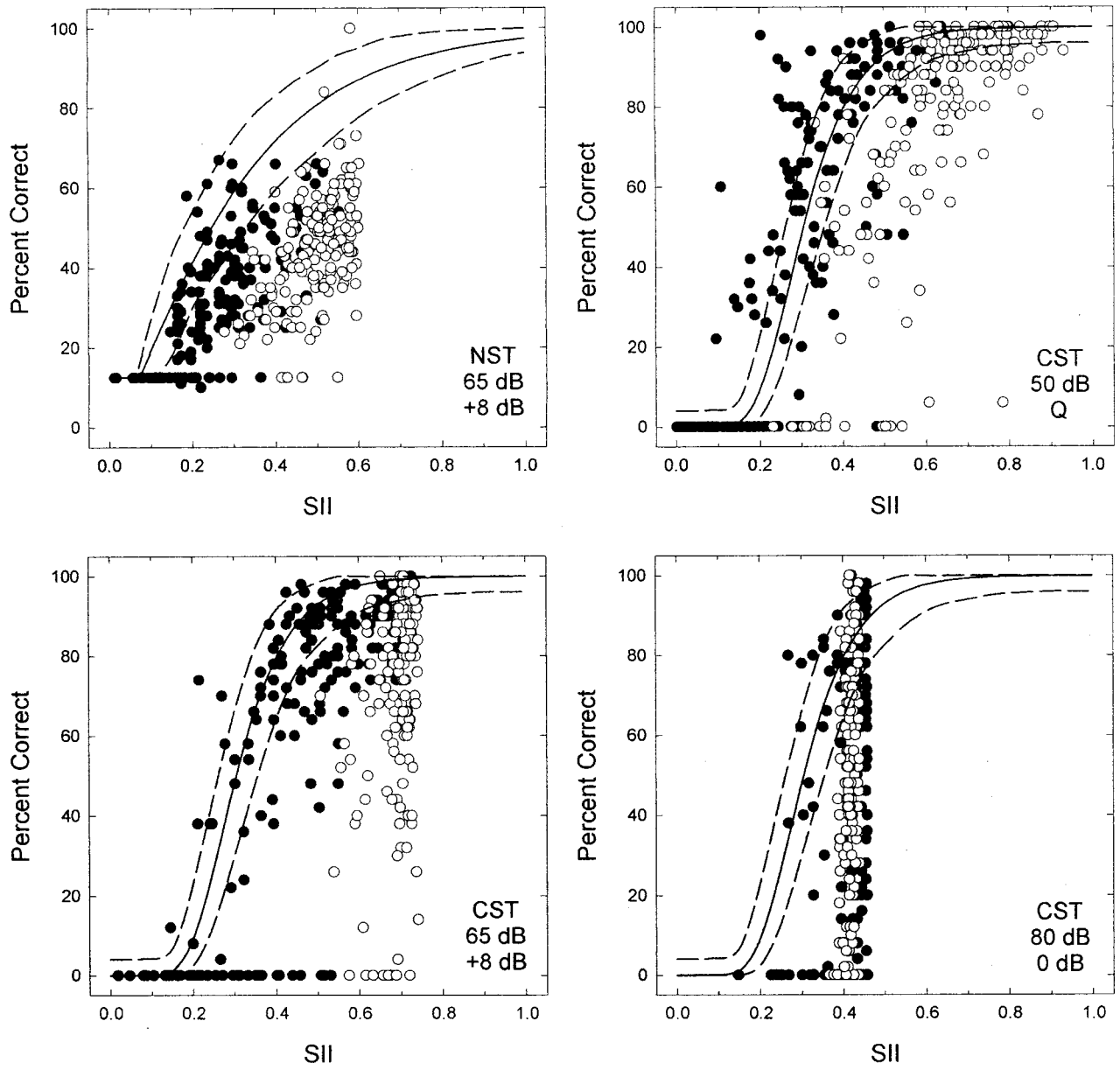


FIG. 9. Scatterplots of percent-correct as a function of SII for each of the four listening conditions. Each panel depicts the results for a different listening condition and provides the appropriate transfer function from Figs. 7 and 8 (solid lines) for reference. The dashed lines above and below each transfer function represent the corresponding 95% critical-difference boundaries for each function. Closed symbols represent unaided listening conditions and open symbols represent aided listening conditions.

counting for 61.0% of the variance in speech-recognition scores, was adequate. Given test–retest correlation coefficients for the NST and CST on the order of 0.8–0.9 (Humes *et al.*, 1987; Cox *et al.*, 1988), a factor analysis accounting for 61% of the total variance accounts for 75%–95% of the systematic variance in speech-recognition performance, and this was considered to be a good fit. Since the eight speech-recognition scores could now be represented accurately by a single general speech-recognition factor score (with mean of 0 and standard deviation of 1.0), a linear-regression approach could be pursued in which the set of prefit variables would be used to predict the general speech-recognition factor score. It should also be noted that when the loadings of each of the eight speech-recognition scores on the lone principal component were examined, they all ranged from a low of 0.70 (CST at 80 dB, aided) to a high of 0.87 (CST at 65 dB, unaided) indicating that this lone factor represented all eight of the scores adequately and with roughly equivalent weight.

There were a total of 33 prefit variables considered as possible predictors of speech-recognition factor scores. These included the eleven raw scale scores and three measures of IQ (verbal, nonverbal, and total) from the WAIS-R, five measures of auditory processing, ABR wave-V latencies for the right and left ears, DPOAE SNR values for each ear and for low, mid, and high frequencies, PTA and HFPTA for the right and left ears, age, and a measure of aided speech audibility (aided HFPTA, or HFPTA minus the average measured hearing-aid gain at 1000, 2000, 3000, and 4000 Hz, was computed for each ear and the minimum value was selected). Recognition of time-compressed NU-6 words and

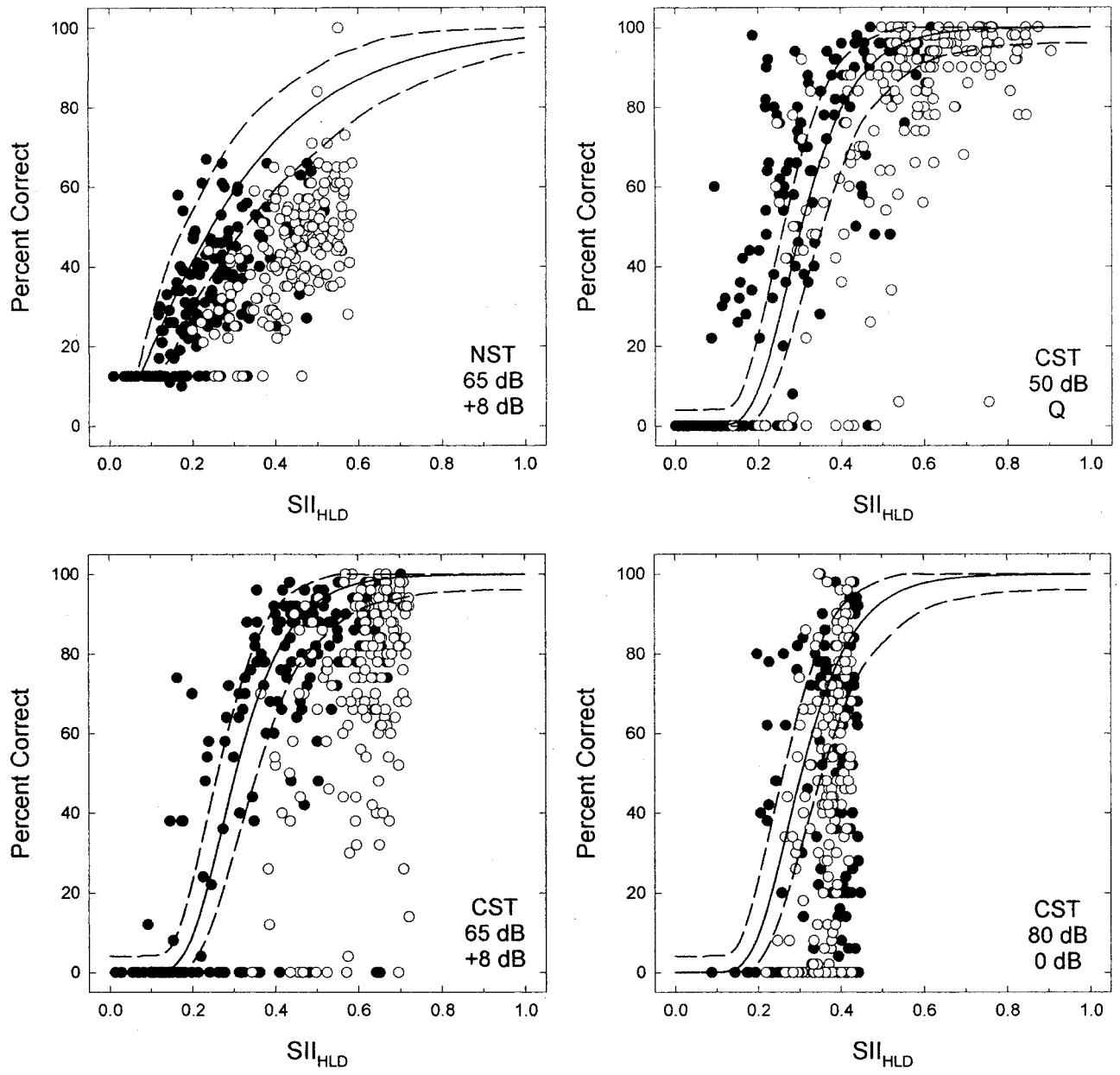


FIG. 10. Same as Fig. 9, but for plots of percent-correct versus the SII adjusted by the hearing-loss-desensitization (HLD) factor, or SII_{HLD}.

NU-6 PI-PB measures were not included as predictor variables because these measures were thought to be too close to the measures being predicted. That is, predicting speech-recognition performance from other measures of speech recognition was not of interest. The 33 prefit variables were then subjected to principal-components factor analysis to eliminate redundancy. Orthogonal rotation of factors was also employed to minimize co-linearity among the resulting factor scores. A total of seven orthogonal principal components were identified and accounted for 69.9% of the total variance among the set of 33 predictor variables.

The loadings of each predictor variable on each of the rotated orthogonal factors are provided in Table III. The first factor accounted for 15.4% of the total variance and was labeled ''Verbal IQ'' based on the factor loadings in Table III. The second factor was labeled ''Hearing Loss'' and accounted for 13.8% of the variance. The third factor accounted for 12.7% of the variance and was labeled ''Nonverbal IQ & Aging.'' The fourth factor was labeled ''DPOAE'' and accounted for 10.5% of the total variance. The fifth, sixth, and seventh factors were labeled ''auditory processing,'' ''ABR,'' and ''miscellaneous'' and accounted for 7.8%, 5.5%, and 4.2% of the total variance, respectively. Of course, as the percentage of variance accounted for diminishes, so does the robustness of the factor identified. This, in turn, makes the labeling of the factor more difficult, as in the case of the ''miscellaneous'' factor identified in Table III.

Next, the association between the seven predictor principal components and the general speech-recognition principal component was examined using stepwise linear regression. The following regression equation yielded a multiple correlation (r_m) of 0.82, accounting for 67.7% of the total variance in the general speech-recognition (SR) factor:

TABLE III. Principal-component factor loadings for the 33 prefit predictor variables. Orthogonal rotation of factors was accomplished using the varimax criterion. Factor identification was determined using a criterion of eigenvalue ⩾1.0. Only values ⩾0.30 shown.

Prefit variable	Verbal IQ	Hearing Loss	Nonverbal IQ & Aging	DPOAE	Auditory Processing	ABR	Miscellaneous
WAIS-information	0.784						
WAIS-digit span	0.464				0.343		0.505
WAIS-vocabulary	0.873						0.579
WAIS-arithmetic	0.367		0.353				
WAIS-comprehension	0.671						
WAIS-similarities	0.693						
WAIS-picture completion			0.655				
WAIS-picture arrangement	0.390		0.641				
WAIS-block design			0.764				
WAIS-object assembly			0.749				
WAIS-digit symbol			0.480				
WAIS-verbal IQ	0.913						
WAIS-nonverbal IQ	0.461		0.786				
WAIS-total IQ	0.813		0.509				
Dichotic CV			0.413		0.466		
Pitch pattern	0.344				0.547		
Syllable sequence					0.615		−0.321
Duration discrimination					0.708		
Temporal order-tones					0.804		
ABR-V right						0.895	
ABR-V left						0.887	
DPOAE-lf right		−0.460		0.620			
DPOAE-lf left		−0.505					0.509
DPOAE-mf right				0.851			
DPOAE-mf left				0.862			
DPOAE-hf right				0.887			
DPOAE-hf left				0.853			
PTA left		0.890					
PTA right		0.875					
HFPTA left		0.858					
HFPTA right		0.894					
aided HFPTA best		0.842					
Age			−0.632				

$$\begin{aligned} SR = &-0.73\ \text{Hearing Loss} \\ &+0.27\ \text{Nonverbal IQ \& Aging} + 0.23\ \text{Verbal IQ} \\ &+0.10\ \text{DPOAE} + 0.10\ \text{Miscellaneous}, \end{aligned} \quad (1)$$

where all of the coefficients are standardized beta coefficients (i.e., range from −1 to +1). The Hearing Loss factor is the only one in Eq. (1) that is inversely related to the general speech-recognition factor. This indicates that as hearing loss increased, unaided and aided speech-recognition decreased. This factor alone accounted for 53.2% of the total variance in SR factor scores and was the single largest contributing factor among the predictor variables. Next, the factor associated with Nonverbal IQ & Aging accounted for an additional 7.1% of the variance in the best-fitting function. Nonverbal IQ measures were positively loaded on this factor whereas age was negatively loaded on this factor (see Table III). Consequently, the positive beta coefficient for this factor indicates that general speech-recognition performance increased as nonverbal IQ increased and age *decreased*. Individual differences in Verbal IQ accounted for another 5.4% of the variance in general speech-recognition ability such that higher Verbal IQ scores yielded higher speech-recognition performance. Each of the remaining factors in Eq. (1), DPOAE and Miscellaneous, although significant, only accounted for an additional 1.0% of the variance and are not discussed further. Clearly, the Hearing Loss factor is the strongest contributor to these predictions accounting for nearly 80% of the explained variance (53.2% for the Hearing Loss factor divided by 67.7% for the entire regression equation).

The top panel of Fig. 11 presents a scatterplot of predicted general speech-recognition factor scores from the best-fitting regression model [Eq. (1)] versus the observed general speech-recognition factor scores for the 171 elderly participants in this study. As noted, the correlation between predicted and observed general speech-recognition factor scores is 0.82 and the rms error of the predictions is 0.58 (in standardized Z-score units). How does this compare to the predictive accuracy of SII-based predictions? To examine this in a manner that would facilitate comparison to the results in the top panel of Fig. 11, a principal-components factor analysis of the eight SII_{HLD} values for each listener was conducted and a one-factor solution was identified that accounted for 87.1% of the individual variations in values among the 171 participants. These values were saved as SII_{HLD} factor scores and the middle panel of Fig. 11 plots the predicted speech-recognition factor scores against the ob-

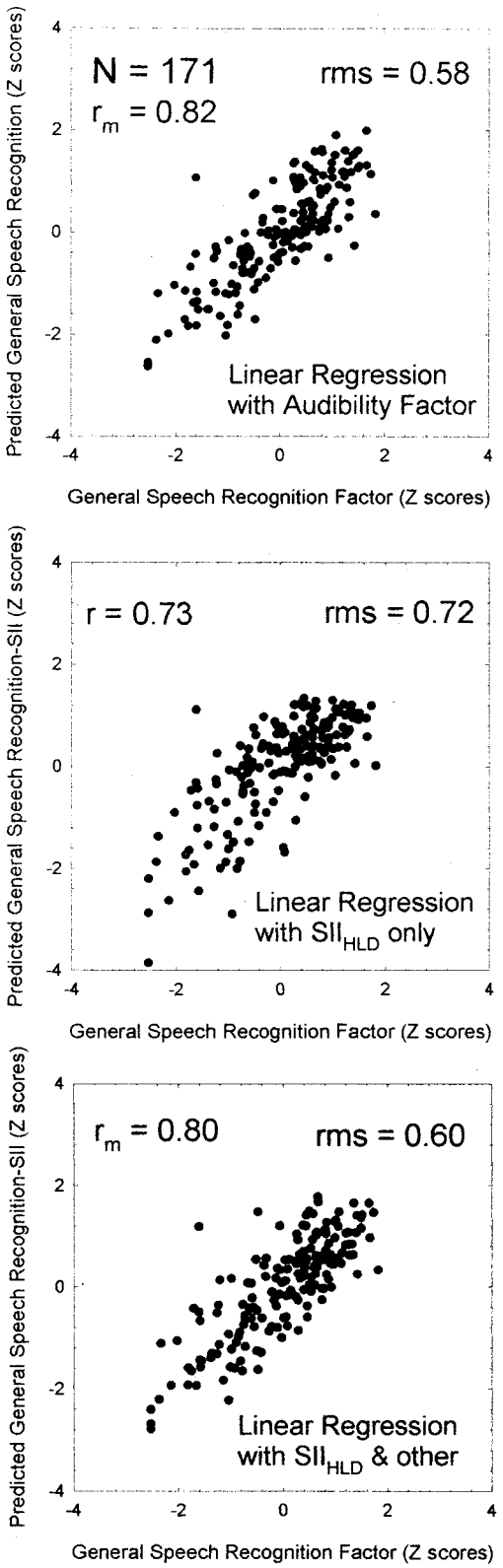


FIG. 11. Scatterplots of various predicted general speech-recognition factor scores, in standardized (z-score) form, as a function of the observed general speech-recognition factor scores in standardized form. Perfect agreement between predicted and observed scores would be represented by all data falling along positive-going diagonal. Top panel: predictions generated with linear-regression Eq. (1); middle panel: predictions from SII_{HLD} only; bottom panel: predictions from SII_{HLD} combined with six other predictor variables used in top panel (i.e., SII_{HLD} replaced the Audibility factor in regression analysis for top panel). Strength of association and accuracy of predictions are summarized in each panel by correlation coefficients (r) and rms crror values (in z-score units).

served scores when SII_{HLD} is the *only* predictor variable in the regression analysis. Clearly, these predictions in the middle panel are not as good as those in the top panel for Eq. (1), an observation that is supported by the correlation coefficients and rms error values appearing in each panel. Next, the SII_{HLD} factor score was included with the other six predictive factor scores to generate predictions. That is, the SII_{HLD} factor score simply replaced the Hearing Loss factor score in the linear-regression analysis. The other six prefit factor scores from the prior regression analysis remained unchanged. The resulting regression equation was nearly identical to Eq. (1) with SII_{HLD} accounting for 53.2% of the total variance in speech-recognition performance, followed by 5.8% of the variance accounted for by the factor of Nonverbal IQ & Aging, 2.4% by the Verbal IQ factor, 1.8% by the Miscellaneous factor, and 1.4% by the ABR factor [instead of the DPOAE factor in Eq. (1)]. The standardized beta coefficients for each predictor variable in this equation were also very similar to those in Eq. (1). However, the beta coefficient for the SII_{HLD} factor, although of identical magnitude (0.73) to the coefficient for Hearing Loss, was now of opposite sign, reflecting the fact that higher SII_{HLD} values result from lower amounts of hearing loss. A scatterplot of predicted versus observed general speech-recognition factor scores for the full regression equation incorporating SII_{HLD} is shown in the bottom panel of Fig. 11. Clearly, the addition of other predictive factors, especially both age-related and non-age-related cognitive factors, improves the predictive power of the SII_{HLD} considerably. This is entirely consistent with the concept of age-corrected SII_{HLD} advocated previously by Studebaker and colleagues. Note, however, that there is little difference between the predictive accuracy of the top and bottom panels suggesting that less elaborate measures of speech audibility, such as PTA, HFPTA, and aided HFPTA, may suffice when generating such predictions.

VI. GENERAL DISCUSSION

The results presented in Table I and Figs. 1–5 regarding prefit measures of hearing loss, ABR wave-V latency, DPOAE SNR, auditory processing, and cognitive function (WAIS-R) demonstrated that the 171 elderly hearing-impaired listeners in this study were typical of others tested previously. In addition, the accuracy of SII and SII_{HLD} predictions for the CST and *unaided* listening was found to be very similar to that reported recently by Sherbecoe and Studebaker (2002b) for a group of 76 elderly hearing-impaired listeners. Thus, the participants in the present study can be considered to be representative of elderly hearing-impaired listeners in general. In addition, the agreement of these results with prior SII analyses, at least for the CST, suggests that the procedures followed in calculating the SII, including the use of a restricted number of bands, were valid.

One of the significant findings in this study was that individual differences in speech-recognition performance for the wide range of conditions sampled, including aided and unaided listening, materials ranging from nonsense syllables to meaningful sentences, and test conditions varying from soft speech in quiet to loud speech in noise, could be de-

scribed by a single general speech-recognition factor. This is similar to the findings in a previous study by Humes *et al.* (1994) in which 50 elderly hearing-impaired listeners were tested unaided in 20 different conditions that included factorial combinations of five different types of test materials (including nonsense syllables, words, and sentences), two different speech levels (70 and 90 dB SPL), and two different background conditions (quiet and +7 dB SNR for a speech-shaped noise). One set of monosyllabic words was also spectrally shaped to provide gain equivalent to that of the NAL-R prescription for the average audiogram of the group. In that study, however, two principal components were found to underlie the speech-recognition measures, one that was interpreted as a general speech-recognition factor that captured 74.1% of the total variance and one that had heavier factor loadings for test conditions involving high speech levels in noise, which accounted for an additional 7.3% of the total variance. Although two speech-recognition factors were identified in that prior study of 50 elderly hearing-impaired listeners, the first factor, on which 17 of the 20 measures of speech recognition were loaded heavily (factor loadings >0.6), clearly accounted for the majority of the variance in speech-recognition performance. In addition, although the study by Humes *et al.* (1994) was concerned with unaided speech-recognition, the use of a 90 dB SPL presentation level for half of the conditions (and spectrally shaped speech for 20% of the conditions) resulted in many listening conditions that were not unlike aided listening conditions for the participants. Still, 74.1% of the total variance in speech-recognition performance was accounted for by one factor in that study.

In the prior study by Humes *et al.* (1994), given the presence of a two-factor representation of speech-recognition performance in that study, canonical correlation was used to determine the association between various prefit variables and speech recognition. In that study, as in this one, an audibility or hearing-loss factor was the primary factor underlying a strong relationship (canonical correlation$=0.89$) between the set of predictor variables and the set of speech-recognition factors. A cognitive factor, one associated with four of the five measures of nonverbal performance from the WAIS-R, was the most substantial secondary contributor to the association between the sets of prefit and speech-recognition factors.

Similar findings were obtained by van Rooij, Plomp and colleagues in a series of studies examining various auditory and cognitive contributors to unaided speech recognition in elderly hearing-impaired listeners (van Rooij, Plomp, and Orlebeke, 1989; van Rooij and Plomp, 1990, 1992). Multiple measures of speech recognition were employed in those studies and ranged from closed-set recognition of vowels and consonants in noise to sentence recognition in quiet and noise. A single general speech-recognition factor was identified that could account for approximately 70% of the variance. Predictor variables examined in that series of studies included measures of hearing loss, frequency selectivity, temporal resolution, and cognitive function. Across this series of studies, two predictive factors, one related to hearing loss and the other to cognitive function, were needed to account for individual differences in speech-recognition performance among elderly hearing-impaired listeners. The hearing-loss factor accounted for 67%–90% of the systematic variance, with the balance (10%–33%) accounted for by the cognitive factor. Although differing in details, the same general findings regarding the primary contribution of hearing loss to unaided speech recognition in elderly hearing-impaired listeners have been observed by others as well (Jerger, Jerger, and Pirozollo, 1991; Divenyi and Haupt, 1997a, 1997b, 1997c). In summary, several prior studies have identified a general speech-recognition factor underlying speech perception in the elderly hearing-impaired for *unaided* speech stimuli and have found hearing loss to be the single best predictor of individual differences in this speech-recognition factor. Other predictive factors, primarily cognitive ones, appear to play a consistent, but secondary, role. In this study, as well as the prior studies by Humes and colleagues and van Rooij and colleagues, supplementary physiological or perceptual measures of auditory function did not play a role in explaining individual differences in speech recognition for the listening conditions evaluated.

Perhaps, given the focus on *unaided* speech-recognition performance in these prior studies, these findings are not too surprising. Figure 12, for example, provides plots of the median hearing thresholds for the 171 elderly participants in this study (solid line), together with the one-third-octave-band speech and babble levels (closed and open circles, respectively), with each panel depicting one of the four speech-recognition test conditions in this study. Clearly, except for the 80 dB SPL CST test condition (lower right panel), hearing thresholds limited the audibility of the speech (and babble) stimuli considerably. From Fig. 12, it is also easy to envision how variations in threshold above and below the median values depicted will create variations in audibility of speech concurrently across all four unaided test conditions. Consequently, as has been observed in this study and prior studies, hearing loss is the primary contributor to unaided speech-recognition performance regardless of the manner in which speech-recognition performance is measured.

Consider now the aided speech and babble one-third-octave levels (triangles) in Fig. 12. To facilitate evaluation of speech audibility here, dotted lines 15 dB above the aided speech rms levels have been added to each panel. In the SII model, this dotted line represents the top of the useful speech range that contributes to speech understanding (i.e., the speech peaks). (This is also the case for unaided speech-recognition, but the corresponding peak values were omitted for clarity.) It is apparent that, although the aided speech (and babble) one-third-octave-band levels are restricted much less so by the median hearing thresholds than the corresponding unaided values, the audibility of the speech stimuli may still be less than optimal from 1000 through 4000 Hz, even for the median hearing thresholds. This is especially true for the CST at 50 dB in quiet (top right panel). Only for the CST at 65 and 80 dB (two bottom panels) is the babble competition the factor limiting audibility of amplified speech from 250 to 4000 Hz, at least for the median hearing loss.

To evaluate the role of hearing loss further, only the data

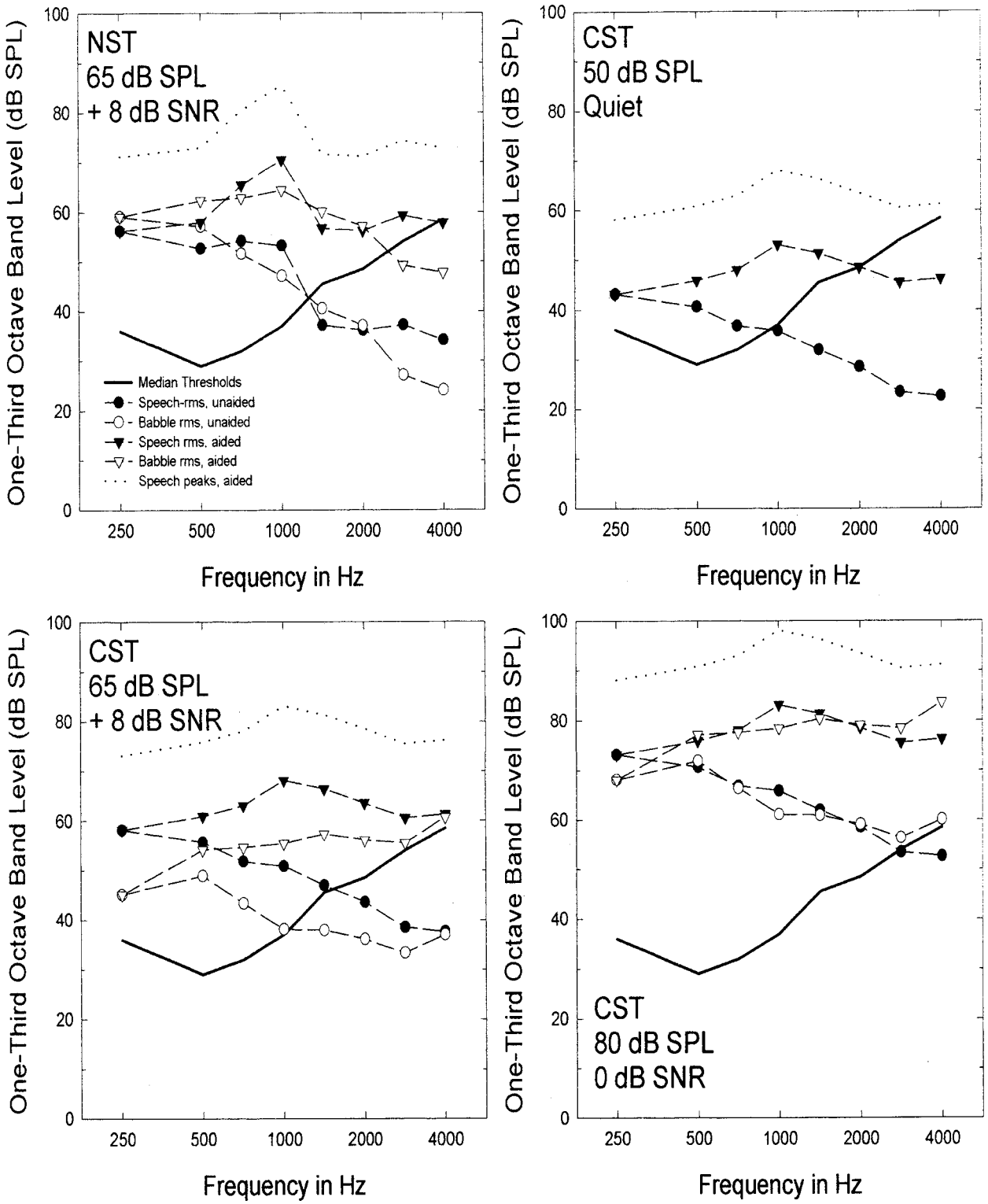


FIG. 12. Plots of the one-third-octave-band levels, in dB SPL, as a function of frequency, for the speech (closed symbols) and babble (open symbols) stimuli for unaided (circles) and aided (triangles) listening conditions. Each panel depicts one of the four speech-recognition test conditions in this study and also provides a plot of the group median hearing thresholds on the same coordinates (heavy solid line). The dotted line in each panel depicts the speech peaks for the aided speech stimuli and is located 15 dB above the rms speech levels for the aided condition (filled triangles).

from those participants with bianaural-averaged HFPTA values less than or equal to the median value of 50 dB HL were selected for analysis. Linear-regression analysis of the general speech-recognition factor for these 87 participants with milder high-frequency hearing loss revealed four predictive factors very similar to the ones identified previously for the entire group in Eq. (1). The predictor variables identified, and the percentage of total variance accounted for by each, were as follows: Hearing Loss (14.8%), Nonverbal IQ & Aging (9.4%), Verbal IQ (5.9%), and Temporal Processing (3.8%). A total of 33.9% of the total variance is accounted for by these four factors, which is quite a bit less than that

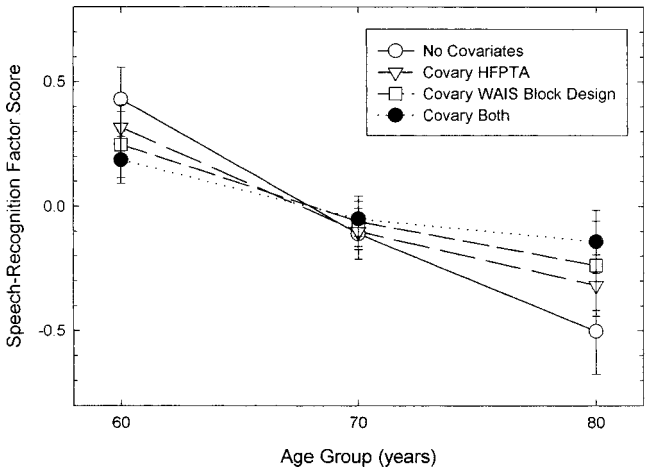


FIG. 13. Estimated marginal means from the GLM analysis of the effects of age group on the standardized general speech-recognition factor scores. Closed circles depict means for no covariates; triangles, for covariation of HFPTA with age group; squares, for covariation of nonverbal cognitive performance, as measured by the WAIS Block Design scale; and closed circles, for covariation of both HFPTA and WAIS block-design performance. Thin vertical lines above and below each symbol represent one standard error about the corresponding means.

accounted for in the entire group and with similar factors. This is to be expected since the variance of the dependent measure has been reduced considerably by selecting only those participants with milder amounts of hearing loss. Nonetheless, audibility or hearing loss still accounts for nearly half of the systematic variance, with cognitive factors, both age-related and non-age-related, accounting for the bulk of the remaining systematic variance.

In all of the regression analyses performed in this study, a predictive factor representing the combined influence of age and nonverbal IQ emerged as the second most powerful predictor of aided and unaided speech-recognition performance (after the factor associated with hearing loss). As noted previously, in the context of the SII or similar acoustical indices, others have suggested that the resulting AI or SII values should be corrected based on the listener's age (Studebaker *et al.*, 1997; Sherbecoe and Studebaker, 2002b). Such correction factors often have been established by examining the dependence of the errors in AI- or SII-predicted speech-recognition performance on listener age for various age groups. Generally, such corrections have been advocated only for listeners 70 years of age or older.

How much of the observed age effect is due to age *per se* or to correlated changes in either high-frequency hearing loss (Fig. 1) or nonverbal IQ? To address this issue, the 171 participants in this study were grouped into three age groups: 60–69 years ($N=55$), 70–79 years ($N=85$), and 80–89 years ($N=30$). Between-group GLM analyses were then conducted to examine the effects of age on the general speech-recognition factor score. Figure 13 provides a plot of the estimated marginal means from the GLM analysis for the speech-recognition factor score as a function of age group. Since factor scores have a mean of 0 and standard deviation of 1.0, the absence of an effect of age on the speech-recognition factor score would be represented in Fig. 13 as a horizontal line at 0. The open circles depict the means for each age group without covariation of any other variables. The effect of age was found to be significant and each age group differed significantly from the other two ($p<0.05$ for t tests with Bonferroni adjustment of p for multiple comparisons). The triangles in Fig. 13 depict the estimated marginal means when HFPTA served as a covariate. The means for each age group were evaluated at HFPTA (average of right and left ears combined)=50.9 dB HL. The mean for the 60-year-old listeners was significantly higher than that of both of the other age groups, but the means for the two older groups no longer differed. Thus, controlling for HFPTA eliminated the significant age effect for those 70 years of age or older. The squares in Fig. 13 depict the estimated marginal means for the speech-recognition factor score when performance on one of the nonverbal-IQ measures from the WAIS-R, the block-design scale, was the only covariate. The estimated marginal means for the speech-recognition factor score (estimated at a block-design score of 25.75) no longer differed significantly as a function of age group. Thus, when differences in nonverbal IQ were controlled, there was no significant effect of age on the speech-recognition factor score. Finally, covarying both HFPTA and block-design scores moved the estimated marginal means still closer toward a flat line at 0 and also resulted in no significant effect of age on the speech-recognition factor score. The results in Fig. 13, therefore, indicate that there is indeed an effect of age on the speech-recognition factor score, but that this effect is due to underlying changes in cognitive function (block design) and, to a lesser extent, hearing loss (HFPTA). In the regression analyses conducted previously in this study [Eq. (1)], since the effects of HFPTA were already accounted for in the regression equations by the Hearing Loss factor, it appears likely that the underlying age-related changes in general speech-recognition performance observed here were mediated through associated cognitive changes (nonverbal IQ).

It should be noted that Eq. (1) only predicts the relative standing of a given elderly hearing-impaired individual among a population of such individuals and does so in terms of general speech-recognition ability. A predicted SR value of 0, for example, simply means that the elderly listeners scored the same as the ''average'' elderly hearing-impaired listener. This is useful in identifying factors underlying individual differences in performance, as in this study, but is not helpful when attempting to evaluate the score of a particular hearing-aid wearer for a specific test condition. The work of Studebaker and colleagues in this regard appears particularly promising (Studebaker Gray, and Branch, 1999).

VII. SUMMARY

The following list summarizes the main findings of this study of 171 ''typical'' elderly hearing-impaired listeners fit with linear amplification:

(1) both unaided and aided measures of speech-recognition performance, obtained for a variety of materials and listening conditions, were related to a single, underlying speech-recognition factor;

(2) approximately two-thirds of the variance in general speech-recognition ability could be accounted for using a linear multiple-regression approach whereas only about half of the variance was accounted for by the SII (without age corrections);
(3) the predominant contributor to the regression equation was a measure of hearing loss or speech audibility (SII); and
(4) additional variance was accounted for by the inclusion of age-related (non-verbal) and non-age-related (verbal) measures of cognitive function.

ACKNOWLEDGMENTS

The author would like to thank Dana Wilson, Nancy Barlow, Carolyn Garner, and Amy Arthur for their assistance with data collection, and Sneha Patel, Mini Narendran, Brian Gygi, Andy Humes, and Kevin Caudill for their assistance with data entry and analysis. The author also thanks Christine Mullen and David Reetz for their help with the WAIS-R testing. This research was supported, in part, by Research Grant No. R01-AG08293 from the National Institute on Aging.

ANSI (**1991**). ANSI S3.1-1991, ''Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms'' (American National Standards Institute, New York).
ANSI (**1996**). ANSI S3.6-1996, ''Specifications for Audiometers'' (American National Standards Institute, New York).
ANSI (**1997**), ANSI S3.5-1977 ''Methods for the Calculation of the Speech Intelligibility Index'' (American National Standards Institute, New York).
Byrne, D., Parkinson, A., and Newall, P. (**1990**). ''Hearing aid gain and frequency response requirements for the severely/profoundly hearing impaired,'' Ear Hear. **11**, 40–49.
Christopherson, L. A., and Humes, L. E. (**1992**). ''Some psychometric properties of the Test of Basic Auditory Capabilities (TBAC),'' J. Speech Hear. Res. **35**, 929–935.
Cox, R. M., Alexander, G. C., Gilmore, C. G., and Pusakulich, K. M. (**1988**). ''Use of the Connected Speech Test (CST) with hearing-impaired listeners,'' Ear Hear. **9**, 198–207.
Divenyi, P. L., and Haupt, K. M. (**1997a**). ''Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. I. Age and lateral asymmetry effects,'' Ear Hear. **18**, 42–61.
Divenyi, P. L., and Haupt, K. M. (**1997b**). ''Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. II. Correlational analysis,'' Ear Hear. **18**, 100–113.
Divenyi, P. L., and Haupt, K. M. (**1997c**). ''Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. III. Factor representation,'' Ear Hear. **18**, 189–201.
Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York).
Fletcher, H., and Galt, R. H. (**1950**). ''The perception of speech and its relation to telephony,'' J. Acoust. Soc. Am. **22**, 89–151.
Frank, T., and Richards, W. D. (**1991**). ''Hearing aid coupler output level variability and coupler correction levels for insert earphones,'' Ear Hear. **12**, 221–227.
French, N. R., and Steinberg, J. C. (**1947**). ''Factors governing the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **19**, 90–119.
Gorsuch, R. L. (**1983**). *Factor Analysis*, 2nd ed. (Lawrence Erlbaum, Hillsdale, NJ).
Hawkins, D. B., Walden, B. E., Montgomery, A., and Prosek, R. (**1987**). ''Description and validation of an LDL procedure designed to select SSPL90,'' Ear Hear. **8**, 162–169.
Humes, L. E. (**1986**). ''An evaluation of several rationales for selecting hearing aid gain,'' J. Speech Hear. Disord. **51**, 272–281.
Humes, L. E. (**1991**). ''Prescribing gain characteristics of linear hearing aids,'' in *The Second Vanderbilt/VA Hearing Aid Report*, edited by G. A. Studebaker, F. H. Bess, and L. Beck (York, Parkton, MD).
Humes, L. E. (**1996**). ''Evolution of prescriptive fitting approaches,'' Am. J. Audiol. **5**, 19–23.
Humes, L. E., Barlow, N. N., Garner, C. B., and Wilson, D. L. (**2000**). ''Prescribed clinician-fit versus as-worn coupler gain in a group of elderly hearing aid wearers,'' J. Speech Lang. Hear. Res. **43**, 879–892.
Humes, L. E., and Christopherson, L. (**1991**). ''Speech-identification difficulties of hearing-impaired elderly persons: The contributions of auditory-processing deficits,'' J. Speech Hear. Res. **34**, 686–693.
Humes, L. E., Coughlin, M., and Talley, L. (**1996**). ''Evaluation of the use of a new compact disc for auditory perceptual assessment in the elderly,'' J. Am. Acad. Audiol **7**, 419–427.
Humes, L. E., Dirks, D. D., Bell, T. S., Ahlstrom, C., and Kincaid, G. E. (**1986**). ''Application of the Articulation Index and the Speech Transmission Index to the recognition of speech by normal-hearing and hearing-impaired listeners,'' J. Speech Hear. Res. **29**, 447–462.
Humes, L. E., Dirks, D. D., Bell, T. S., and Kincaid, G. E. (**1987**). ''Recognition of nonsense syllables by hearing-impaired listeners and noise-masked normal hearers,'' J. Acoust. Soc. Am. **81**, 765–773.
Humes, L. E., Garner, C. B., Wilson, D. L., and Barlow, N. N. (**2001**). ''Hearing-aid outcome measures following one month of hearing aid use by elderly participants,'' J. Speech Lang. Hear. Res. **44**, 469–486.
Humes, L. E., and Halling, D. C. (**1993**). ''Overview, rationale and comparison of suprathreshold-based gain-prescription procedures,'' in *Strategies for Selecting and Verifying Hearing Aid Fittings*, edited by M. Valente (Thieme Medical, New York).
Humes, L. E., and Roberts, L. (**1990**). ''Speech-recognition difficulties of the hearing-impaired elderly: The contributions of audibility,'' J. Speech Hear. Res. **33**, 726–735.
Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., and Lee, L. (**1994**). ''Factors associated with individual differences in clinical measures of speech recognition among the elderly,'' J. Speech Lang. Hear. Res. **37**, 465–474.
Humes, L. E., Wilson, D. L., Barlow, N. N., and Garner, C. B. (**2002**). ''Changes in hearing-aid benefit following one or two years of hearing aid use by the elderly,'' J Speech Lang. Hear. Res. (in press).
Jerger, J., Jerger, S., and Pirozzolo, F. (**1991**). ''Correlational analysis of speech audiometric scores, hearing loss, age and cognitive abilities in the elderly,'' Ear Hear. **12**, 103–109.
Jerger, J., and Johnson, K. (**1988**). ''Interactions of age, gender and sensorineural hearing loss on ABR latency,'' Ear Hear. **9**, 168–176.
Kalikow, D. N., Stevens, K. N., and Elliot, 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.
Levitt, H., and Resnick, S. B. (**1978**). ''Speech reception by the hearing impaired: Methods of testing and development of materials,'' Scand. Audiol. Suppl. **6**, 107–129.
Loven, F. (**1985**). Some effects of reverberation, filtering, masking and speech level on the speech-identification ability of normal-hearing listeners. Unpublished Ph.D. dissertation, The University of Iowa, Iowa City, IA.
Moscicki, E., Elkins, E., Baum, H., and McNamara, P. (**1985**). ''Hearing loss in the elderly: An epidemiologic study of the Framingham Heart Study Cohort,'' Ear Hear. **6**, 184–190.
Musiek, F. E. (**1994**). ''Frequency (pitch) and duration pattern tests,'' J. Am. Acad. Audiol **5**, 265–268.
Noffsinger, D., Martinez, C., and Wilson, R. H. (**1994**). ''Dichotic listening to speech: Background and preliminary data for digits, sentences, and nonsense syllables,'' J. Am. Acad. Audiol **5**, 248–254.
Noffsinger, D., Wilson, R. H., and Musiek, F. E. (**1994**). ''Department of Veterans Affairs Compact Disc (VA-CD) recording for auditory perceptual assessment: Background and introduction,'' J. Am. Acad. Audiol **5**, 231–235.
Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). ''An articulation-index based procedure for predicting the speech-recognition performance of hearing-impaired individuals,'' J. Acoust. Soc. Am. **80**, 50–57.
Pearsons, K. S., Bennett, R. L., and Fidell, S. (**1977**). ''Speech levels in various noise environments,'' EPA Report No. 600/1-77-025.
Schuknecht, H. (**1964**). ''Further observations on the pathology of presbycusis,'' Arch. Otolaryngol. **80**, 369–382.
Sherbecoe, R. L., and Studebaker, G. A. (**2002a**). ''Audibility-index functions for the Connected Speech Test,'' Ear Hear. (in press).
Sherbecoe, R. L., and Studebaker, G. A. (**2002b**). ''Audibility-index predic-

tions of normal-hearing and hearing-impaired listeners' performance on the Connected Speech Test,'' Ear Hear. (in press).
Starr, A., Picton, T. W., Sininger, Y. S., Hood, L. J., and Berlin, C. I. (**1996**). ''Auditory neuropathy,'' Brain **119**, 741–753.
Studebaker, G. A. (**1985**). ''A rationalized arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.
Studebaker, G. A., Gray, G. A., and Branch, W. E. (**1999**). ''Prediction and statistical evaluation of speech recognition test scores,'' J. Am. Acad. Audiol. **10**, 355–370.
Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gray, G. A. (**1997**). ''Age-related changes in monosyllabic word recognition performance when audibility is held constant,'' J. Am. Acad. Audiol **18**, 150–162.
Thornton, A. R., and Raffin, M. J. (**1978**). ''Speech discrimination scores modeled as a binomial variable,'' J. Speech Hear. Res. **21**, 507–518.
Tillman, T. W., and Carhart, R. (**1966**). ''An expanded test for speech discrimination utilizing CNC Monosyllabic Words (N.U. Auditory Test No. 6),'' Technical Report No. SAM-TR-66-55, USAF School of Aerospace Medicine, Brooks Air Force Base, TX.
Van Rooij, J. C. G. M., and Plomp, R. (**1990**). ''Auditive and cognitive factors in speech perception by elderly listeners. II. Multivariate analyses,'' J. Acoust. Soc. Am. **88**, 2611–2624.
Van Rooij, J. C. G. M., and Plomp, R. (**1992**). ''Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion,'' J. Acoust. Soc. Am. **91**, 1028–1033.
Van Rooij, J. C. G. M., Plomp, R., and Orlebeke, J. F. (**1989**). ''Auditive and cognitive factors in speech perception by elderly listeners. I. Development of test battery,'' J. Acoust. Soc. Am. **86**, 1294–1309.
Walden, B. E. (**1997**). ''Toward a model clinical-trials protocol for substantiating hearing aid user-benefit claims,'' Am. J. Audiol., **6**, 13–24.
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 (Lawrence Erlbaum, Hillsdale, NJ).
Wechsler, D. (**1981**). *The Wechsler Adult Intelligence Scale-Revised* (The Psychological Corporation, New York).
Wilson, R. H., Preece, J. P., Salmon, D. L., Sperry, J. L., and Bornstein, S. P. (**1994**). ''Effects of time compression plus reverberation on the intelligibility of Northwestern University Auditory Test No. 6,'' J. Am. Acad. Audiol **5**, 269–277.

Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies[a)]

Thomas Baer, Brian C. J. Moore,[b)] and Karolina Kluk[c)]
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 29 January 2002; revised 9 May 2002; accepted 14 June 2002)

People with high-frequency sensorineural hearing loss differ in the benefit they gain from amplification of high frequencies when listening to speech. Using vowel–consonant–vowel (VCV) stimuli in quiet that were amplified and then low pass filtered with various cutoff frequencies, Vickers *et al.* [J. Acoust. Soc. Am. **110**, 1164–1175 (2001)] found that the benefit from amplification of high-frequency components was related to the presence or absence of a cochlear dead region at high frequencies. For hearing-impaired subjects without dead regions, performance improved with increasing cutoff frequency up to 7.5 kHz (the highest value tested). Subjects with high-frequency dead regions showed no improvement when the cutoff frequency was increased above about 1.7 times the edge frequency of the dead region. The present study was similar to that of Vickers *et al.* but used VCV stimuli presented in background noise. Ten subjects with high-frequency hearing loss, including eight from the study of Vickers *et al.*, were tested. Five had dead regions starting below 2 kHz, and five had no dead regions. Speech stimuli at a nominal level of 65 dB were mixed with spectrally matched noise, amplified according to the ''Cambridge'' prescriptive formula for each subject and then low pass filtered. The noise level was chosen separately for each subject to give a moderate reduction in intelligibility relative to listening in quiet. For subjects without dead regions, performance generally improved with increasing cutoff frequency up to 7.5 kHz, on average more so in noise than in quiet. For most subjects with dead regions, performance improved with cutoff frequency up to 1.5–2 times the edge frequency of the dead region, but hardly changed with further increases. Calculations of speech audibility using a modified version of the articulation index showed that application of the Cambridge formula was at least partially successful in making high-frequency components of the speech audible for subjects with dead regions, and that such subjects often failed to benefit from increased audibility of the speech at high frequencies. [DOI: 10.1121/1.1498853]

PACS numbers: 43.71.Ky, 43.72.Kb, 43.66.Ts, 43.66.Sr [CWT]

I. INTRODUCTION

There have been reports over a period of many years suggesting that people with moderate-to-severe hearing loss at high frequencies often do not benefit from amplification of high frequencies, or even perform more poorly when high frequencies are amplified (Villchur, 1973; Moore *et al.*, 1985; Murray and Byrne, 1986; Hogan and Turner, 1998; Ching *et al.*, 1998; Turner and Cummings, 1999; Amos and Humes, 2001); for reviews see Moore (2001) and Vickers *et al.* (2001). These studies clearly show that, when the hearing loss exceeds about 55 dB at high frequencies, amplification of high frequencies is often not beneficial, although it usually is beneficial for people with milder losses (Skinner and Miller, 1983; Vickers *et al.*, 2001).

The lack of benefit of high-frequency amplification, when found, has often been interpreted in terms of damage to inner hair cells (IHCs) in the basal region of the cochlea. When the IHCs or neurons over a specific region of the cochlea are not functioning, we refer to this as a dead region (Moore and Glasberg, 1997; Moore *et al.*, 2000; Moore, 2001). It has been suggested that subjects who do not benefit from amplification have dead regions (or at least extensive damage to the IHCs) at high frequencies, while subjects who do benefit from amplification have surviving IHCs and neurons with high CFs.

Vickers *et al.* (2001) were among the first to examine the benefits of high-frequency amplification for people with diagnosed dead regions. Both psychophysical tuning curves (PTCs) (Moore and Alcántara, 2001) and the TEN test (Moore *et al.*, 2000) were used to detect and define the limits of any dead regions. The TEN test involves measurement of masked thresholds for pure tones presented in a threshold-equalizing noise. An abnormally high threshold at a specific frequency is taken as indicating a dead region at that frequency. All subjects tested by Vickers *et al.* had high-frequency hearing loss, but some had high-frequency dead regions and some did not; generally, the subjects with dead regions had more severe high-frequency hearing losses than those without dead regions. The speech stimuli were vowel–

[a)]Small portions of the data described in this paper were presented in Baer *et al.* (2002).
[b)]Electronic mail: bcjm@cus.cam.ac.uk
[c)]Current address: Institute of Acoustics, Adam Mickiewicz University, 85 Umultowska Street, 61-614 Poznan, Poland.

TABLE I. Audiometric (air conduction) thresholds of the test ears of the subjects, in dB HL. An asterisk indicates frequencies falling within the estimated dead region. An arrow indicates that audiometric thresholds were too high to be measured with the available audiometer. The age of each subject (years) is also shown.

			Frequency (kHz)									
Subject	Ear	Age	0.25	0.5	0.75	1.0	1.5	2.0	3.0	4.0	6.0	8.0
JC	L	82	10	5		20	30	35	35	50	60	70
	R		20	10		20	30	35	40	40	60	70
KC	L	65	10	10		10	40	65	70	75	70	65
	R		10	10		10	15	45	60	60	60	55
VW	L	73	50	60		70		60		70	60	65
	R		40	55		70	75	65	75	85	75	90
MR	L	49	5	0		25	55	60	60	45	40	30
	R		5	0		10	65	75	70	65	55	45
DT	L	76	25	15		15		45		45		50
	R		20	10		5		60		65	60	60
CA	L	55	25	30	45	65*	110*	↑*	↑*	↑*	↑*	↑*
MW	L	60	30	55	60	65	85*	110*	110*	100*	↑*	↑*
RC	L	71	15	10		5	60*	75*		105*		90*
LR	R	48	10	55		65	80	90*	110*	95*	↑*	↑*
NC	L	76	20	25	30	35	60*	70*	95*	110*	↑*	↑*
	R		25	25	55	70*	90*	95*	95*	110*	↑*	↑*

consonant–vowel (VCV) nonsense syllables, using one of three vowels and 21 different consonants. In a baseline condition, subjects were tested using broadband stimuli with a nominal level of 65 dB SPL. Prior to presentation via Sennheiser HD580 earphones, the stimuli were subjected to the frequency-gain characteristic prescribed by the ''Cambridge'' formula (Moore and Glasberg, 1998; see the following details). The stimuli for all other conditions were initially subjected to this same frequency-gain characteristic. Then, the speech was low pass filtered with various cutoff frequencies.

For subjects without dead regions, performance generally improved progressively with increasing cutoff frequency up to the highest frequency tested (7.5 kHz). For subjects with dead regions, two patterns of performance were observed. For most subjects, performance initially improved with increasing cutoff frequency and then reached an asymptote. The asymptote was reached when the cutoff frequency was about 1.5–2 times the estimated edge frequency of the dead region. For a few subjects, performance initially improved with increasing cutoff frequency, and then worsened with further increases. This indicates that amplification of high frequencies impaired performance.

For the six ears without dead regions, the mean score for the broadband speech (77.2%) was significantly higher ($p = 0.015$) than the mean score for speech low pass filtered at 2000 Hz (65.1%). However, for nine ears with dead regions starting below 2000 Hz, the mean scores did not differ significantly for the broadband speech (48.5%) and for the speech low pass filtered at 2000 Hz (47.1%).

The study of Vickers *et al.* (2001) used speech stimuli presented in quiet. However, in everyday life, speech often occurs in the presence of background noise, and hearing-impaired people generally have the greatest difficulty when noise is present. It is not clear whether the results of low pass filtering would be the same in quiet and in the presence of background noise. In the present paper, we determined the effect of low pass filtering on the intelligibility of speech in noise, using subjects with high-frequency hearing loss. The TEN test (Moore *et al.*, 2000) was used to detect the presence of dead regions and PTCs were used to define the limits of any dead regions more precisely (Moore and Alcántara, 2001). Subjects both with and without dead regions were tested.

II. METHOD

A. Subjects

Ten subjects with high-frequency hearing loss were tested. Eight of them had previously served in the study of Vickers *et al.* (2001), where they were comprehensively tested for dead regions. Of these eight, five had high-frequency dead regions extending to frequencies below 2 kHz and the other three had no dead regions. Two additional subjects without dead regions were tested. They were assessed in the same way as the other subjects. For all subjects, losses were diagnosed as being sensorineural (presumably cochlear) based on normal tympanometry and lack of an air-bone gap. Testing was done separately for each ear, but some subjects were tested using one ear only. In total, six ears with dead regions and ten ears without dead regions were tested. The audiograms of the ears tested are given in Table I. Table I also shows the ages of the subjects. For all of the subjects with dead regions, the dead region appeared to extend from the estimated edge frequency up to the highest frequency that was evaluated (10 kHz). Generally, the subjects with dead regions had more severe high-frequency hearing losses than those without dead regions. Audiometric frequencies falling within the diagnosed dead regions are indicated by asterisks in Table I. Edge frequencies of the dead regions, estimated from the PTC measurements, are given in the figures showing the data (presented later).

B. Stimuli and conditions

The speech stimuli were the same as those used by Vickers *et al.* (2001), namely, vowel–consonant–vowel (VCV)

nonsense syllables, using one of three vowels (/i/, /a/ or /u/) and 21 different consonants. These were: /p, t, k, b, d, g, f, θ, s, ʃ, h, v, z, r, l, j, w, tʃ, dʒ, n, and m/. The syllables were spoken by a female speaker with a British accent. The initial vowel was always the same as the final vowel. Each combination of vowel and consonant was presented once in each list in a randomized order, giving a total of 63 tokens per list. There were four different tokens of each VCV combination; the tokens used varied across lists. Twelve randomized lists were prepared. All lists were recorded on one channel of a CD and were replayed from the CD for presentation to the subjects.

Data were collected in three stages. In the first stage, stimuli were presented in quiet with amplification and low pass filtering. This is the condition used by Vickers *et al.* (2001), so data were already available for 8 of the 10 subjects. For the other 2 subjects, the methods of Vickers *et al.* (2001) were followed. A summary of the stimulus processing methods follows.

Stimuli were presented monaurally via Sennheiser HD580 earphones. The nominal level of the broadband speech (7500 Hz bandwidth) was 65 dB SPL. Before presentation to a subject, the stimuli were subjected to the frequency-gain characteristic prescribed by the ''Cambridge'' formula (Moore and Glasberg, 1998). We refer to this as ''Camfiltering.'' The formula is:

$$IG(f) = 0.48\mathrm{HL}(f) + \mathrm{INT}(f), \tag{1}$$

where IG(f) is the insertion gain in dB at frequency f, HL(f) is the hearing loss in dB at frequency f, and INT(f) is an intercept whose value depends on frequency. For frequencies of 1000 Hz and above, the value of INT is approximately 0 dB, so the insertion gain is approximately 0.48 times the hearing loss. The goal of the frequency-dependent amplification was to restore audibility as far as possible, while avoiding excessive loudness.

The gain specified by the Cambridge formula was calculated and applied separately for each ear tested. The maximum insertion gain applied was 50 dB. If the Cambridge formula called for an insertion gain greater than 50 dB, then the insertion gain was limited to 50 dB. In practice, this occurred when the hearing loss at high frequencies was 105 dB or more. The gain limit was reached in the following cases: LR, for frequencies of 3000 Hz and above; CA, for frequencies of 1500 Hz and above; NC, both ears for frequencies of 4000 Hz and above; MW, for frequencies of 2000 Hz and above.

The amplified speech was presented to the subject either broadband (upper frequency limit 7500 Hz) or after low pass filtering with various cutoff frequencies. The cutoff frequencies were chosen separately for each ear tested to cover a range from slightly below to well above the estimated edge frequency of the dead region, when present. For subjects without a dead region, the cutoff frequencies were chosen to span the range 800–7500 Hz. The number of cutoff frequencies varied across subjects, depending on the amount of time that they were willing to be tested.

The HD580 earphones are intended to have a ''diffuse field'' response, i.e., they give a response at the eardrum similar to what would be obtained listening with an open ear in a diffuse sound field. As described in Vickers *et al.* (2001), measurements using a KEMAR manikin (Burkhard and Sachs, 1975) were used to determine deviations from the desired diffuse field response, and these deviations were compensated using digital filtering [Tucker–Davis Technologies (TDT) Power DAC]. The digital filter also implemented the Camfiltering and the low pass filtering.

For the second and third stages of the experiment, noise was mixed with the speech stimuli. Noise having the same long-term spectrum as the speech was synthesized for this purpose. The noise was unmodulated and had a Gaussian distribution of instantaneous amplitudes. A continuous sample of the noise was recorded along with each of the twelve lists on the second channel of a compact disk (CD).

In the second stage of the experiment, the 65 dB speech stimuli were combined with the noise at various levels in order to choose a signal-to-noise ratio (SNR) to be used in the third stage. Both speech and noise levels were specified in terms of their root-mean-square values. The relative level of the noise was controlled by a TDT PA4 attenuator, and the speech and noise were combined with a TDT SM3 summer. The mixed stimuli were subjected to Camfiltering and to low pass filtering, as before. For ears with dead regions, the low pass cutoff frequency was set to a frequency close to that giving the highest speech score in quiet, which was somewhat (20%–80%) above the estimated edge frequency of the dead region. For ears with no dead region, the cutoff frequency was chosen to be somewhat above the frequency at which the hearing loss began to increase rapidly. Table II shows the cutoff frequency used for each ear of each subject and percent correct scores for the different SNRs. Generally, the scores worsened as the SNR decreased, as would be expected. These scores were used to find a SNR at which consonant intelligibility was reduced by 10%–15% relative to that measured in quiet for the chosen cutoff frequency (shown in the right-most column). The underlining in Table II indicates the SNR chosen for each ear tested.

Conditions for the third stage were exactly like those for the first, except that the input speech stimuli were combined with noise, at the SNR determined in stage two, before Camfiltering and low pass filtering with various cutoff frequencies.

C. Procedure

Subjects were given a list of the possible consonants, with examples of their pronunciation. They were asked to respond by writing down the consonant they heard for each token, ignoring the vowel. They were given 10 min of practice at the beginning of each session and the first test list was considered practice and was not scored. At least two lists were used for each condition. These were always presented in different test sessions. In stages 1 and 3, the order of testing of the different cutoff frequencies was randomized in the first test session. This order was reversed for the second session to balance the effects of practice and fatigue. In some cases, especially when the results appeared to be unusually

TABLE II. Results of stage 2, showing mean scores for the test ears of the subjects for the chosen cutoff frequencies, for various SNRs. Underlining indicates the SNR chosen for stage 3 for the given ear of each subject.

Subject	Ear	Cutoff freq (Hz)	Signal-to-noise ratio (dB)											
			−3	0	3	5	6	7	9	10	12	15	30	No noise
JC	L	2400		48	55		59							70
JC	R	2000		45	48		59		58					62
KC	L	1400		48	48		54		59					59
KC	R	1200		44	46		49		52					59
VW	L	2000		48	53		57		60					63
VW	R	1400		35	43		40		40					46
MR	L	2000		50	56		63		64					71
MR	R	2000		58	57		65		67					68
DT	L	2000		50	49		58		59					68
DT	R	2000		58	67		65		67					70
CA	L	900		35	37		45			51			57	56
MW	L	1680	21	24	23		32		33					44
RC	L	2000	37	44	54		55		59					68
LR	R	2000		30		31		41		45	41	42		47
NC	L	1400		36	41	38				41				45
NC	R	980		21	29	30	26			32			36	42

variable, additional lists were used, up to a maximum of five. For the additional tests, a new random order was used for the different cutoff frequencies.

III. RESULTS

A. Overall percent correct

Figures 1 and 2 show the percentage correct scores on the VCV test, both in quiet (closed symbols) and in noise (open symbols), for subjects without dead regions; they are plotted as a function of the lowpass-filter cutoff frequency. Error bars indicate ±one standard deviation across runs. Figure 1 contains results for the three subjects (six ears) without any dead region who were tested by Vickers *et al.* (2001). For these subjects, the results in quiet were taken from that study. Figure 2 contains results for the additional two subjects (four ears), who were tested both in quiet and in noise in the present study. It also shows the overall mean scores across subjects without dead regions; means were taken only for cutoff frequencies that were used for at least eight of the ten ears tested.

Results in noise are similar in form to those in quiet; scores generally improve monotonically with increasing cutoff frequency, indicating that listeners without dead regions are able to make effective use of high-frequency information in speech. Subject DT is a slight exception in that his curves tended to flatten off above 3 kHz. In general, scores in noise are below those in quiet, as expected, although for subjects DT and KC, performance tended to be slightly better in noise at the highest one or two cutoff frequencies. This appears to reflect a practice effect; the noise condition always followed the quiet condition. With extended practice, DT and KT may have learned to make more effective use of the high-frequency information provided by Camfiltering (Gatehouse, 1992).

To quantify the extent to which the subjects without dead regions were able to make use of high-frequency information, scores for the speech in noise were compared for the broadband speech (7500 Hz bandwidth) and the speech low pass filtered at 2000 Hz. The mean scores for these two conditions were 79.1% and 55.2%, respectively. A one-way within-subjects analysis of variance (ANOVA) of the scores for these two conditions showed a significant effect of cutoff frequency; $F(1,9)=143.2$, $p<0.001$.

Figure 3 shows results both in quiet (closed symbols) and in noise (open symbols) for the six ears with dead regions. The estimated edge frequencies of the dead regions are given in each panel, and are also indicated by the arrows. All these ears were tested by Vickers *et al.* (2001), and the results in quiet are taken from that study. Results in noise show a similar pattern to those in quiet. Scores tend to increase with increasing cutoff frequency up to somewhat above the estimated edge frequency of the dead region. For higher cutoff frequencies, scores tend to remain more nearly constant or to decrease slightly. For one subject (NC left ear), the highest score is obtained at a frequency about three times the estimated edge frequency of the dead region, but in general a peak or plateau is reached within an octave of the edge of the dead region. For the two ears that show decreases in score at higher cutoff frequencies in the quiet condition (RC left ear, CA left ear), scores in noise are about the same as those in quiet at the highest cutoff frequencies, but are lower at lower frequencies. Thus, for these ears, amplification of the high-frequency components of speech has a deleterious effect on speech intelligibility, but the effect is smaller in noise than in quiet. In fact, for CA there is no clear decrease when noise is present. For the other four ears with dead regions, scores in noise at the highest cutoff frequencies remain significantly below scores in quiet. For subject LR, in fact, scores in noise show a small peak at a frequency of about 1.7 times the estimated edge frequency of the dead region and remain constant at higher frequencies, although scores in quiet actually increase somewhat up to the highest cutoff frequency.

To assess the extent to which the subjects with dead regions were able to make use of high-frequency informa-

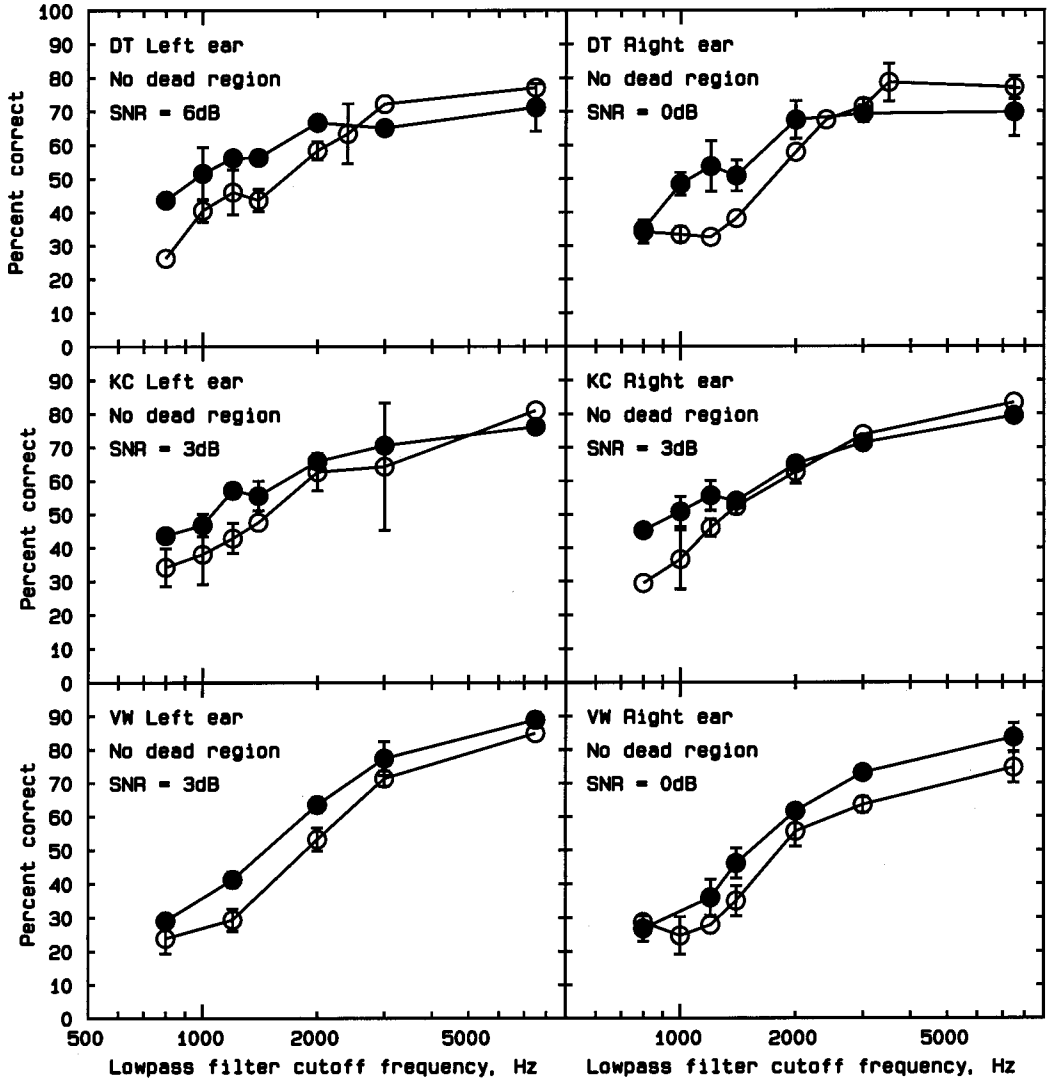


FIG. 1. Results for six ears without dead regions. The percentage correct score on the VCV test is plotted as a function of the low pass filter cutoff frequency. Scores in quiet and in noise are indicated by closed and open symbols, respectively. Error bars indicate ± one standard deviation across runs.

tion, scores for speech in noise were compared for the broadband speech (7500 Hz bandwidth) and the speech low pass filtered at 2000 Hz. The mean scores for these two conditions were 41.1% and 39.3%, respectively. A one-way within-subjects ANOVA of scores for these two conditions showed no significant effect of cutoff frequency; $F(1,5)=5.98$, $p=0.058$.

To assess the statistical significance of the worsening with increasing cutoff frequency found for the results in noise for LR and RC, scores were compared for the cutoff frequency giving the highest score and the cutoff frequency of 7500 Hz. This was done using both t-tests and chi-square tests (under the null hypothesis that the expected scores were equal for the two cutoff frequencies). The worsening was not significant in either case.

To give an overall impression of the pattern of results for the subjects with dead regions starting below 2 kHz, the cutoff frequencies used for each ear of the ears tested were expressed relative to the estimated edge frequency of the dead region for that ear. The data for each ear (percent correct versus relative frequency) were fitted with a cubic spline function. The cubic spline functions were then averaged across ears. This was done separately for the quiet and noise conditions; results are shown in the left-hand panel of Fig. 4 (broken and solid lines, respectively). On average, the noise has a roughly equal effect on intelligibility across the range of cutoff frequencies tested. For both noise and quiet conditions, average scores increase systematically with cutoff frequency up to relative frequencies between 1 and 2, and remain roughly constant at higher cutoff frequencies.

A similar analysis was applied to the results for the subjects without dead regions; these results are plotted in the right-hand panel of Fig. 4 as a function of absolute (not relative) frequency in kHz. For both conditions (quiet and noise), average scores increase monotonically with cutoff frequency. The difference between conditions decreases at higher cutoff frequencies; on average, scores in noise approach those in quiet for cutoff frequencies above 3 kHz. This may partly reflect the increasing ability of the subjects to make use of high-frequency information with increasing practice.

The left- and right-hand panels in Fig. 4 allow comparison of scores for subjects with and without dead regions. The two frequency scales are roughly comparable, as a relative

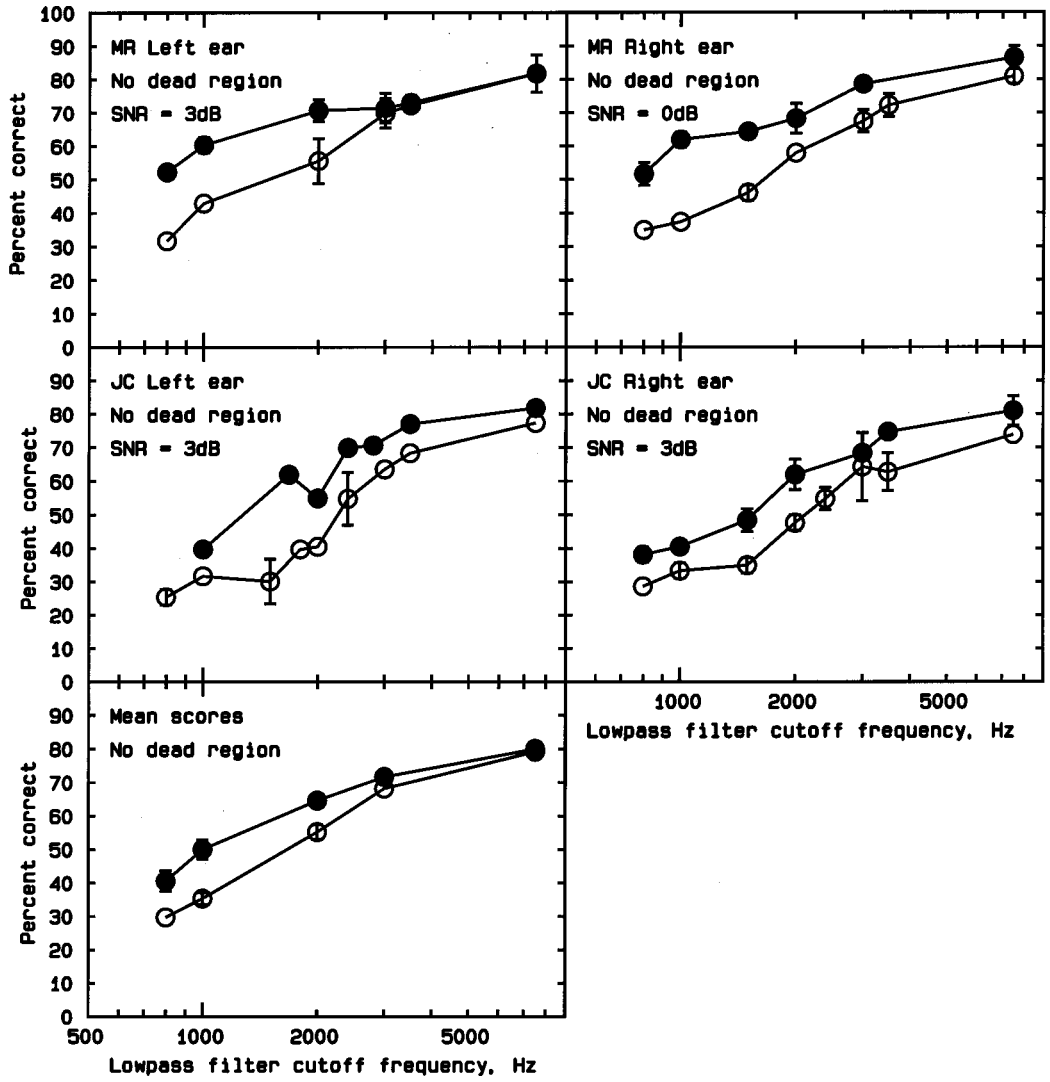


FIG. 2. As in Fig. 1, but showing results for a further four ears without dead regions. The lower left-hand panel shows mean scores across all subjects without dead regions; in this panel error bars indicate ± one standard error.

frequency of 1 corresponds, on average, to a frequency near 1 kHz; the geometric mean of the estimated edge frequencies of the dead regions was 1.007 kHz. In quiet (dashed lines), performance is similar for subjects with and without dead regions for the lowest abscissa values. The curves begin to diverge at an abscissa value of 1.3 to 1.4, and differ by 10% at about 1.7. In noise (solid lines), performance is essentially the same for the two groups up to abscissa values of 1.3–1.5, and the difference is again about 10% at 1.7. Thus, the cutoff frequencies for which scores begin to diverge for subjects with and without dead regions are about the same in quiet and in noise. For the subjects with dead regions, the fitted function increases with increasing relative frequency up to about 1.7, and then flattens off. This indicates that there is typically some benefit to intelligibility of amplifying frequencies up to about 70% above the estimated edge frequency of the dead region. For the subjects without dead regions, the fitted function increases progressively with increasing cutoff frequency, indicating that broadband amplification gives the best performance.

To assess the extent to which the subjects with dead regions were able to make use of information from frequencies just above the estimated edge frequency of the dead region for speech in noise, scores were compared for the low pass filter cutoff frequency closest to the estimated edge frequency, and for the cutoff frequency closest to 1.7 times the estimated edge frequency. The mean scores for these two conditions were 33.1% and 41.1%, respectively. A one-way within-subjects ANOVA of scores for these two conditions showed that the difference was significant; $F(1,5)=8.19$, $p=0.035$.

B. Transmission of phonetic features

It is of interest to determine which phonetic features are affected most by changes in bandwidth, and to determine whether the effect of bandwidth on transmission of phonetic features differs for subjects without and with dead regions. To examine these issues, the results for each ear of each subject were analyzed to determine the percentage information transmission for the phonetic features of voicing, place, and manner (Miller and Nicely, 1955) as a function of filter cutoff frequency. The method used was exactly as described by Vickers *et al.* (2001). The results were very similar to

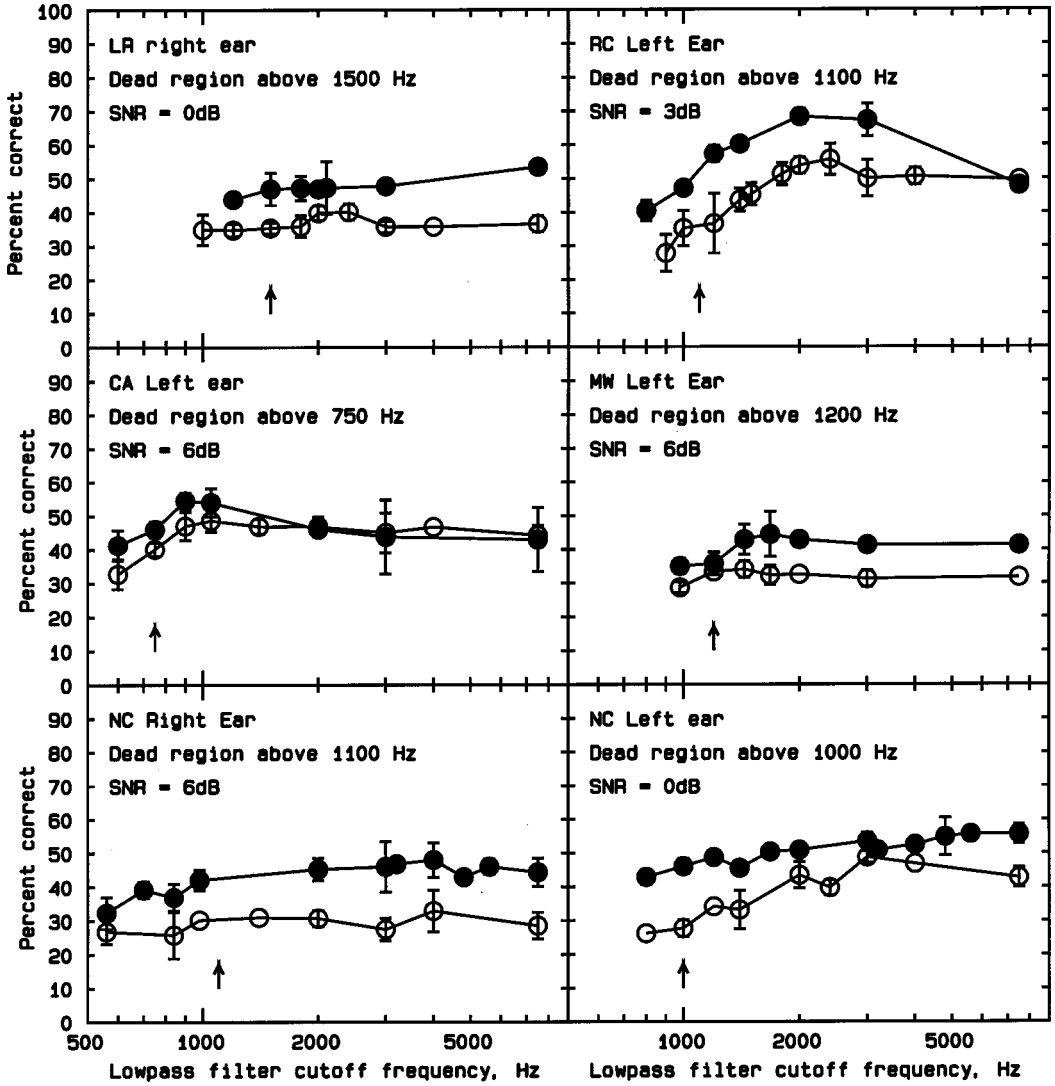


FIG. 3. Results for six ears with dead regions. Otherwise as in Fig. 1.

those presented by Vickers *et al.* (2001), so they will not be presented in detail. Briefly, the percentage of information transmitted was nearly always highest for voicing and lowest for place. Scores for voicing remained above 70% even for cutoff frequencies as low as 800 Hz, indicating that information about voicing can be extracted effectively from low-frequency components of speech. For the subjects without dead regions, scores for manner tended to increase with increasing cutoff frequency, and scores for place increased even more, at least for cutoff frequencies up to 3000 Hz. For the subjects with dead regions, scores for manner and place tended to improve initially with increasing cutoff frequency, but did not improve once the cutoff frequency was more than 50%–100% above the estimated edge frequency of the dead region.

C. Assessment of speech audibility

One important issue is the extent to which amplification according to the Cambridge formula was successful in providing audibility (Rankovic, 2002; Moore, 2002). It might be the case that the failure to find a benefit of amplification for

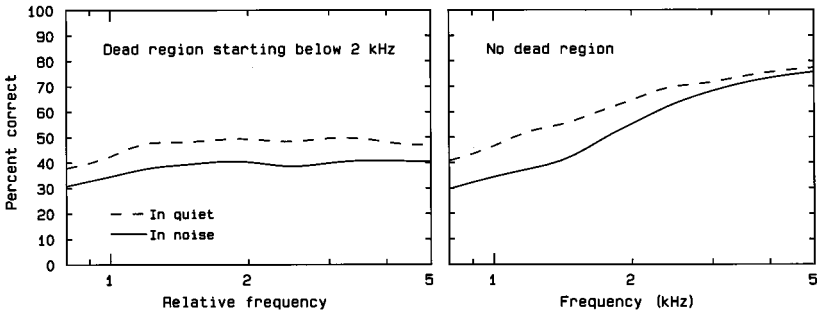


FIG. 4. The left-hand panel shows mean cubic-spline functions fitted to the data for subjects with dead regions, tested in quiet (dashed curve) and in noise (solid curve). Percent correct is plotted as a function of the filter cutoff frequency, relative to the estimated edge frequency of the dead region. The right-hand panel shows results for subjects without dead regions. In this case, the filter cutoff frequency is expressed in kHz.

frequencies more than a factor of 1.7 above the estimated edge frequency of a dead region was caused by insufficient amplification of those frequencies; if the speech is not made audible, there can be no benefit from amplification. To address this issue, we calculated a form of the articulation index (AI), using a modification of the procedure described by Moore and Glasberg (1998). This procedure was specifically designed for use with hearing-impaired people, and it takes into account the broadening of the auditory filters that typically occurs in people with cochlear hearing loss (Moore, 1998). When the filters are broader than normal, spectral energy is integrated over a wider range than normal, so the threshold for detecting a broadband sound may be lower than would be anticipated from the pure-tone audiogram.

In the procedure described by Moore and Glasberg (1998), the first stage is the calculation of an excitation pattern for an input with the same long-term average spectrum as the speech signals under consideration. For this purpose, we used the long-term average spectra of our VCV stimuli, taking into account the effects of the digital filtering used to implement the Cambridge formula and the low pass filtering, and the transfer function of the HD580 earphone. The excitation patterns were calculated using the procedure described by Moore and Glasberg (1997), except that the stage simulating the transfer function through the middle ear was modified to be the same as described in Moore *et al.* (1997). This procedure allows for the possibility of dead regions, by setting the excitation to a very low value within a specified dead region. However, for the present analyses, the program was ''told'' that no dead region was present. The excitation patterns were specified on an ERB scale (Glasberg and Moore, 1990). For points with equal ERB-spacing, we calculated the excitation level relative to the excitation required for threshold. From this we calculated the proportion of the speech dynamic range that was above threshold in a given frequency band, assuming that this range extends from 15 dB below the rms level to 15 dB above it (ANSI, 1997). For example, if the rms excitation is 15 dB above threshold, then the proportion is 1, while if the rms excitation is at threshold, the proportion is $15/30=0.5$, and if the rms excitation is 5 dB below threshold the proportion is $10/30=0.333$. Band importance functions were adapted from those for average speech as specified in the ''critical band SII procedure'' (ANSI, 1997), but values were adjusted to allow for the difference between traditional critical bandwidth values (Zwicker and Terhardt, 1980) and ERB values (Glasberg and Moore, 1990). The importance values are specified in Table 2 of Moore and Glasberg (1998). The AI was calculated by multiplying the proportion of the speech dynamic range that was above threshold in a given band, by the importance value for that band, and then summing across all bands from 189.7 Hz to 9.302 kHz.

This method of calculating the AI is based on the assumption that the audibility of excitation at a particular center frequency is closely related to the audibility of the speech spectrum at that frequency. In other words, we assume that the excitation at a specific frequency is determined by the speech energy close to that frequency, rather than the spread of excitation from adjacent frequencies. To check the validity of this assumption, we assessed the effect of sharply low pass filtering the input spectrum used to calculate the excitation pattern. In all cases, removal of spectral energy in a specific frequency region (for example by decreasing the lowpass cutoff frequency from 2 to 1.8 kHz) resulted in a reduction in the excitation level in that frequency region, confirming that the excitation was produced by ''local'' rather than by ''remote'' frequency components.

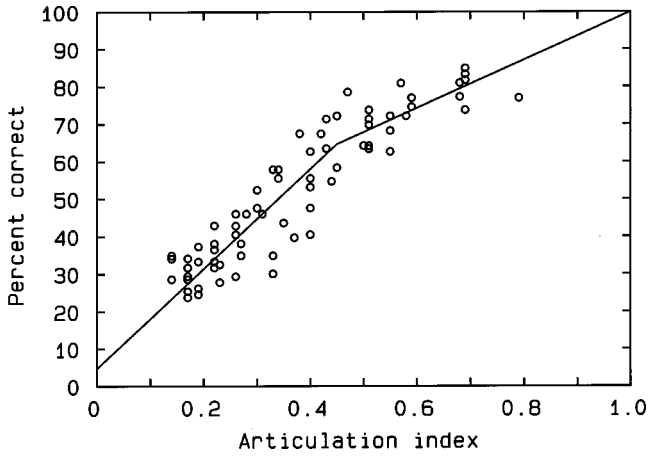


FIG. 5. Scatter plot of the scores for VCV syllables in noise as a function of the AI, for subjects without dead regions, calculated as described in the text. Each point represents one cutoff frequency for one subject. The two straight lines used to fit the data are also shown.

Moore and Glasberg (1998) did not apply their procedure for the calculation of the AI to the case of speech presented in noise. To extend their procedure to this case, we assumed that the noise limited the audibility of the lower part of the speech dynamic range. Experiments on the detection of tones in noise suggest that tones can be detected when the SNR at the output of the auditory filter centered on the signal frequency is about -3 dB (Patterson and Moore, 1986). We assumed that this was also the case for speech in noise. Thus, for example, when the SNR used with a given subject was 0 dB, we assumed that useful information could be extracted from the speech when the short-term speech level was 3 dB below that of the noise. For this example, the audible proportion of the speech dynamic range (assuming that the absolute threshold is not a limiting factor) is $18/30=0.6$. More generally, the audible proportion of the speech dynamic range in a specific frequency band was taken as the proportion above the absolute threshold, or as $(\mathrm{SNR}+18)/30$, whichever was the smaller.

Figure 5 shows a scatter plot of the percent correct scores for speech in noise as a function of the AI as calculated previously, for the subjects without dead regions. This plot can be considered as representative of the performance expected from a given AI value when a dead region is not present. The data were fitted with two straight lines. One line was constrained to pass though 4.76% (corresponding to the score obtainable by guessing) when $\mathrm{AI}=0$. The other line was constrained to pass through 100% when $\mathrm{AI}=1$, as data from a normally hearing subject indicated that a score of 100% was obtained for the wideband speech (a case when $\mathrm{AI}=1$). Least-squares minimization was used to find the best-fitting slopes of the pair of lines. The fitted lines were

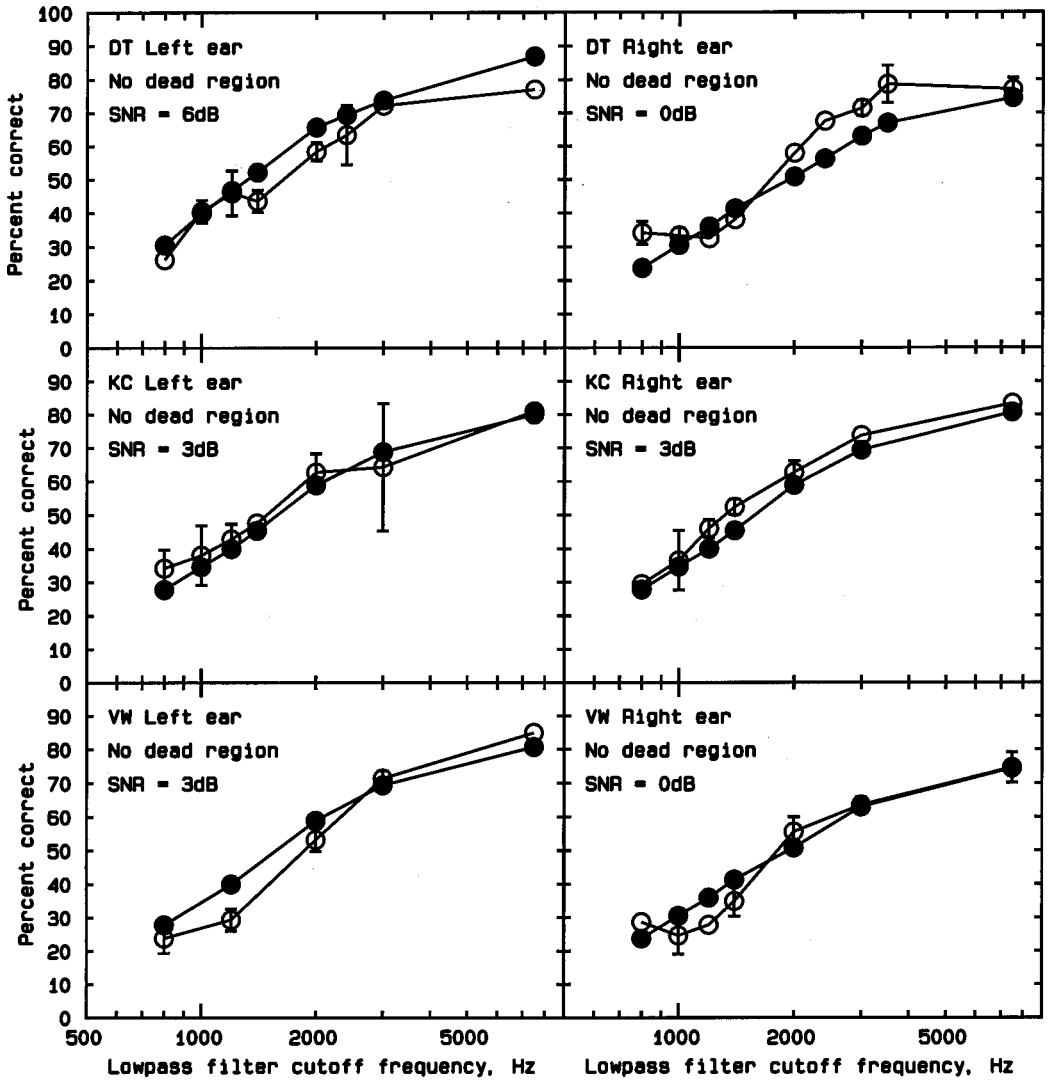


FIG. 6. Comparison of obtained scores (open circles) and scores predicted from the AI (closed circles) for six ears without dead regions.

then used to predict the results for the subjects without dead regions; the AI score for each subject and each cutoff frequency was used to derive a predicted percent correct value.

The obtained scores (open circles) and predicted scores (closed circles) for the subjects without dead regions are shown in Figs. 6 and 7. Generally, there is a good correspondence between the obtained and predicted scores. In particular, both the predicted scores and the obtained scores tend to increase monotonically and in a similar way with increasing cutoff frequency, although the obtained scores show some minor irregularities.

We now describe a test of the hypothesis that subjects with dead regions make as effective use of audible speech information as subjects without dead regions. If this is the case, then the fitted lines in Fig. 5 should correctly predict the pattern of scores for the subjects with dead regions. The obtained scores (open circles) and predicted scores (closed circles) for the subjects with dead regions are shown in Fig. 8. It is clear from the predicted scores that application of the Cambridge formula did lead to at least partial audibility of frequency components of the speech well inside the dead regions. The percent correct scores predicted from the AI values increase with increasing cutoff frequency for all subjects up to at least 2000 Hz, and for LR (right ear) and NC (left ear) predicted scores increase up to 4000 Hz.

It is also clear that there are substantial discrepancies between the obtained and predicted scores. First, for cutoff frequencies greater than or equal to 1500 Hz, the predicted scores are always above the obtained scores. The discrepancies for high cutoff frequencies range from about 10% to over 30%. These discrepancies are far larger than those for the subjects without dead regions, which range from 0 to 10%. The large discrepancies between obtained and predicted scores for subjects with dead regions indicate that these subjects do not make as effective use of audible high-frequency speech information as subjects without dead regions. Second, for all of the subjects with dead regions, when the cutoff frequency is above the estimated edge frequency of the dead region, there is a range of cutoff frequencies over which predicted scores increase but obtained scores either decrease or remain roughly constant. For example, for subject LR (right ear), the predicted score increases from 50% to 63% when the cutoff frequency is increased from 2500 to 7500 Hz, while the obtained score decreases from 40% to

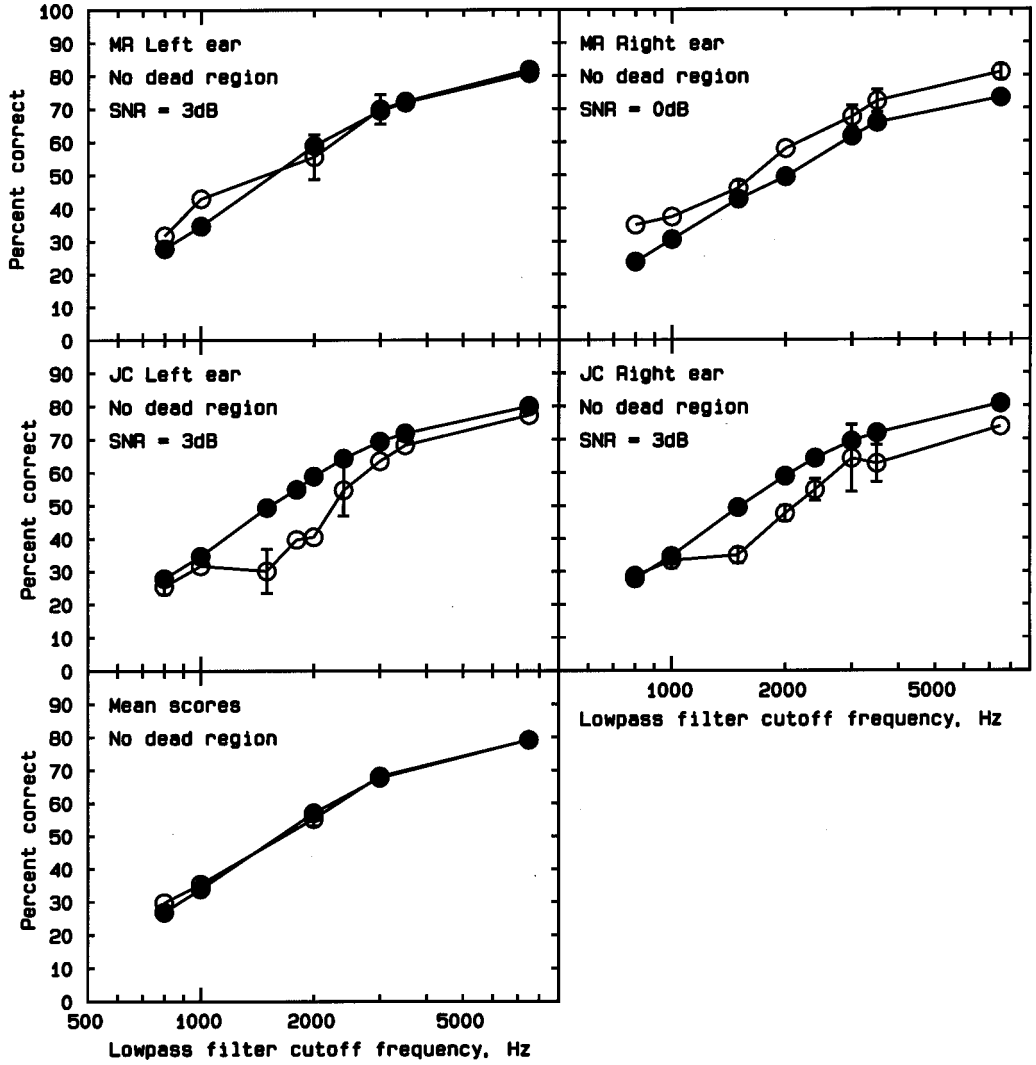


FIG. 7. As in Fig. 6, but showing results for a further four ears without dead regions. The lower left-hand panel shows mean obtained and predicted scores across all subjects without dead regions.

37% over the same range. For subject CA (left ear), the predicted score increases from 40% to 55% when the cutoff frequency is increased from 1000 to 7500 Hz, while the obtained score decreases from 47% to 44% over the same range. For subject NC (left ear), the predicted score increases from 52% to 59% when the cutoff frequency is increased from 1500 to 7500 Hz, while the obtained score decreases from 30% to 28% over the same range.

Overall, these results support the idea that subjects with dead regions at high frequencies do not make as effective use of audible speech information at high frequencies as subjects without dead regions. Furthermore, the results support the idea that increasing the audibility of speech for frequencies well inside a dead region does not lead to concomitant increases in speech intelligibility.

IV. DISCUSSION

Our results extend the findings of Vickers *et al.* (2001) by showing that, when trying to understand speech in noise, people with high-frequency hearing loss but without a dead region at high frequencies benefit from amplification of high frequencies, whereas people with a dead region generally do not benefit, except for frequencies that are 50%–100% above the estimated edge frequency of the dead region. Some possible reasons for the lack of benefit for the latter are discussed by Vickers *et al.* (2001).

Our results have important implications for the fitting of hearing aids to people with high-frequency dead regions. They suggest that there will be little or no benefit to speech perception from amplifying frequencies *well inside* a dead region. However, there may be some benefit in amplifying frequencies up to 50%–100% above the estimated edge frequency of the dead region. The present results suggest that the frequency up to which amplification should be applied is about the same for speech in quiet and speech in noise.

It should be emphasized that, for patients *without* high-frequency dead regions, amplification of the high frequencies can be (and usually is) beneficial. In the present study, most of the subjects without dead regions gave higher scores for the 7.5 kHz cutoff frequency than for the 3 kHz cutoff frequency. Furthermore, the results suggest that the ability of these subjects to use high-frequency information improves with extended practice. Hence, before deciding what form of amplification should be provided for a patient with high-

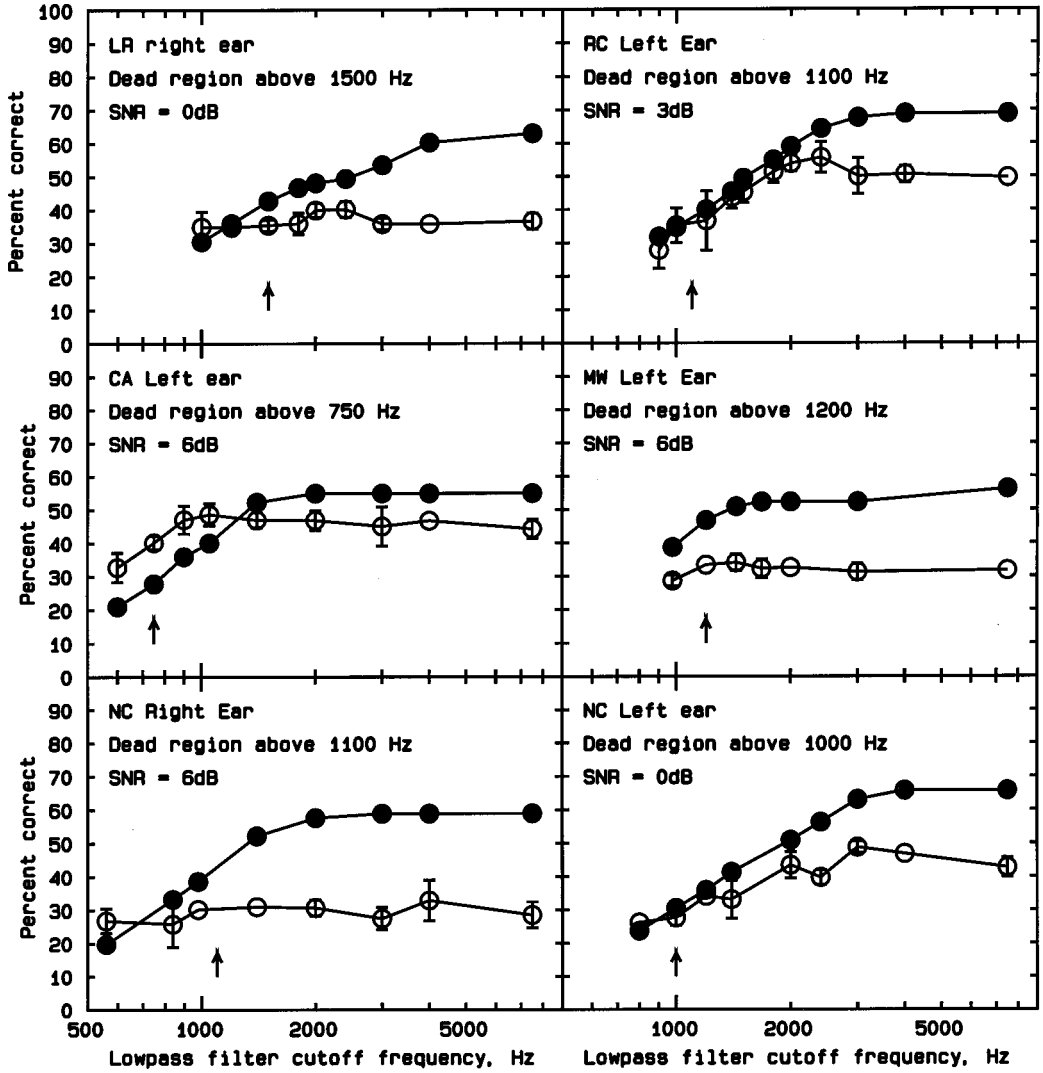


FIG. 8. As in Fig. 6, but for six ears with dead regions.

frequency hearing loss, it is important to determine whether that patient has a high-frequency dead region, and, if so, what its extent is. The TEN test (Moore *et al.*, 2000) is recommended for this purpose.

V. CONCLUSIONS

For VCV syllables presented in noise, the benefit of high-frequency amplification (according to the ''Cambridge'' formula) for subjects with high-frequency hearing loss differed according to whether or not there was a dead region at high frequencies. The pattern of results was similar to that found for speech in quiet by Vickers *et al.* (2001). For subjects without dead regions, identification scores improved progressively with increasing low pass filter cutoff frequency. For subjects with dead regions, scores improved with increasing cutoff frequency until the cutoff frequency was 50%–100% above the estimated edge frequency of the dead region. With further increases in cutoff frequency, scores remained roughly constant. Analyses based on a modified version of the articulation index indicated that amplification according to the Cambridge formula was at least partially successful in restoring the audibility of high-frequency components in the speech for subjects with dead regions, and that such subjects often failed to benefit from increased audibility of the speech at high frequencies.

ACKNOWLEDGMENTS

This work was supported primarily by the MRC (UK) with additional support from Starkey (USA), Defeating Deafness (UK), and the RNID (UK). We thank José Alcántara, Henry Chubb, Brian Glasberg, Martina Huss, Josephine Marriage, and Debi Vickers for their contributions to this work. We also thank Chris Turner for helpful comments.

Amos, N. E., and Humes, L. E. (**2001**). ''The contribution of high frequencies to speech recognition in sensorineural hearing loss,'' in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven (Shaker, Maastricht).

ANSI (**1997**). ANSI S3.5-1997, ''Methods for the Calculation of the Speech Intelligibility Index'' (American National Standards Institute, New York).

Baer, T., Moore, B. C. J., and Kluk, K. (**2002**). ''Effects of low pass filtering on the intelligibility of speech in noise for listeners with and without dead regions at high frequencies,'' Int. J. Audiol. (in press).

Burkhard, M. D., and Sachs, R. M. (**1975**). ''Anthropometric manikin for acoustic research,'' J. Acoust. Soc. Am. **58**, 214–222.

Ching, T., Dillon, H., and Byrne, D. (**1998**). ''Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification,'' J. Acoust. Soc. Am. **103**, 1128–1140.
Gatehouse, S. (**1992**). ''The time course and magnitude of perceptual acclimatization to frequency responses: Evidence from monaural fitting of hearing aids,'' J. Acoust. Soc. Am. **92**, 1258–1268.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Hogan, C. A., and Turner, C. W. (**1998**). ''High-frequency audibility: Benefits for hearing-impaired listeners,'' J. Acoust. Soc. Am. **104**, 432–441.
Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).
Moore, B. C. J. (**2001**). ''Dead regions in the cochlea: Diagnosis, perceptual consequences, and implications for the fitting of hearing aids,'' Trends Amplif. **5**, 1–34.
Moore, B. C. J. (**2002**). ''Response to 'Articulation index predictions for hearing-impaired listeners with and without cochlear dead regions,' '' J. Acoust. Soc. Am. (in press).
Moore, B. C. J., and Alcántara, J. I. (**2001**). ''The use of psychophysical tuning curves to explore dead regions in the cochlea,'' Ear Hear. **22**, 268–278.
Moore, B. C. J., and Glasberg, B. R. (**1997**). ''A model of loudness perception applied to cochlear hearing loss,'' Aud. Neurosci. **3**, 289–311.
Moore, B. C. J., and Glasberg, B. R. (**1998**). ''Use of a loudness model for hearing aid fitting. I. Linear hearing aids,'' Br. J. Audiol. **32**, 317–335.
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). ''A model for the prediction of thresholds, loudness and partial loudness,'' J. Audio Eng. Soc. **45**, 224–240.
Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcántara, J. I. (**2000**). ''A test for the diagnosis of dead regions in the cochlea,'' Br. J. Audiol. **34**, 205–224.
Moore, B. C. J., Laurence, R. F., and Wright, D. (**1985**). ''Improvements in speech intelligibility in quiet and in noise produced by two-channel compression hearing aids,'' Br. J. Audiol. **19**, 175–187.
Murray, N., and Byrne, D. (**1986**). ''Performance of hearing-impaired and normal hearing listeners with various high-frequency cut-offs in hearing aids,'' Aust. J. Audiol. **8**, 21–28.
Patterson, R. D., and Moore, B. C. J. (**1986**). ''Auditory filters and excitation patterns as representations of frequency resolution,'' in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).
Rankovic, C. M. (**2002**). ''Articulation index predictions for hearing-impaired listeners with and without cochlear dead regions,'' J. Acoust. Soc. Am. (in press).
Skinner, M. W., and Miller, J. D. (**1983**). ''Amplification bandwidth and intelligibility of speech in quiet and noise for listeners with sensorineural hearing loss,'' Audiology **22**, 253–279.
Turner, C. W., and Cummings, K. J. (**1999**). ''Speech audibility for listeners with high-frequency hearing loss,'' Am J. Audiol. **8**, 47–56.
Vickers, D. A., Moore, B. C. J., and Baer, T. (**2001**). ''Effects of lowpass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies,'' J. Acoust. Soc. Am. **110**, 1164–1175.
Villchur, E. (**1973**). ''Signal processing to improve speech intelligibility in perceptive deafness,'' J. Acoust. Soc. Am. **53**, 1646–1657.
Zwicker, E., and Terhardt, E. (**1980**). ''Analytical expressions for critical band rate and critical bandwidth as a function of frequency,'' J. Acoust. Soc. Am. **68**, 1523–1525.

A speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity

J. Lyzenga,[a] J. M. Festen, and T. Houtgast
Department of Otolaryngology, VU Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands

(Received 19 June 2001; accepted for publication 23 May 2002)

Hearing-impaired listeners often suffer from supra-threshold speech perception deficits. One such deficit is reduced frequency selectivity. We applied a speech enhancement scheme that incorporated spectral expansion in an attempt to reduce the effects of this deficit. The speech processing could contain up to three stages, a first in which the peak–valley ratio of the speech spectrum was enlarged to counteract the broadening of the auditory filtering, and a second in which the overall speech spectrum was modified to counteract the effects of upward-spread-of-masking, using a linear filter. The third stage was a noise suppression stage, applied before the spectral enhancement. The effectiveness of the speech processing with and without noise suppression was evaluated for various parameter settings by measuring the speech reception threshold (SRT) in noise, i.e., the signal-to-noise ratio at which listeners repeat 50% of presented sentences correctly. We used normal-hearing subjects. To simulate the loss of frequency selectivity we applied spectral smearing to the stimuli presented to the subjects. The speech material of the SRT tests was mixed with the noise before processing, and, when present, the smearing was applied last. The results indicated that for one specific parameter setting the SRT values decreased (i.e., improved) by approximately 1 dB when incorporating the spectral expansion together with the linear filtering. Employing either of these two stages separately did not improve the SRT. The application of the noise suppression stage did not further improve the SRT. A pilot study using hearing-impaired listeners showed more promising results for a female than for a male speaker.
[DOI: 10.1121/1.1497619]

PACS numbers: 43.72.Ew, 43.71.Es, 43.66.Ts [DOS]

I. INTRODUCTION

Hearing-impaired listeners frequently experience problems understanding speech, especially under noisy listening conditions. In addition to threshold elevation they often suffer from supra-threshold deficits (Plomp, 1978, 1986) such as loss of compression resulting in loudness recruitment (Huizing, 1948; Villchur, 1973), and reduced frequency selectivity resulting in a loss of detail in the internal excitation (Glasberg and Moore, 1986; Noordhoek *et al.*, 2000; van Schijndel *et al.*, 2001) Glasberg and Moore found considerably wider auditory filter shapes for hearing-impaired than for normal-hearing listeners. Noordhoek *et al.* measured the effect on speech reception of band-filtering speech in a band around 1 kHz for normal-hearing and hearing-impaired listeners. They found that the bandwidth needed for a 50%-correct score in sentences, presented above absolute threshold, was considerably larger for many hearing-impaired listeners than for the normal-hearing listeners. This indicates that these hearing-impaired listeners suffered from supra-threshold distortions. In the study of van Schijndel *et al.* a distortion sensitivity approach was used to investigate the amount of impairment-related distortion in the temporal, spectral, and intensity domain. For normal-hearing and hearing-impaired listeners, they measured the influence on speech reception of perturbations in each domain. At low perturbation levels, they found that speech reception thresholds were better for normal-hearing than for hearing-impaired listeners. This difference disappeared at high perturbation levels, which indicates that for the low perturbation levels the hearing-impaired listeners suffered from internal supra-threshold distortion, while at high perturbation levels the applied perturbations were dominant for all listeners. Such a distortion sensitivity effect was not observed in the temporal and the intensity domains.

In this study we investigate a scheme in which speech is processed prior to presentation in an attempt to alleviate intelligibility problems with speech in noise resulting from reduced frequency selectivity. Previous attempts using speech enhancement to improve speech intelligibility in noise as measured in word or sentence scores have not been very successful (Bunnell, 1990; Clarkson and Bahgat, 1991; Cheng and O'Shaughnessy, 1991; Stone and Moore, 1992; Baer *et al.*, 1993; Baer and Moore, 1997; Franck *et al.*, 1999). To enhance spectral contrast, Bunnell used an adaptive digital filter that expanded the spectral magnitudes between 0 and 5 kHz using a Hamming-window-shaped weight function. Using ten hearing-impaired subjects, he found very small, but significant, increases in consonant identification scores in quiet. Clarkson and Bahgat used a noise reduction scheme in which the temporal envelopes in separate frequency bands were expanded. They reported very small, but consistent, improvements in recognition scores at a signal-to-noise ratio (SNR) of 0 dB. Cheng and O'Shaughnessy used a noise reduction scheme using a digital filter that mimicked the behavior of spatial lateral inhibition between internal fre-

[a]Electronic mail: johannes.lyzenga@mrc-cbu.cam.ac.uk

quency channels. They achieved enhanced contrasts in short term signal spectra and substantial SNR improvements for speech in white noise, but did not perform any intelligibility measurements. Stone and Moore used an adaptive bank of hardware filters that behaved very much like an analog version of the digital method of Cheng and O'Shaughnessy. They achieved enhanced contrasts in short-term signal spectra, but no consistent improvements in intelligibility scores. Baer *et al.* used an adaptive digital difference-of-Gaussians filter to enhance spectral contrasts. They reported some small, but significant, increases in word-recognition scores for enhancement combined with a moderate amount of compression. Baer and Moore used a variation of the method used by Baer *et al.* In the adaptive filter of Baer and Moore the enhancement function was based on modeled excitation patterns. Using spectral smearing (as described in Baer and Moore, 1993) to simulate reduced spectral acuity in 12 normal-hearing listeners, they found a 5% increase in intelligibility scores. Franck *et al.* combined the scheme of Baer *et al.* (1993) with phonemic compression. They found for consonant–vowel–consonant (CVC) complexes that spectral enhancement could improve the scores for vowels, but degraded those for consonants. This provides one possible reason why enhancement schemes have not been very successful to date.

Another possible explanation for the limited successes of spectral enhancement schemes may lie in the limited increase in peak-to-valley ratios they can achieve without introducing a large amount of distortion. Leek *et al.* (1987) have shown that hearing-impaired listeners need more than three times larger peak-to-valley ratios (in dB) for correct vowel identification in quiet than normal-hearing listeners. Furthermore, in most natural vowels the peak of the first formant is more intense than that of the second, which in its turn is more intense than that of the third, the result of which is that upward-spread-of-masking (USOM) decreases the effective depth of the valleys between the formants. When applying a scheme for spectral enhancement the valley between these peaks can be deepened, but that will not produce a substantial reduction in USOM of the first formant on the second and of the second formant on the third. Recently, Miller *et al.* (1999) found in acoustically traumatized cat that auditory-nerve spiking patterns only represented the first formant of a presented vowel /ɛ/. After the second formant was raised substantially in level (+30 dB), firing patterns representing this formant reappeared. So, some reduction of the effects of USOM might be achieved by adapting the spectral shape of speech in such a way that the levels of the higher formants are enhanced relative to that of the first formant. Such a process will have to be progressive, i.e., enhance the second formant relative to the first, the third relative to the enhanced second, and likewise enhance the fourth relative to the enhanced third formant. In the present study we have added a stage of linear filtering that progressively emphasizes a frequency region that roughly spans the region from the second to the fourth formant, while avoiding the introduction of a very sharp overall sound quality from too much emphasis of high-frequency regions.

When attempting to increase the relative energy of formants or, in other words, to decrease the relative energy between the formants, the presence of noise can reduce the efficiency of the enhancement algorithm. When noise fills the gaps between the formants, the effective formant amplitude will be adversely affected and enhancement schemes will be less effective. In the second experiment, we added a noise suppression stage before the enhancement stage in an attempt to avoid this problem. Many attempts at monaural noise reduction have used techniques that use estimates of spectral properties of the noise in the suppression process, such as spectral subtraction, Wiener filtering, or alternatives that minimize mean-square errors (e.g., Boll, 1979; McAulay and Malpass, 1980; Ephraim and Malah, 1984). With such algorithms, various rather large degrees of SNR improvements have been reached, but often at the cost of clearly audible artifacts, and speech intelligibility has seldom been improved. For a recent overview, see Levitt (2001). The present noise suppression is a novel scheme in which the consistency, over a period of 29 to 58 ms, of consecutive short-term Fourier spectra is used to estimate the importance of signal components. Most noise suppression schemes, which use some form of spectral attenuation of noise, incorporate a separate noise estimator and speech-pause detector and attempt to update the estimate of the noise spectrum during gaps in the speech. The main disadvantage of such schemes is the delay present in the estimate of the noise, which can cause switching problems when the noise changes in character (which happens very often in every day listening situations). In the present scheme that problem is avoided by using a much smaller delay in the noise estimation, and a closer proximity of noise estimator and noise filter in the algorithm. To minimize distortion of the speech sounds and other artifacts, we do not aim for a large increase in the SNR. The goal is to achieve an improvement of just a couple of dB, mainly for the benefit of the spectral enhancement scheme. Expectations of the present noise suppression scheme are discussed at the end of Sec. III.

The effect of the speech processing is evaluated by measuring the speech reception threshold (SRT) for sentences in noise. This test estimates the SNR at which listeners are able to correctly reproduce 50% of the presented sentences. To simulate, in our normal-hearing listeners, the hearing impairments resulting from a loss of frequency selectivity, we spectrally smeared the stimuli using a separate computer program that was applied after the processing and before presentation of the stimuli. Earlier studies (ter Keurs *et al.*, 1992, 1993; Baer and Moore, 1993, 1994) have shown that such smearing schemes affect speech perception in noise in normal-hearing listeners in a way that is not unlike the problems experienced by many hearing-impaired listeners.

In the present project we employ three processing stages, a first in which the peak-to-valley ratio of the spectrum is enlarged to counteract the broadening of the auditory filtering, a second in which the overall spectrum is modified to counteract the effects of USOM, and a third in which the noise in the stimuli is suppressed before the other two stages take place. The present scheme uses more subtle spectral modifications than those of Stone and Moore (1992), Baer *et al.* (1993), Baer and Moore (1997), and Franck *et al.*

(1999), with which we hope to reduce distortions. The scheme of Bunnell (1990) is fairly similar to the present one, but it lacks the present ripple filtering and spectral shaping. In two of the above cited enhancement papers combined enhancement and noise reduction were used: Cheng and O'Shaughnessy (1991) and Clarkson and Bahgat (1991). However, the schemes applied in those two papers were based on speech in white noise, while at present we wish to suppress speech-shaped noise. In the first experiment of the present study, only the two spectral enhancement stages are used (the ''expansion'' and the ''lift''), and in the second experiment all stages are used. The various schemes are evaluated with SRT measurements in which spectrally smeared stimuli are presented to normal-hearing subjects. A pilot study using three hearing-impaired listeners is performed to study the more successful versions of the enhancement scheme and to check the validity of the model of normal-hearing listeners and spectral smearing to emulate hearing impairment.

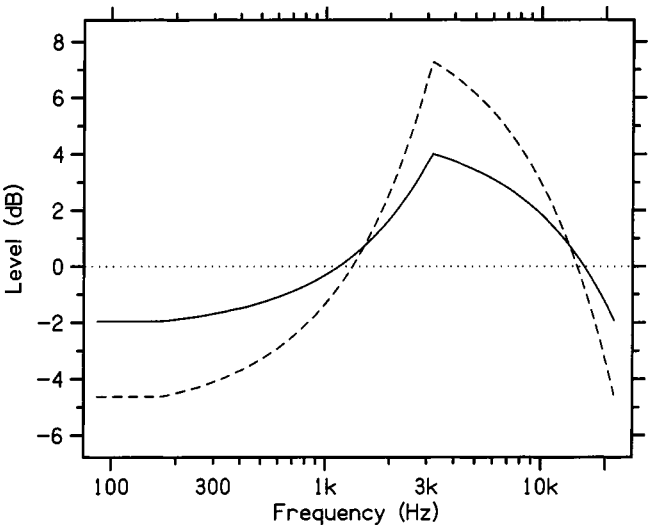


FIG. 1. The shape of the linear lift filter for the adaptation ranges of 6 (solid line) and 12 dB (dashed line) as a function of frequency.

II. ALGORITHMS

A. Speech enhancement

The speech processing scheme was incorporated in an overlap-and-add (OLA) procedure. The sample rate was 44100 Hz. Consecutive segments of 512 (duration: 11.6 ms) or 1024 samples (duration: 23.2 ms) were taken from the source signal, processed, and then added to create the output signal. Segments were taken with a 50% overlap. Unless mentioned otherwise, no zero padding was applied, and the segments were windowed using a raised-cosine (Tukey) window that contained a long midsection (96% of its length) at maximum amplitude. Next, a fast Fourier transform (FFT) was performed, after which the actual processing stages took place on the spectral frames. The processed signal segments were then created from the spectra with the inverse FFT (IFFT). To form a smooth continuous output signal, a raised-cosine (Hanning) window, in accordance with the 50% overlap, was used before adding the segments together. With the combination of the Tukey window before the FFT and IFFT, and the Hanning window after it, the sum of the overlapped windows does not exactly equal unity (Allen, 1977), but the deviation is less than 0.05%. See Appendix A 1 for more details about the side effects of the chosen windowing strategy.

The first part of the actual enhancement process was the spectral expansion stage. To enhance spectral features, the amplitude spectrum of each frame was raised to the power of the expansion factor (either 1.5 or 2.0). These calculations were performed by means of an adaptation function in dB that initially was set to contain the difference between the original and the expanded spectrum. To prevent the overall spectral shape from tilting in the procedure, as well as to prevent an increase in the statistical variations of the noisy content of each frame, the expansion was only applied on those spectral variations that might specify vowel formants; we only expanded spectral ripples with periods between about 350 and 1400 Hz. This was achieved by convolving the adaptation function with the impulse response of an appropriate band-pass filter. Expansion was only performed in the frequency region where natural formants normally occur: we used a region from 400 to 4500 Hz. Using a raised-cosine shaped window, the adaptation function was forced to zero outside this frequency range, after which the original spectrum was modified according to the adaptation function. Finally, the root-mean-square (rms) content of the frame was readjusted to its original value. We named this processing stage the expansion stage.

The second stage of the enhancement process implemented spectral adaptation using a fixed linear filter. Figure 1 shows the shape of this filter for two ranges of the spectral adaptation (6 and 12 dB). With the shape of the filter we emphasized the region of the third formant relative to that of the second and, in the same way, the region of the second formant relative to that of the first. With this emphasis we attempted to maintain the presence of second and third formants in excitations with large degrees of upward-spread-of-masking. We named this the lift stage.

In the second experiment we used an additional noise suppression stage. This stage consisted of a linear noise reduction filter applied before the enhancement stage. The amplitude shape of this filter was set high for spectral components with a regular behavior over time, and low for irregularly behaving, noisy, spectral components. The regularity of the components was determined over three or four consecutive frames (see Appendix A 2 for details). We named this the noise suppression stage.

With the intent to implement these algorithms in portable hearing aids in the future, the speed of the processing is a crucial factor. The size of the signal segments creates a fixed delay; therefore, segments should be as small as possible. On the other hand, good spectral resolution is needed to accurately enhance spectral features. A frame size of 512 samples seemed a good compromise between these two factors with an intrinsic delay of 11.6 ms and a spectral resolution of 86 Hz. For more details concerning the speech processing algorithms, see Appendix A.

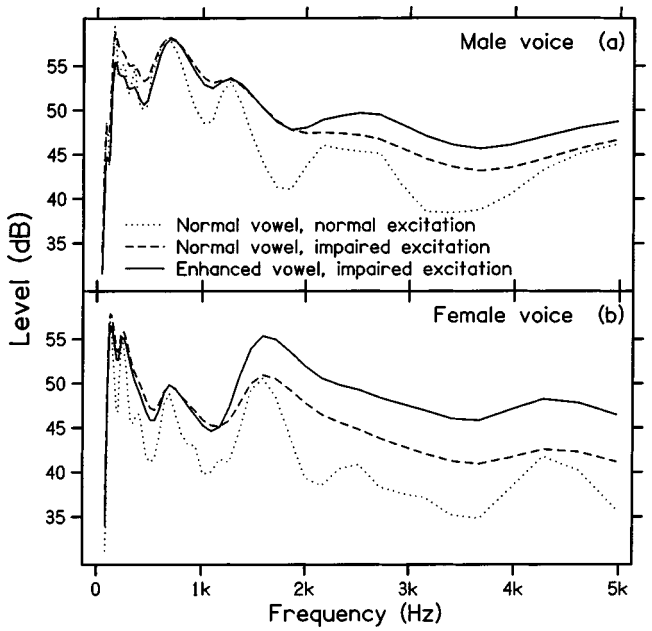


FIG. 2. Excitations for widened auditory filters (RoEx filter bank with a Q_{-10dB} of 2.0 and a skew factor of 2.0) of unprocessed (dashed lines) and processed (solid lines) vowels in quiet. For comparison normal excitations of the unprocessed vowels are shown in the dotted lines.

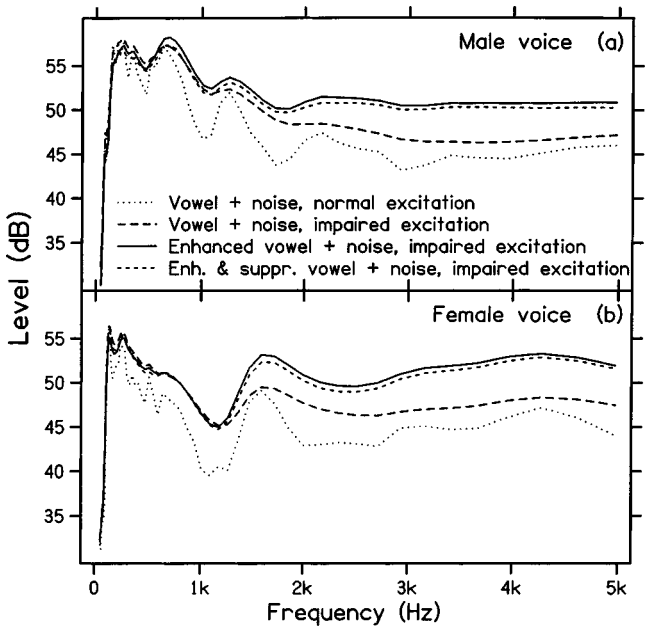


FIG. 3. The same as Fig. 2 for speech plus noise at a signal-to-noise ratio of 3 dB.

To examine the enhancement effects of the processing for various settings, we used a simple linear place model to calculate expected excitation patterns. The place model consisted of 70 overlapping rounded exponential (RoEx) filters (Patterson *et al.*, 1982) with a quality factor ($Q_{-10\,dB}$) of 2.0. These filters were roughly twice as wide as for the $Q_{-10\,dB}$ of 4 to 5 that would be expected for normal-hearing listeners (Patterson *et al.*, 1982; Patterson and Moore, 1986; Glasberg and Moore, 2000). As it is well known (Hoekstra and Ritsma, 1977; Patterson *et al.*, 1982; Glasberg and Moore, 1986) that for hearing-impaired listeners the auditory filters tend to tilt and become more like lowpass filters, we have introduced such a tilt in the RoEx filters. For this we introduced a skew factor that determines the relative steepness of the low- and high-frequency slopes, while keeping the $Q_{-10\,dB}$ fixed. The results in Figs. 2 and 3 have been calculated with a skew of 2.0 (high-frequency slope twice as steep as the low-frequency one). Figure 2 shows the expected excitations for the Dutch vowel /a/ as in ''vader'' for normal-hearing listeners ($Q_{-10\,dB}=5$, no skewing, dotted line), and hearing-impaired listeners ($Q_{-10\,dB}=2.0$, skew$=2.0$, dashed line). The solid line shows the expectations for the enhanced vowel using an expansion factor of 1.5 and a 6-dB lift-range (these settings are used in condition E_1L_1S of measurement 1). It can be seen that many spectral features are reduced with the widened auditory filters. The difference between the solid and the dashed lines shows that the presence of the higher-order formants is more pronounced in the processed version of the vowels, especially for the female speaker. Figure 3 shows excitations for the same vowels at a SNR of 3 dB. Adding the noise greatly reduces the amount and size of the spectral features. Nevertheless, the excitations for the enhanced vowels in the solid curves still show an increase in the amount of detail as compared to the curves for the original vowels in the long dashed curves. The excitations for enhancement and noise suppression (short dashed curves) show approximately the same amount of detail as those for just the enhancement.

B. Spectral smearing

The spectral smearing was performed in a separate computer program applied after the speech enhancement. Like the enhancement, the smearing was implemented in an OLA algorithm in which the actual processing took place in the spectral domain. From the frequency spectrum of each frame an expected excitation was calculated by convolving the amplitude spectrum with a simulated auditory filter (we used the RoEx filter). To prevent the occurrence of a rattling artifact produced in earlier versions of such algorithms (ter Keurs *et al.*, 1992; Baer and Moore, 1993), we generated random phase values for the regions with small excitation values (see Appendix B for further details). Excitation and phase spectrum were then retransformed using the IFFT and added to the resultant waveform using the Hanning window. The size of the segments was 512 samples. This is equal to the size used under most conditions in the enhancement scheme. However, the two programs were performed one after the other and no interactions from equal frame sizes have been observed. A 75% overlap of consecutive segments was used. Using 75% instead of 50%, we found that the smeared sounds contained more energy in the region of the fundamental[1], which made them sound more natural. On the other hand, the amount of effective smearing was slightly smaller (see below).

As explained in detail by Baer and Moore (1993), the smearing of such an algorithm is not as large as would be expected from the Q value of the simulated auditory filters. This is caused by the overlap of the segments. While each frame is well smeared, spectral features reappear in the re-

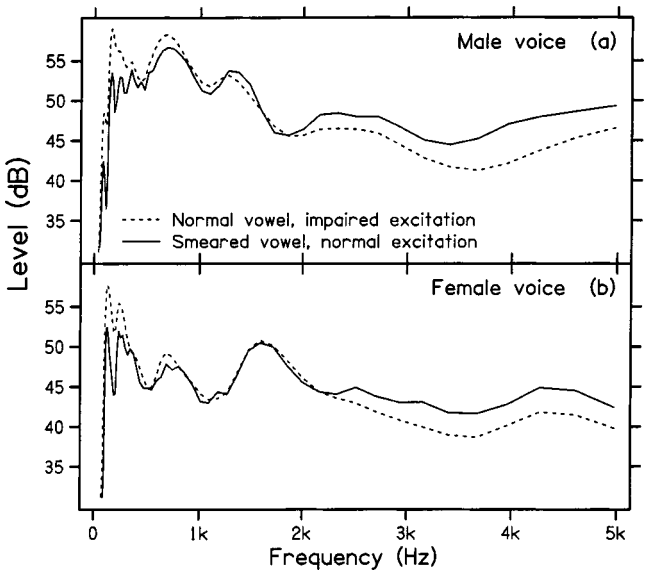


FIG. 4. Normal excitations of smeared vowels and impaired excitations of unprocessed vowels. The smeared excitations for the original vowels were calculated using a RoEx filter bank with a $Q_{-10\,dB}$ of 2.5 and a skew factor of 2.0. For the normal excitations of the smeared vowels a $Q_{-10\,dB}$ of 5.0 and a skew factor of 1.0 were used.

gions of overlap of the segments. In particular, the harmonics tend to reappear when different ''versions'' of the same stimulus piece are added together. With the partial random phase we applied in the present study, this reappearance of spectral features was much reduced since the correlation between consecutive frames was broken by the different random phase values in each frame. Nevertheless, the effective smearing of the algorithm was still somewhat smaller than expected for a $Q_{-10\,dB}$ of the RoEx filters of 2.5 (two times smaller than the $Q_{-10\,dB}$ of 5 that is often used for normal-hearing listeners). To achieve the desired amount of smearing we had to decrease the $Q_{-10\,dB}$ to 1.6. This is illustrated in Fig. 4, which shows normal excitations of two smeared vowels and excitations with reduced selectivity of the original vowels. The excitations of the smeared vowels (solid lines) show some overall tilt; they are relatively low below 400 Hz, and they show a slight rise for frequencies above 2 kHz. Over all, they show a little more spectral detail than the low-resolution excitations, but in the region of natural formants (400 to 4500 Hz) the spectral details are fairly similar.

III. METHODS AND STIMULI

A. Methods and conditions

We measured the speech reception threshold (SRT) in noise (Plomp and Mimpen, 1979). In this test lists of 13 everyday sentences were used in an adaptive procedure that estimates the SNR at which listeners are able to reproduce 50% of the sentences entirely correctly. We performed these measurements double blind; the experimenter, who judges the correctness of the sentences, did not hear the sentences and was not informed about the measurement condition. We used eight normal-hearing subjects with ages from 22 to 51 years, with an average of 35. The measurements contained eight conditions each which were organized in an 8×8 Latin square setup so that listeners, sentence lists, and measurement condition were balanced evenly. A female and a male speaker were used in each condition, resulting in 16 thresholds per subject. The different speakers and conditions were presented to the listeners in a randomized order.

The sentence lists used were taken from the material presented and tested by Versfeld *et al.* (2000). These lists were recorded using one male and one female speaker. The speech stimuli were presented at a fixed level of 61 dB SPL, the noise level varied during the measurements. The spectral shape of the noise was equal to the long-term average spectrum of the speech. A fresh sample of noise was taken for each sentence. Noise and speech were mixed before applying the processing scheme and the smearing (when present). The stimuli were calculated on a PC before the measurements and stored on hard disk. During the measurements the stimuli were retrieved from disk, transferred to analog using a TDT digital-to-analog converter DD1 followed by a TDT FT5 anti-aliasing filter with a cutoff of 16 kHz and a slope of 48 dB/oct. They were attenuated with a TDT PA4 attenuator and buffered with a TDT HB6 amplifier. The stimuli were presented to the right ear over Sony MDR-CD999 headphones.

Experiment 1 contained one unsmeared and seven smeared conditions, comprising two reference and six measurement conditions. The first reference condition was the normal SRT (Org), the second one was the SRT for smeared speech plus noise (S). For these conditions the expansion stage has been applied with an expansion factor of 1.0 (no expansion) and a flat lift filter. The six measurement conditions contained (apart from smearing): lift (L_1S), expansion (E_1S), expansion and lift (E_1L_1S), increased expansion (E_2S), increased expansion and lift (E_2L_1S), and increased expansion and increased lift (E_2L_2S). The expansion factors were 1.5 and 2.0 in the conditions E_1 and E_2, respectively. In the conditions L_1 and L_2, the ranges of the spectral lift were 6 and 12 dB, respectively.

In experiment 2 the noise suppression stage was added (see Appendix A 2 for a detailed description of this stage). To enhance the spectral resolution between FFT and IFFT from 86 Hz (for a frame size of 512 samples) to 43 Hz, we used conditions with a frame size of 1024 samples. With this resolution we tried to give the noise suppression filter the opportunity to suppress any noise between the harmonics of the voiced sentence parts. So far we had not used zero padding, because tests had shown that it was detrimental for the present nonlinear signal processing (see Appendix A 1). To put this assumption to the test, we included a condition with zero padding. Experiment 2 contained three unsmeared and five smeared conditions, comprising two reference and six measurement conditions. In the two reference conditions, the size of the segments was 1024 samples. The unsmeared conditions were a reference condition and two measurement conditions that contained only noise suppression (expansion factor: 1.0, lift range: 0 dB) with the averaging performed over four frames with a frame size of 512 (condition N_{4s}) or 1024 samples (condition N_{4l}). The five smeared conditions were a reference condition (S) and four measurement conditions that all contained noise suppression and enhancement

TABLE I. Evaluation of the enhancement schemes for simulated loss of frequency selectivity. The increases in excitation valleys between fundamental and first formant, and between first and second, and second and third formants, are given in columns ΔD1, ΔD2, and ΔD3, respectively. These values are averaged over ten vowels (male and female speaker, five vowels each) and three SNR values (3, 0, and −3 dB). The values for condition Org in the bottom row indicate these values for normal relative to reduced frequency selectivity. The total level of the voiced sentence parts relative to the unvoiced parts are given under V-UN for male and female speaker separately. The average speech distortion caused by the noise is given under Distortion. Under ΔSD the changes in standard deviation of the noise amplitudes, and under ΔTR the changes in the temporal ripples (in dB) of the noise, are given. Under ΔSN the changes in SNR are given for the male and female speaker separately.

Condition	Frame size	Expansion	Lift (dB)	ΔD1 (dB)	ΔD2 (dB)	ΔD3 (dB)	V-UN male (dB)	V-UN female (dB)	Distortion (dB)	ΔSD noise	ΔTR noise (dB)	ΔSN male (dB)	ΔSN female (dB)
S	512	1.0	0	0.0	0.0	0.0	14.2	5.3	−∞	0.0	0.0	0.0	0.0
L_1S	512	1.0	6	0.1	1.5	−0.1	9.6	1.3	−∞	0.0	−0.3	0.3	1.2
E_1S	512	1.5	0	0.1	2.3	0.3	14.2	5.3	−23	0.3	2.1	0.2	0.2
E_1L_1S	512	1.5	6	0.2	3.7	0.3	9.8	1.4	−22	0.3	1.8	0.7	1.5
E_2S	512	2.0	0	0.0	4.4	0.6	14.3	5.3	−17	0.6	5.7	0.4	0.3
E_2L_1S	512	2.0	6	0.1	5.3	0.7	10.1	1.7	−16	0.6	5.3	1.2	1.8
E_2L_2S	512	2.0	12	0.2	6.2	0.5	6.4	−0.8	−16	0.6	5.1	2.1	3.4
$N_{4s}ELS$	512	1.5	6	0.2	3.6	0.2	10.8	3.3	−16	0.4	3.3	1.1	1.6
$N_{4l}ELS$	1024	1.5	6	0.5	3.4	0.4	11.1	3.3	−15	0.4	2.3	1.1	1.4
$N_{3l}ELS$	1024	1.5	6	0.4	3.3	0.4	10.8	2.8	−15	0.4	2.9	1.2	1.5
$N_{4sz}ELS$	512	1.5	6	0.1	2.3	0.1	11.0	3.8	−13	0.7	5.9	1.2	1.3
Org	512	1.0	0	2.8	4.2	1.5	14.2	5.3	−∞	⋯	⋯	⋯	⋯

(expansion factor: 1.5, lift range: 6 dB). In the first of these measurement conditions, the averaging was performed over four frames of 512 samples ($N_{4s}ELS$), in the second the frame size was increased to 1024 samples ($N_{4l}ELS$). In the third condition the averaging took place over three frames, and the frame size was 1024 samples ($N_{3l}ELS$). The fourth condition ($N_{4sz}ELS$) had an input frame size of 512 samples and zero padding with a factor of 2, resulting in a frame size of 1024 samples in all spectral processing. In the noise reduction stage, the averaging was performed over four frames.

B. Acoustic analysis

Tables I and II present overviews of several factors regarding the measurement conditions. For each condition we applied the scheme to two Dutch sentences at various signal-to-noise ratios, one sentence spoken by a male speaker and one by a female speaker. From the results we estimated average changes in the depths of the valleys in vowel excitations, and we estimated the overall levels of the voiced relative to the unvoiced parts of the sentences. The valley depth between the fundamental and the first formant was calculated (using a $Q_{-10\,dB}$ of 2.5) for five vowels in each of the two sentences and for three values of the SNR: 3, 0, and −3 dB. For each measurement condition, the average change in this depth relative to that of the reference condition is given under "ΔD1." Under "ΔD2" and "ΔD3," the same are given for the valleys between first and second, and between second and third formants, respectively. By determining the voiced and the unvoiced parts of the sentences in quiet, their relative level differences could be estimated. These differences are given separately for the male and female voice under "V-UN male" and "V-UN female." The smaller these values, the louder the, usually relatively weak, consonants will be. To analyze the distortions produced by the enhancement scheme, we stored all the involved manipulations while processing sentence plus noise at 0 dB SNR. Subsequently, we applied those manipulations to sentence and noise separately. These signals were then compared with sentence and noise processed individually. From this comparison we determined distortion caused by the noise in the processing of the speech, expressed as the level of the difference between the two processed sentences relative to the overall level, given under "Distortion." The influence of the sentence on the processed noise was expressed as the standard deviation of the changes in noise amplitude (the larger this value, the more speechlike modulations have appeared in the noise), given under "ΔSD noise," and as the change in the temporal ripple of the noise. To enumerate the strength of this temporal ripple we calculated a series of short-term Fourier transformations along the length of the signal which were stored in a two-dimensional matrix (forming a spectrogram). Next, a two-dimensional FFT was performed on this matrix, and

TABLE II. Numerical evaluation of the noise suppression scheme. The format of this table is the same as that of Table I. The expected excitations of normal-hearing listeners were used to calculate the estimates.

Condition	Frame size	Expansion	Lift (dB)	ΔD1 (dB)	ΔD2 (dB)	ΔD3 (dB)	V-UN male (dB)	V-UN female (dB)	Distortion (dB)	ΔSD noise	ΔTR noise (dB)	ΔSN male (dB)	ΔSN female (dB)
Org	1024	1.0	0	0.0	0.0	0.0	14.2	5.3	−∞	0.0	0.0	0.0	0.0
N_{4s}	512	1.0	0	0.1	0.0	−0.1	15.3	7.4	−17	0.4	2.2	0.5	0.5
N_{4l}	1024	1.0	0	0.2	−0.1	0.0	15.6	7.5	−16	0.4	1.6	0.5	0.3

the total level in its first row along the time axis was taken as a measure of the temporal ripple power. The changes in these values are given in dB under "ΔTR noise." Finally, the changes in overall SNR produced by the enhancement were estimated by comparing the power levels of the processed and unprocessed sentences and noises. Since these SNR changes could show large variations between the two speakers, they are given for the male and the female speaker separately under "ΔSN male" and "ΔSN female."

In the fields along the top of Table I it can be seen that the depth between the first and the second formant (D2) increases for all conditions with expansion, but at the cost of higher distortions (Distortion), and less smooth noise (larger SD and TR). The changes in the depths below the first (D1) and above the second formants (D3) are very small. For the conditions with lift, the V-UN factors often improve, as do the signal-to-noise ratios. This leads to the expectation that a combination of lift and expansion might be more successful than each of them in isolation. A comparison of the E_1L_1S condition and the fields with the noise suppression shows that the noise suppression increases the SNR for the male voice. With the noise suppression, the depths below the first and above the second formants (D1 and D3) slightly increase in two cases, but at the cost of more distortion. The V-UN factors have increased slightly, but they are still better than found in the reference condition. In comparison with the reference condition, the noise suppression increases the average SNR with just over 1 dB. Without the smearing, the increase in SNR is reduced to about 0.5 dB in Table II. This increase is nearing the limit of the measurement accuracy. Below we will discuss these factors in relation to the measurement results.

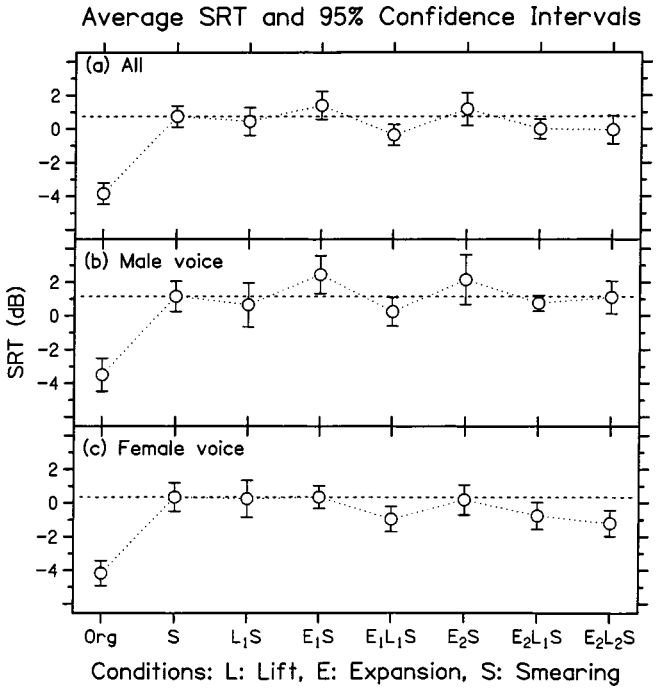


FIG. 5. Results of experiment 1. Plotted is the average SRT for eight measurement conditions. In panel (a) it was averaged over the eight subjects and two speakers. In panels (b) and (c) it was averaged over the eight subjects. These panels show data for the male and female speaker, respectively. The error bars indicate the 95% confidence intervals. The dashed lines indicate the level of the reference SRT for the smeared condition (S). The various conditions are described in the text.

IV. RESULTS

A. The enhancement scheme

The results of experiment 1 are shown in Fig. 5. Plotted are the SRT values for the eight measurement conditions, averaged over the eight subjects. The averages in panel (a) were calculated over both speakers. Panels (b) and (c) show the averages separated for the male and the female speaker, respectively. The error bars indicate the 95% confidence intervals. The dashed lines indicate the level of the reference SRT for the smeared condition (S). For normal-hearing listeners, the average SRT is −3.8 dB. This is in good agreement with the value of −4.1 dB found for this speech material by Versfeld *et al.* (2000). The difference in average SRT for the smeared and the unsmeared reference conditions indicates that the smearing produced a 4.6-dB increase in SRT.

The effects of the various processing schemes are shown in the remaining six conditions. Panel (a) shows that only the conditions with combined spectral expansion and lift produce an improvement in the average SRT. For the conditions E_1L_1S and E_2L_1S the reductions in SRT are 1.1 and 0.8 dB, respectively, and they are found to be significant [paired sample *t*-test $p<0.002$ (E_1L_1S) and $p<0.05$ (E_2L_1S), after Bonferoni correction]. The reduction in SRT for condition E_1L_1S was obtained separately for both the male and the female speaker; see panels (b) and (c). For the conditions E_2L_1S and E_2L_2S, the SRT values show a clear decrease for the female speaker but not for the male speaker. This indicates that, for these more extreme parameter settings, the present enhancement scheme functions better for the female than for the male voice.

B. Addition of the noise suppression scheme

The results of experiment 2, incorporating noise suppression before the enhancement stage, are shown in Fig. 6. The format of this figure is the same as that of Fig. 5. The dashed lines on the left side of each panel indicate the level of the SRT for the unsmeared reference condition, those on the right indicate the smeared reference levels. The results on the left side of panel (a) show that the application of just the noise reduction stage does not improve the SRT of normal-hearing listeners. The results with the smearing, on the right side of panel (a), show that the combination of noise reduction and enhancement only seems to produce a decrease (0.8 dB) in SRT for the $N_{4s}ELS$ condition, which incorporated averaging over four frames and a frame size of 512 samples. However, this decrease in SRT failed to reach significance (paired samples *t*-test, $p=0.11$ after Bonferoni correction). The results with the frame size of 1024 samples and those with zero padding are all negative. For condition $N_{4sz}ELS$, with zero padding, the average SRT is significantly higher (by 1.5 dB) than for than the corresponding $N_{4s}ELS$ condition without zero padding (paired samples *t*-test, $p<0.05$

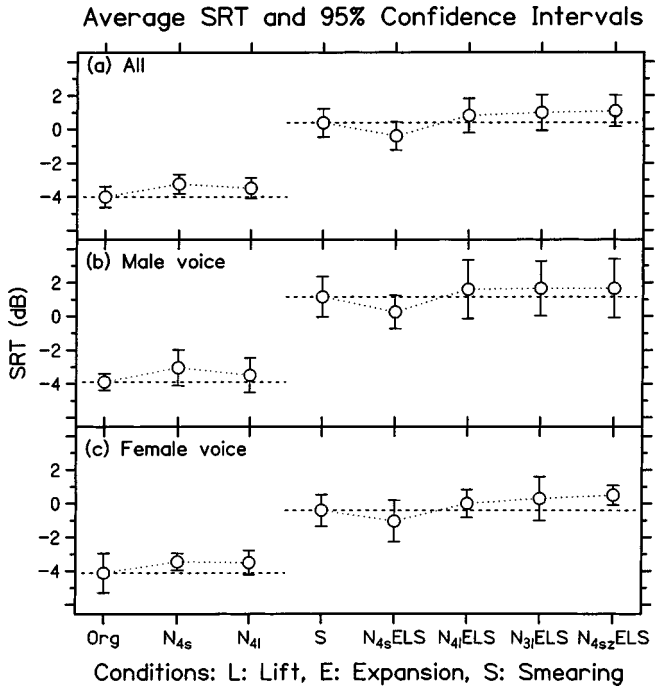


FIG. 6. Results of experiment 2. The format of this figure is the same as that of Fig. 5. The dashed lines on the left and right sides of the panels indicate reference SRT values for the unsmeared and the smeared conditions, respectively.

after Bonferoni correction). This implies that the application of zero padding degrades rather than improves the quality of the processing.

C. Pilot study with hearing-impaired listeners

For the hearing-impaired subjects, an audiogram was recorded for each ear, after which the best ear was chosen for the measurements. The audiogram of subject HI1 showed a loss of around 10 dB up to 1 kHz, which gradually increased to 60 dB at 4 kHz, and stabilized from there. HI2 had a 30-dB loss up to 2 kHz, increasing to 70 dB at and above 4 kHz. HI3 had a 25-dB loss up to 500 Hz, increasing to 40 dB at and above 1 kHz, with a narrow peak of 60 dB at 4 kHz. To present as much speech information as possible in the dynamic range of the subjects, all stimuli were spectrally adapted according to a half-gain rule. This spectral adaptation function was shaped according to the half-way point between the absolute threshold and a 110-dB SPL uncomfortable loudness level. The stimuli were presented at such a level that a calculation of the speech intelligibility index (SII) (ANSI, 1997) predicted optimal performance (i.e., a maximal SII). This approach optimizes presentation levels between loud enough to present most of the important speech information above threshold, and soft enough to prevent large amounts of self-masking from occurring in the speech. Without noise, the SII values for the reference condition (averaged over the male and female speakers) were 0.77, 0.59, and 0.54 for subjects HI1, HI2, and HI3, respectively. These values represent the fraction of the total speech information, corrected for spectral masking, available to the subjects. At the SRT these SII values dropped to 0.38, 0.35, and 0.36, respectively, indicating a clear masking effect of the noise. Comparable SII values were found for the spectrally adapted SRT (SRTa) by Noordhoek (2000) for hearing-impaired listeners with supra-threshold impairments.

To estimate the steepness of auditory filters, we used detection measurements of 1-kHz pure tones in notched noises with several notch widths. This method estimates the shallowest auditory filter slope that passes through the 1-kHz region. For normal-hearing listeners, we generally find values around 90 dB/oct when measuring at 1 kHz. For subject HI1 we found a slope of 80 dB/oct, indicative of near normal auditory filters in an area around 1 kHz. For subjects HI2 and HI3 we found slopes of 32 and 12 dB/oct, respectively, indicating that these subjects have strongly broadened auditory filters near 1 kHz. So, HI2 and HI3 form part of the target group of the present speech enhancement scheme, while HI1 does not seem to suffer from reduced spectral resolution.

Table III shows the results of the SRT measurements for the three hearing-impaired subjects. The top row contains the reference SRT values of the sentence lists of the male and female speakers. Averaged over speakers, these reference SRT values are −1.6, −1.0, and −1.8 dB for subjects HI1, HI2, and HI3, respectively. This indicates that all subjects have an elevated SRT in noise. The lower three rows present the relative SRT values for three enhancement conditions. These three conditions had been used before, albeit in combination with the smearing, in either experiment 1 or 2. They are indicated with the same labels as in those experiments (Figs. 5 and 6), but now without the ''S'' for smearing. The SRT differences show that there was little or no gain for subject HI1. For subjects HI2 and HI3, the enhancement schemes often improved SRT values for the female speaker, while they were less successful for the male voice. The most

TABLE III. The results of the pilot study using hearing-impaired listeners. The top row marked ''Reference'' shows the SRT values for the reference conditions, the lower three rows show the SRT differences between the reference and the indicated conditions. For subject HI1, the width of the auditory filter near 1 kHz was found to be normal. For subjects HI2 and HI3, the auditory filters had broadened considerably. The last two columns show their average results.

Condition	Frame size	Expansion	Lift (dB)	HI1 male	HI1 female	HI2 male	HI2 female	HI3 male	HI3 female	Average 2 & 3 male	Average 2 & 3 female
Reference	512	1.0	0	−1.8	−1.4	−0.6	−1.4	−1.8	−1.8	−1.2	−1.6
ΔE_1L_1	512	1.5	6	0.0	0.0	0.0	−2.0	−0.8	−1.6	−0.4	−1.8
$\Delta N_{4s}EL$	512	1.5	6	2.4	−0.4	0.8	0.0	2.0	−0.8	1.4	−0.4
ΔE_2L_2	512	2.0	12	1.6	0.4	−1.2	−1.2	2.8	−0.4	0.8	−0.8

consistent improvements are found for enhancement scheme E_1L_1.

V. DISCUSSION

The results of Experiment 1 in Fig. 5 show a 4.6-dB increase in SRT from the unsmeared to the smeared conditions. For hearing-impaired subjects increases in SRT between 1 and 10 dB have been recorded (Plomp, 1986; Noordhoek *et al.*, 2000). More specifically, for eight subjects with decreased spectral and reasonably good temporal resolution, as measured in upward and downward spread of masking as compared to forward masking, Noordhoek (2000) found an average increase of 5.5 dB relative to 12 normal-hearing subjects. Therefore, the present increase is in good agreement with the expectations which means that, in terms of produced increase in SRT, the effect of the smearing scheme is comparable to that of a loss of frequency selectivity.

The 4.6-dB increase in SRT obtained with the present smearing scheme is much closer than previous results to the value of 5 dB that is expected for a ratio of 2 to 3 loss in spectral resolution. Ter Keurs *et al.* (1993) only smeared the spectral envelope while retaining the original spectral fine structure. They found a 3-dB increase in SRT when convolving with a filter of 1 oct, which is roughly equal to the presently used filter width. Interpolation of their data indicates that they would need filter width of about 1.5 oct for a 5-dB increase in SRT. Baer and Moore (1993) smeared the whole spectrum and, at a -3 dB SNR, they lost about 30% in word scores after convolving with a filter that was three times wider than the normal ERB. A similar loss was found at SNR of 0 dB and a filter width of six times the ERB. Their results indicate a detrimental effect of smearing on speech intelligibility that was almost as large as reported by ter Keurs *et al.* That the effects of the present smearing scheme are larger can probably be attributed to the partly randomized phase spectra that we used.

As found previously (Baer *et al.*, 1993; Franck *et al.*, 1999), the results of the first experiment indicate that the application of just spectral expansion to counteract loss of frequency selectivity does not lead to improved SRT values (though it may still improve the subjective listening quality for speech in noise). The same is true for the lift filter with which we tried to counteract the upward-spread-of-masking. An improved SRT could, however, be achieved when combining these two strategies (condition E_1L_1S). This may be because spectral expansion improves the vowel quality while it decreases the quality of the consonants (Franck *et al.*, 1999), where the spectral lift increases the overall energy of the unvoiced parts of speech, and thus may prevent consonants from losing too much of their quality in the noise.

These notions are corroborated by the factors in the top section of Table I. This table shows that the increase in valley depth between formants is larger for combined application of expansion and lift than for these factors separately (the largest increases are found between the first and the second formant). The relative levels of the unvoiced parts of the sentences, that primarily contain consonants, are larger for the conditions with lift (i.e., smaller values for the V-UN factors). The most successful condition of experiment 1, E_1L_1S, has reasonably good increases in valley depth and consonant levels, while not showing large distortions in the speech (under Distortion) or decreases in the temporal regularity of the noise (expressed as increases in the ΔSD and ΔTR factors).

The introduction of the noise suppression stage in experiment 2 has no positive effects. This may well be due to the increased levels of distortion that the application of this stage introduces. The middle section of Table I shows that, in relation to condition E_1L_1S, adding the noise suppression stage slightly improves the valley depth between fundamental and first formant, but at the cost of larger speech distortion and decreased temporal regularity of the noise. The speech distortion is smallest for the more successful condition of experiment 2 (N_{4s}ELS). For this condition the SRT values are smaller than those found for the more formally correct processing scheme that includes zero padding (N_{4sz}ELS). This is probably due to the smaller valley depths of condition N_{4sz}ELS in combination with its decreased speech quality and noise regularity.

For the unsmeared conditions, the noise suppression does not improve the SRT. This is corroborated by the results in Table II, which shows that the noise suppression stage does not produce large increases in SNR while it does introduce distortions in both speech and noise. When comparing the SRT changes in Fig. 6 between conditions with smearing and those without, a small improvement is seen for the N_{4s}ELS relative to the S condition (treated in Table I) that is not found without smearing and processing for the N_{4s} relative to the Org condition (treated in Table II). Given a positive effect of about 1 dB for the enhancement, as found in experiment 1, this would imply that the effects of the noise suppression are less detrimental when it is followed by smearing. This suggests that the noise suppression may stand a better chance for listeners with reduced spectral resolution than for normal-hearing listeners. This is in agreement with the notion that more noise enters an internal frequency channel when its associated auditory filter is wider.

The effectiveness of the different components of the algorithm seems to differ for the male and the female speaker. Several differences between the male and the female voice may be involved, such as differences in fundamental frequency, in the spread of the formants, in the overall tilt of the vowel spectrum, and in the energy in the voiced versus the unvoiced parts of the speech. The simulations in Figs. 2 and 3 indicate an effect of spacing and level of formants: because the second formant is larger in the female version of the vowel it receives more emphasis. The overall downward tilt of the spectrum is larger for the sentence of male speaker. This may make it more susceptible to upward-spread-of-masking, so more lift will be needed. Table I shows that the consonants carry more of the total energy for the female than for the male speaker, which could render the female voice more robust to spectral expansion. A further difference could lie in a different sensitivity of the male and the female voice to the distortions produced by the enhancement. When large, such distortions quickly overshadow any positive effects the enhancement scheme may have.

The speech enhancement scheme only produces subtle changes in cues in limited parts of the sentences. Any positive effect of the processing scheme on the SRT will depend on listeners being able to use these very subtle increases in number and size of formant peaks. The SRT is not a very sensitive instrument for recording such subtle differences. Word scores are probably more sensitive, and they also give the opportunity to examine the effects of enhancement schemes for different phonemes. From the SRT, a division between vowel and consonant scores cannot be made. A further limitation of the SRT is that it does not need to relate directly to subjective listening quality; two counteractive forces in this are, first, the notion that more ease of understanding will increase the subjective listening quality, while, second, a large amount of high-frequency energy often improves intelligibility scores in hearing-impaired listeners while this is experienced as highly unpleasant. Nevertheless, the SRT is a really tough test that provides a measure that is more representative of speech understanding in real life situations than word scores do. When an improvement in SRT is recorded, this is a good indication that a positive effect will also be found in real-life listening situations.

For the hearing-impaired subjects, the combined strategy E_1L_1 produced the most consistent SRT improvements. For the subjects with reduced frequency selectivity (HI2 and HI3), the effects are of the same size as found for the normal-hearing listeners with smearing. However, for the measurements with the hearing-impaired subjects, the nominal standard deviations of the results are of the same size as the recorded improvements. Therefore, we cannot draw any positive conclusions from this pilot experiment, but the results for the female speaker do lead to careful optimism.

The average SRT differences for the hearing-impaired subjects are -1.1, 0.5, and 0.0 for the conditions E_1L_1, $N_{4s}EL$, and E_2L_2, respectively. For the normal-hearing listeners we found a significant difference of -1.1 dB for the first condition and no significant differences for the latter two conditions. So, there is some similarity in the behavior of the normal-hearing listeners with spectral smearing on the one side and of the hearing-impaired listeners on the other side. However, from the limited results of the pilot study we cannot positively conclude that the present spectral smearing produces an accurate model of loss of spectral resolution as in hearing-impaired listeners, but the model is clearly not very far from the mark.

VI. CONCLUSIONS

The present spectral smearing scheme, incorporating a partial random phase, produced an increase in SRT for normal-hearing listeners that was very close to the expected increase for the hearing impaired suffering from reduced frequency selectivity.

Spectral expansion combined with spectral filtering (lift) can improve the intelligibility of speech in noise, as measured in the SRT while using simulated loss of frequency selectivity. However, care has to be taken in the development of such algorithms, as the positive effects of such a scheme can quickly be overshadowed by the detrimental effects of the distortions it inherently produces.

The addition of the present noise suppression stage did not give any further improvements.

ACKNOWLEDGMENTS

This research was supported by the Technology Foundation STW, Applied Science Division of NWO, and the technology program of the Ministry of Economic Affairs of the Netherlands. The authors wish to thank the reviewers for their constructive comments.

APPENDIX A: SPEECH PROCESSING ALGORITHM

1. Enhancement: Windowing and zero padding

To maximize processing speed, it was our aim to use as little zero padding as possible in the OLA algorithm. To examine the effects of different windowing strategies, we have plotted the waveforms of processed 23.2-ms segments of a natural utterance of the vowel /e/ as in the Dutch word ''spreek.'' In the examples shown, we have used four times zero padding and both spectral compression and spectral expansion with a ratio of 2.0. The spectra of the original and processed segments are shown in Fig. 7. Figure 8 shows the spread of the waveforms out of the original 23.2-ms segment for various conditions (to be able to plot these relatively small signals, the waveforms in the segments themselves have been forced to zero). The spread is only shown behind the segment; a similar spread occurs before it. The total size of these spreading waveforms is indicative of the amount of problems that can be expected without zero padding, as these portions of waveforms will then be reflected back into the frame as artifacts. Panels (b) and (c) show the spread while using a raised-cosine (Tukey) window with 96% maximal amplitude (solid lines) or a Hanning window (dashed lines) before the compression and the expansion, respectively. Here the Hanning window functions better. Panels (e) and (f) show these same waveforms with a second window [shown

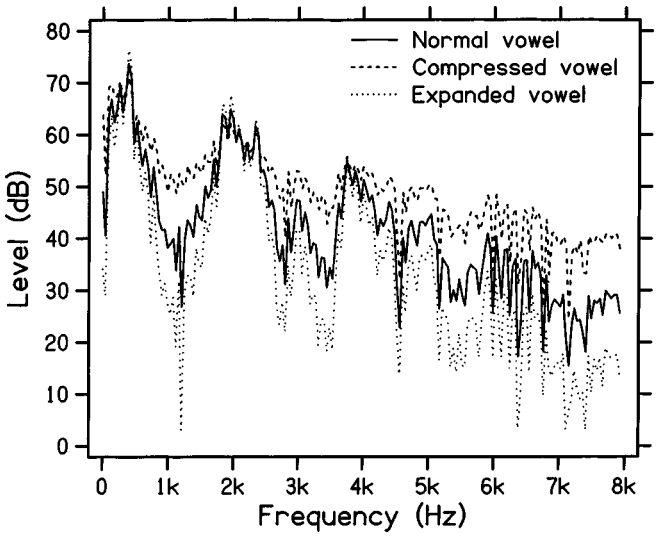


FIG. 7. Spectra of original (solid line), compressed (dashed line), and expanded (dotted line) versions of a 23.2-ms segment of a natural utterance of the vowel /e/ as in the Dutch word ''spreek.'' Spectral compression and expansion were performed with a ratio of 2.0. The spectra were calculated using a Hanning window.

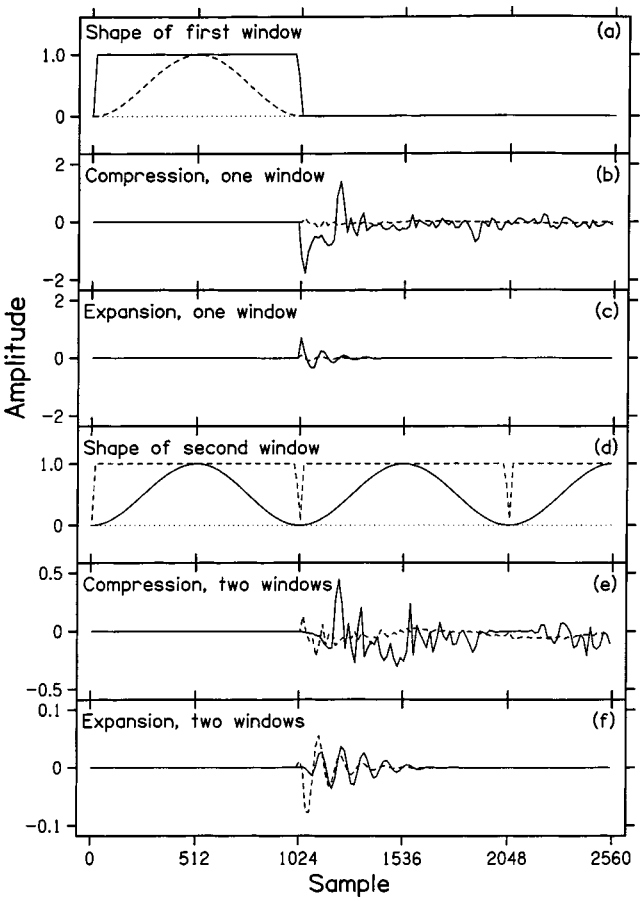


FIG. 8. Spread of the waveform behind the original segment size with four times zero padding. For clarity, the waveforms in the original segments have been forced to zero. On the present scale, the amplitudes in the segments ranged from 6 to −10. The calculations have been performed using the same vowel segments as used in Fig. 7. For the waveforms in panels (b) and (c) windowing was only applied before the FFT using the Hanning (dashed line) or Tukey window (solid line) in panel (a). For the waveforms in panels (e) and (f) a second window, with the alternative shape (Tukey versus Hanning) as the first one, was applied after the IFFT. These windows are shown in panel (d). To illustrate the size of the problems that can arise without zero padding, the second window was mirrored at integer multiples of the segment borders. The line types of the windows and waveforms correspond with each other over the panels.

in panel (d)] applied after the spectral processing. The second window had the interchanged shape of the first window; the solid lines show the spread using a Tukey window before, and a Hanning window after IFFT, and the dashed lines show the spread for reversed order of windows. To study the distortions that occur without zero padding, the second windows were not zero outside the frame, but their shapes were mirrored in the segment borders as shown in panel (d). Without the application of zero padding, the waveforms in panels (e) and (f) should be minimized. Panel (f) shows that, unlike for the spectral compression, for the expansion the better strategy is to use a Tukey before, and a Hanning window after, the spectral processing [note the different scales for panels (e) and (f)].

By listening to separate frames that had been processed with zero padding and without a second window, we were able to determine the nature of the parts of the waveforms that spread out of the original segment [as in Fig. 8(b)]. For the expansion schemes that we used, they sound like metallic echoes and do not carry any speech information. Though they will contribute to the desired spectral alterations, they may introduce new and unwanted sounds in neighboring segments when no second window is used. Excluding this spread by the application of a second window will then reduce distortion. Reducing the operations to the size of the original frame will mean that the delay of the design can be reduced. For these reasons, it seemed a good idea to first try a version of the OLA algorithm without zero padding. In such a strategy the effectiveness of the spectral manipulations will be reduced, but that can easily be compensated for by applying a little stronger expansion and lift.

In an informal binaural listening test using sentences in noise at a SNR of 0 dB, we have compared the results of the present processing scheme, with no zero padding, a Tukey window before and a Hanning window after the IFFT, with those from a more formally correct one with a Hanning window and four times zero padding before the FFT, and without a second window (Allen, 1977; Allen and Rabiner, 1977; Griffin and Lim, 1984). The two types of stimuli were presented simultaneously and dichotically. By alternating the ear of presentation for the two stimuli we could examine some of the differences between them. When switching ears, the speech parts of the stimuli remained located in the center of the head, while the noisy parts were perceived louder on the side of the stimulus created with the zero padding. It seems, therefore, that for the present nonlinear expansion the more formally correct scheme generates more distortion from interactions between speech and noise. Of course, the same need not be true for linear processing.

2. Noise reduction: Averaging in two ways

The shape of the noise reduction filter was determined from the ratio of the average complex amplitude, and the square root of the average intensity (i.e., the rms value), calculated over the last three or four frames. When the contents of the frames are very noisy, the average complex amplitude spectrum will be smaller than the rms spectrum, as the random phase values of the complex amplitudes will lead to a smaller than maximal sum. When averaging over N noise spectra, this will lead to a nominal reduction in amplitude by a ratio $\sqrt{N}$. For sinusoidal components we have to make sure that the average amplitude is as large as possible, so they will lead to clearly larger average amplitudes than the noise. To achieve this we need to realign their phase values over consecutive frame spectra, so that the sum of the complex amplitudes will not suffer too much reduction from phase differences. For this we corrected each phase value with the phase shift caused by the time delay between the frames. For sinusoidal components with frequencies that are equal to integer multiples of the frequency resolution of the FFT (i.e., exactly equal to the frequency of a FFT bin) this is sufficient to equalize their phase values over time. For most sinusoidal components that condition will not be met, of course. Therefore, we next calculated the frame-to-frame phase shifts (correcting with 2π when necessary) after which the phases in each frame were corrected according to the average values of these shifts. In this way we were able to reliably track the

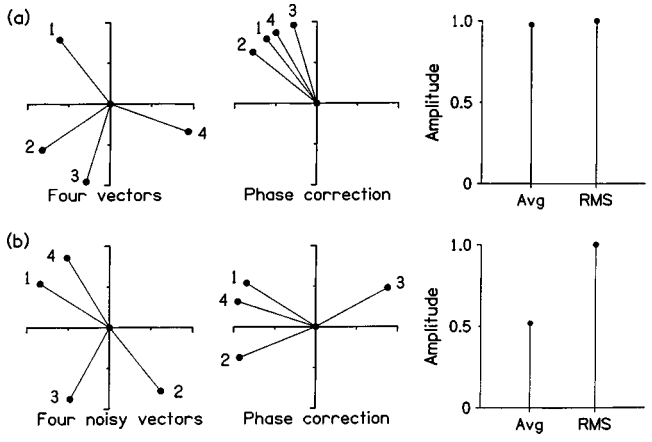


FIG. 9. Phase correction and averages for four vectors representing the behavior over time of a FFT bin with a sinusoid plus a small amount of noise in panel (a) and of a FFT bin with a very noisy content in panel (b). To illustrate the phase effects, all amplitudes are chosen equal to unity. The panels on the left show the original vectors, the mid panels show them after phase correction. The panels on the right show the differences between the absolute value of the average and the rms average.

phase of sinusoids whose frequencies did not fall exactly on a FFT bin. This is illustrated in Fig. 9, which shows two examples of the behavior of a FFT bin, and its complex and rms averages, over four consecutive frames. To form the noise reduction filter, the absolute values of the average amplitude spectrum were then divided by the values of the rms spectrum. The spectral amplitudes of the frame under attention were then multiplied with this noise filter, thus attenuating the amplitudes of FFT bins with noisy content more than those with mainly sinusoidal content.

To estimate the effects of the noise reduction scheme, we have processed pure-tone and complex stimuli in noise. For a slowly changing sinusoid (rising from 20 to 4400 Hz in 1.75 s) in white noise, a 2-dB SNR improvement was achieved. As examples of relevant complex stimuli, we processed two vowels, /e/ and /a/ as in the Dutch words ''spreek'' and ''vader,'' spoken both by a male and by a female speaker. These vowels were taken from sentences in speech noise at 0 dB SNR. The increases in SNR for the noise reduction scheme, with a frame size of 512 and averaging over 4 frames, were 1.4 and -0.1 dB for the two vowels of the male speaker, and 1.9 and 1.2 dB for the two vowels spoken by the female. So, for these four examples we achieved an average increase of 1.1 dB in SNR.

APPENDIX B: SMEARING ALGORITHM: THE PHASE SPECTRUM

To create the expected excitation as used in the smearing algorithm, we convolved the amplitude spectrum with a RoEx filter with a constant Q and a skew factor of 2. Next, the formant peaks in the excitation were located, and the peaks that produced them were traced in the original amplitude spectrum (due to the skewness of the filter shape, the excitation peaks usually occurred at slightly higher frequencies than the original peaks in the amplitude spectrum). Using this information, a phase spectrum for the excitation was

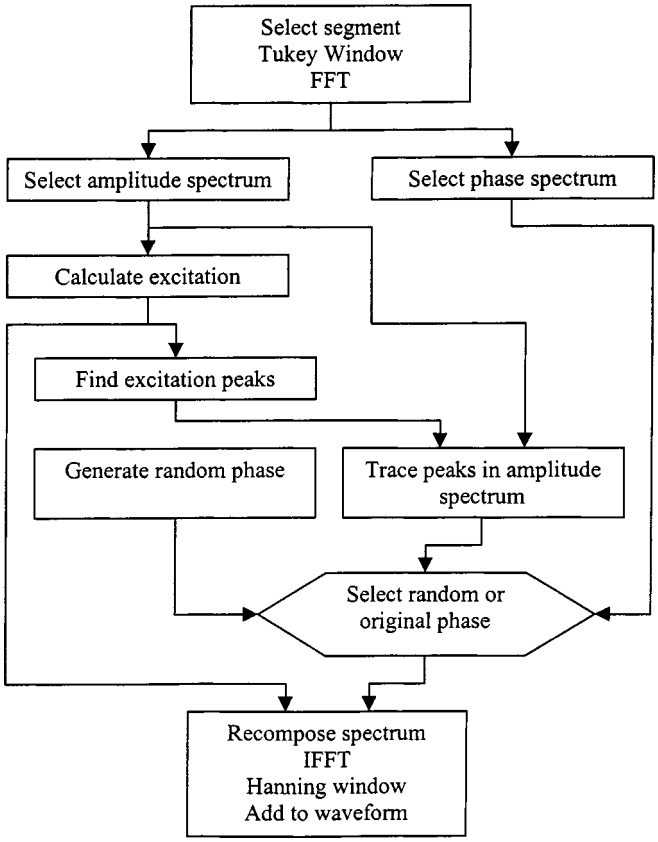


FIG. 10. Schematic of the spectral smearing algorithm.

constructed, so that it could be retransformed to a waveform. We copied the phase spectrum from a region (7 FFT bins wide for a 512 bin frame) around the peaks in the original amplitude spectrum to the areas of the corresponding peak in the excitation. All the remaining values of the phase spectrum were given a random value. In this way we were able to maintain the original group delay of this section of the sound to some extent, and suppress possible artifacts from mismatched group delays. A schematic of this algorithm is given in Fig. 10.

The use of partly randomized phases in the smearing scheme prevents the occurrence of a rattling artifact. Instead it produces noise where the random phase values are introduced in the spectrum. However, these random phase values are all introduced in sections in which the spectral amplitudes are fairly small. Therefore, the noise level produced by the algorithm is rather low. It is, however, clearly audible when processing speech in quiet. So, the results from this smearing scheme do not represent the percepts of hearing-impaired listeners for speech in quiet. However, the present algorithm was designed for use in SRT measurements, which incorporate noise in each stimulus presentation. When noise is added to the speech before the smearing, the noise generated by the algorithm becomes part of, or disappears in, that background noise. For that reason, we consider the application of this smearing scheme a reasonably good approximation of the loss of frequency selectivity in the present SRT measurements.

[1]The fundamental was found to be attenuated by the smearing when its period was of the order of the length of the segments. This attenuation could be partly counteracted by increasing the overlap of the segments.

ANSI (**1997**). ANSI S3.5-1997, ''American National Standard Methods for Calculation of the Speech Intelligibility Index'' (American National Standards Institute, New York).

Allen, J. B. (**1977**). ''Short term spectral analysis, synthesis and modification by discrete Fourier transform,'' IEEE Trans. Acoust., Speech, Signal Process. **25**, 235–238.

Allen, J. B., and Rabiner, L. R. (**1977**). ''A unified approach to short-time Fourier analysis and synthesis,'' IEEE **65**(11), 1558–1564.

Baer, T., and Moore, B. C. J. (**1993**). ''Effects of spectral smearing on the intelligibility of sentences in noise,'' J. Acoust. Soc. Am. **94**, 1229–1241.

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., and Moore, B. C. J. (**1997**). ''Evaluation of a scheme to compensate for reduced frequency selectivity in hearing-impaired subjects,'' in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Hillsdale, NJ), pp. 329–341.

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.

Boll, S. F. (**1979**). ''Suppression of acoustic noise in speech using spectral subtraction,'' IEEE Trans. Acoust., Speech, Signal Process. **27**, 113–120.

Bunnell, H. T. (**1990**). ''On enhancement of spectral contrast in speech for hearing-impaired listeners,'' J. Acoust. Soc. Am. **88**, 2546–2556.

Cheng, Y. M., and O'Shaughnessy, D. (**1991**). ''Speech enhancement based conceptually on auditory evidence,'' IEEE Trans. Signal Process. **39**, 1943–1954.

Clarkson, P. M., and Bahgat, S. F. (**1991**). ''Envelope expansion methods for speech enhancement,'' J. Acoust. Soc. Am. **89**, 1378–1382.

Ephraim, Y., and Malah, D. (**1984**). ''Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,'' IEEE Trans. Acoust., Speech, Signal Process. **32**, 1109–1121.

Franck, B. A. M., van Kreveld-Bos, C. S. G. M., and Dreschler, W. A. (**1999**). ''Evaluation of spectral enhancement in hearing aids, combined with phonemic compression,'' J. Acoust. Soc. Am. **106**, 1452–1464.

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., and Moore, B. C. J. (**2000**). ''Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise,'' J. Acoust. Soc. Am. **108**, 2318–2328.

Griffin, D. W., and Lim, J. S. (**1984**). ''Signal estimation from modified short-time Fourier transform,'' IEEE Trans. Acoust., Speech, Signal Process. **32**, 236–243.

Hoekstra A., and Ritsma, R. J. (**1977**). ''Perceptive hearing loss and frequency selectivity,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 263–271.

Huizing, H. C. (**1948**). ''The symptom of recruitment and speech intelligibility,'' Acta Oto-Laryngol. **36**, 346–355.

Leek, M. R., and Dorman, M. F. (**1987**). ''Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 148–154.

Levitt, H. (**2001**). ''Noise reduction in hearing aids: an overview,'' J. Hear. Res. Dev. **38**, 111–121.

McAulay, R. J., and Malpass, M. L. (**1980**). ''Speech enhancement using a soft-decision noise suppression filter,'' IEEE Trans. Acoust., Speech, Signal Process. **28**, 137–145.

Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999**). ''Contrast enhancement improves the representation of /ɛ/-like vowels in the hearing-impaired auditory nerve,'' J. Acoust. Soc. Am. **106**, 2693–2708.

Noordhoek, I. M. (**2000**). ''Intelligibility of narrow-band speech and its relation to auditory functions in hearing-impaired listeners,'' Ph.D dissertation, Free University, Amsterdam, The Netherlands.

Noordhoek, I. M., Festen, J. M., and Houtgast, T. (**2000**). ''Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. II Hearing-impaired listeners,'' J. Acoust. Soc. Am. **107**, 1685–1696.

Patterson, R. D., and Moore, B. C. J. (**1986**). ''Auditory filters and excitation patterns as representations of frequency resolution,'' in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 123–177.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Millroy, 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.

Plomp, R. (**1978**). ''Auditory handicap of hearing impairment and the limited benefit of hearing aids,'' J. Acoust. Soc. Am. **63**, 533–549.

Plomp, R. (**1986**). ''A signal-to-noise ratio model for the speech-perception threshold of the hearing impaired,'' J. Speech Hear. Res. **29**, 146–154.

Plomp, R., and Mimpen, A. M. (**1979**). ''Improving the reliability of testing the speech reception threshold for sentences,'' Audiology **18**, 43–52.

Stone, M. A., and Moore, B. C. J. (**1992**). ''Spectral feature enhancement for people with sensorineural hearing impairment: Effects on speech intelligibility and quality,'' J. Rehabil. Res. Dev. **29**, 39–56.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). ''Effect of spectral envelope smearing on speech reception I.,'' J. Acoust. Soc. Am. **91**, 2872–2880.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1993**). ''Effect of spectral envelope smearing on speech reception II,'' J. Acoust. Soc. Am. **93**, 1547–1552.

van Schijndel, N. H., Houtgast, T., and Festen, J. M. (**2001**). ''Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **110**, 529–542.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). ''Method for the selection of sentence materials for efficient measurement of the speech reception threshold,'' J. Acoust. Soc. Am. **107**, 1671–1684.

Villchur, E. (**1973**). ''Signal processing to improve speech intelligibility in perceptive deafness,'' J. Acoust. Soc. Am. **53**, 1646–1657.

A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn

Helen M. Hanson[a),b)] and Kenneth N. Stevens[b),c)]
Sensimetrics Corporation, 48 Grove Street, Suite 305, Somerville, Massachusetts 02144-2500

(Received 19 September 2001; accepted 31 May 2002)

The HLsyn speech synthesizer uses models of the vocal tract to map higher-level quasiarticulatory parameters to the acoustic parameters of a Klatt-type formant synthesizer. The benefits of this system are several. In addition to requiring a relatively small number of parameters, the HLsyn model includes constraints on source–filter relations that occur naturally during speech production. Such constraints help to prevent combinations of sources and filter that are impossible to achieve with the human vocal tract. Thus, HLsyn could lead to reductions in the complexity of formant synthesis and result in better quality synthesis. HLsyn can also be a useful tool for speech-science education and speech research. This paper focuses on the generation of acoustic sources in HLsyn. Described in detail are the equations and methods used to estimate Klatt-type source parameters from HLsyn parameters. Several examples illustrating the generation of source parameters for obstruents (voiced and voiceless) and sonorants are provided. Future papers will describe the filtering components of HLsyn. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1498851]

PACS numbers: 43.72.Ja, 43.70.Bk [DOS]

I. INTRODUCTION

A. Source–filter theory and its articulatory correlates

Speech production involves two activities: maintenance of a relatively steady, positive pressure in the lungs, and formation of constrictions at various points throughout the laryngeal, pharyngeal, and oral cavities. For some sounds, the pressure between two of these constrictions may be manipulated by changing the volume between them. The positive lung pressure forces air to flow through the trachea and vocal tract. The formation of constrictions plays two roles: the location along the vocal tract affects the resonant qualities of the cavities, and when a constriction is narrow enough, pressure can increase or decrease in the airway behind it, causing a pressure drop across the constriction. This pressure drop forces air to flow through the constriction, creating sound sources. For example, given a lung pressure greater than atmospheric and appropriately configured vocal folds, subglottal pressure will build up and set the folds into a quasiperiodic mode of vibration. The resulting acoustic energy produced by the modulation of airflow through the glottis is then shaped by the cavities above the glottis. Adding a narrow constriction at some point downstream of the vocal folds not only changes the resonant qualities of the vocal tract but also results in turbulence in the airflow, giving rise to a second source of sound near the point of constriction. Sources above the glottis are usually noise sources, but could also be transient (at the release of stop consonants) or have a periodic component (e.g., trills formed at the tongue blade or lips). In addition to creating a second source of sound, the supraglottal constriction influences the amplitude and other characteristics of the periodic sound source at the glottis; likewise, the degree of adduction of the vocal folds influences the supraglottal source characteristics. Thus, the effects of local constrictions in the vocal tract are not restricted to the sources in the vicinity of the constriction, and the characteristics of one source are not independent of those of another source.

In terms of physiology, the sound sources and the resonant characteristics of the vocal tract are often produced and controlled by the same mechanisms, and are not separable. Acoustically, however, the characteristics of the sources and the transfer functions of the cavities formed by the articulators in the vocal tract can be treated separately, greatly simplifying models and analysis. This separation is the basis of the source–filter theory of speech production (Fant, 1960). Since its introduction, the source–filter theory has proven highly useful, productive, and informative, and our understanding of the acoustic characteristics of speech has been greatly enhanced. Treatment of source production and filtering as entirely independent processes is assumed in many models that form the basis for speech research, and in most automated speech applications.

Formant synthesis is a direct application of source–filter theory. A block diagram of a conventional source–filter speech synthesizer, the Klatt synthesizer (Klatt and Klatt, 1990), is shown in Fig. 1. Three sources are displayed at the left: one represents the periodic source at the glottis, one represents the aspiration noise source at the glottis, and one accounts for frication noise, that is, noise arising from turbulent airflow in the vicinity of a supraglottal constriction. (Figure 1 shows three alternative periodic sources in the Klatt synthesizer; in this paper, only the KLGLOTT88 source is used.) The amplitudes of these sources are con-

[a)]Electronic mail: hanson@speech.mit.edu
[b)]Also at the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139.
[c)]Also at the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

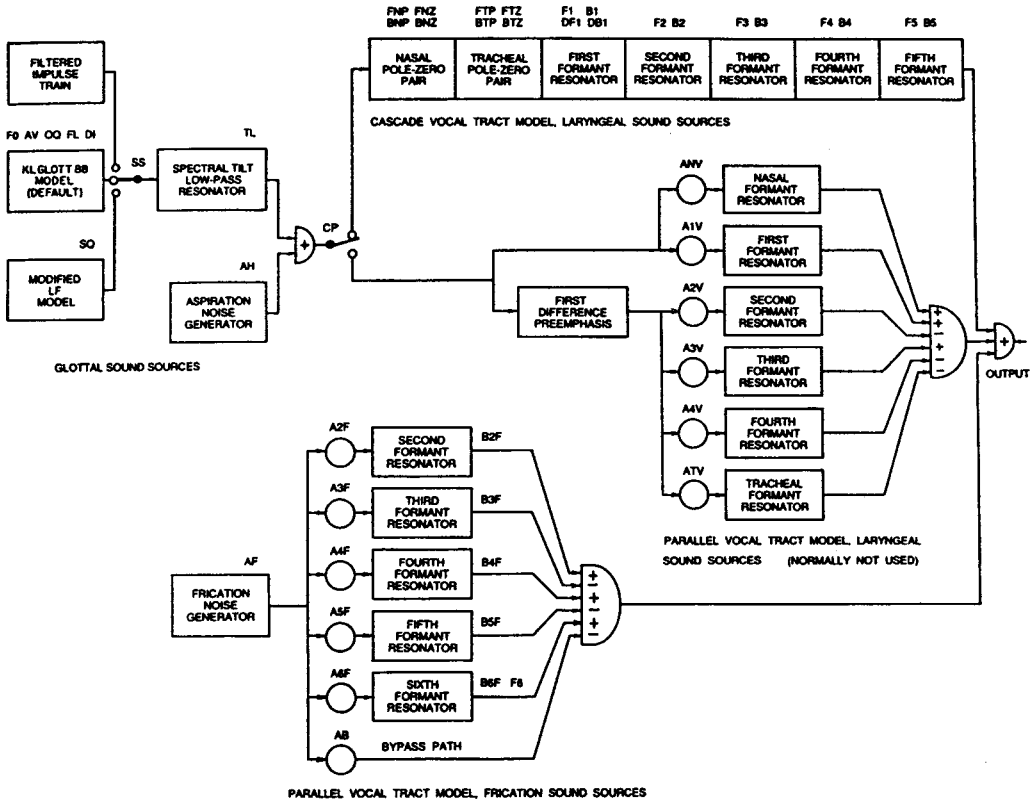


FIG. 1. Block diagram of the Klatt synthesizer (after Klatt and Klatt, 1990).

trolled by three separate parameters: AV (amplitude of the quasiperiodic component of the glottal source), AH (amplitude of the aspiration source), and AF (amplitude of the frication source). The characteristics of the quasiperiodic source are also controlled by a parameter $F0$ that manipulates the fundamental frequency, parameters (TL and OQ) that determine the shape of the glottal pulse waveform, and a parameter (DI) that simulates the amount of diplophonia. These three sources excite either cascade or parallel banks of filters, which represent the vocal-tract resonances. Each resonance has a formant frequency Fi, and resonances in cascade have bandwidths Bi, where $i=1,...,6$. Each of the parallel resonators includes an amplitude (AiV when excited by the glottal source, and AiF when excited by the frication source), as well as bandwidths BiV and BiF. In addition, in the cascade branch there are resonators to represent pole–zero pairs introduced by coupling to the nasal cavity, the trachea, or other side branches. In all, about 50 parameters are necessary to control the synthesizer, though usually they do not all vary at the same time.

Although very high quality speech can be synthesized using formant synthesis (Holmes, 1983; Klatt and Klatt, 1990; Hanson, 1995b), the discussion above suggests a weakness in the system of Fig. 1, namely the lack of constraints that occur naturally among sound sources and the vocal-tract filter during speech production. In the source–filter conception in Fig. 1, the interdependencies and interactions of sources and filter must be implemented by proper selection and control of a multitude of synthesizer parameters. Perhaps more importantly, by treating the sources and vocal tract as completely independent, one could introduce combinations of parameters that result in sources and filters that are impossible to achieve with a human vocal tract. Articulatory synthesizers, which attempt to mimic exactly the production of speech by humans, would provide a solution to these problems. These synthesizers, however, have been found to be computationally expensive. Furthermore, there is a lack of articulatory data which form the basis for the design of control parameters for such a synthesizer.

B. Constraining source–filter relations in HLsyn

Alternative synthesis systems have been suggested. For example, Lin and Fant (1990) describe a quasiarticulatory system, which combines some aspects of articulatory and formant synthesis. Another such system, and the subject of this paper, was first described by Stevens and Bickley (1991). It employs only a small set of higher level (HL) parameters which are closely related to parameters that might be found in an articulatory synthesizer. The HL parameters are mapped to a large set of Klatt-type (KL) parameters. The parameters controlling the parallel resonators with laryngeal sources in Fig. 1 are not used in this implementation; rather, sounds produced with laryngeal sources are generated with the cascade branch. The mapping relations among HL and KL parameters are based on equivalent circuit and transmission line models of the vocal tract that include both aerodynamic and acoustic processes. Hence, it retains the computational simplicity of formant synthesis, while including the constraints among source characteristics and vocal-tract configuration.

This system has been implemented under the name HL-

TABLE I. Description of HLsyn parameters.

f1–f4	First four natural frequencies of vocal tract, assuming no narrow local constrictions (Hz)
f0	Fundamental frequency due to active adjustments of vocal folds (Hz)
ag	Average area of glottal opening between the membranous portion of the vocal folds (mm^2)
ap	Area of the posterior glottal opening (mm^2)
ps	Subglottal pressure (cm H_2O)
al	Cross-sectional area of constriction at the lips (mm^2)
ab	Cross-sectional area of tongue-blade constriction (mm^2)
an	Cross-sectional area of velopharyngeal port (mm^2)
ue	Rate of increase of vocal-tract volume (cm^3/s)
dc	Change in vocal-fold or wall compliances (%)

syn (Bickley *et al.*, 1997; Hanson *et al.*, 1997). It controls an existing formant synthesizer, SenSyn™ (Sensimetrics Corp., 1992), which is modeled after KLSYN88 (Klatt and Klatt, 1990). The current set of 13 parameters is listed in Table I and the parameters are illustrated in Fig. 2(a). A block diagram of the HLsyn system is shown in Fig. 2(b). Each parameter represents either an articulatory attribute or an acoustic parameter that is directly related to articulation.[1] The parameters **f1**, **f2**, **f3**, and **f4** represent the natural frequencies of the vocal tract, assuming that there is no coupling to the tracheal or nasal cavities. The cross-sectional areas of constrictions formed by the lips or tongue blade are controlled by parameters **al** and **ab**, respectively, while the cross-sectional area of a tongue-body constriction is controlled by **f1**.[2] The parameter **ag** represents the cross-sectional area at the membranous portion of the vocal folds, while **ap** represents that at the cartilaginous portion. An initial fundamental-frequency contour is set using the parameter **f0**. This contour represents intentional prosodic variations in fundamental frequency that are achieved by varying vocal-fold tension. Perturbations on this contour are made in the mapping relations to account for intrinsic pitch of vowels, changes in subglottal pressure, and changes in vocal-fold compliance that may occur when obstruent consonants are produced. The remaining parameters control the opening at the velopharyngeal port (**an**), the subglottal pressure (**ps**), the rate of expansion of the vocal tract (**ue**) (used during the production of obstruent consonants), and changes in the compliance of the vocal-tract walls and vocal folds, expressed as a percentage change relative to the default compliances (**dc**).

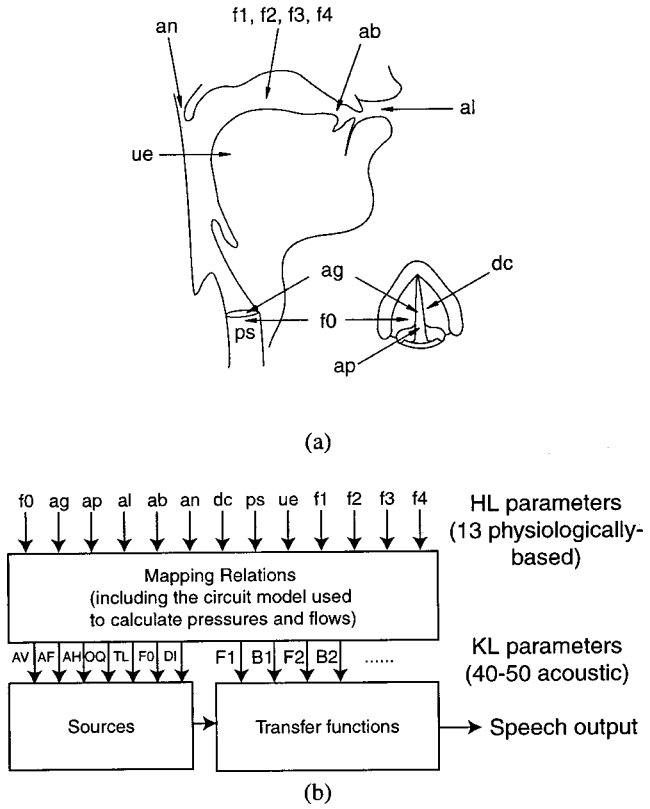


FIG. 2. (a) Illustration of the HLsyn parameters and their relation to the vocal tract. (b) Schematic of the HLsyn synthesizer.

As shown in Fig. 2(b), these parameters are used to first determine the pressures and flows that would develop in a vocal tract having the specified time-varying configuration. These pressures and flows, along with the original parameters, are then used to calculate the KL parameters for sources at the vocal folds and in the supraglottal tract (including the amplitudes AV, AH, and AF), as well as those for the transfer functions. Because the Klatt synthesizer is language independent, HLsyn is not oriented towards one language. However, some of the constants in the mapping relations, to be described later, may need to be adjusted for certain languages.

Figure 3 illustrates the calculation of source amplitudes for two simple time-varying adjustments of the glottis and of a supraglottal constriction that might occur for a consonant in intervocalic position.[3] In the first example [Figs. 3(a) and (b)] the vocal folds remain in an adducted position throughout the consonant (that is, **ag** remains constant at a value appropriate for a vowel), and a narrow constriction is formed at the alveolar ridge. We model this configuration in HLsyn by first setting the average glottal area **ag** to be 4 mm^2. The cross-sectional area of the constriction formed by the tongue blade, **ab**, is initially set to 100 mm^2, to indicate the lack of a significant constriction there. At 100 ms, however, **ab** begins to decrease. It reaches a minimum of 10 mm^2 at 125 ms, and is held there for 50 ms. At 175 ms, **ab** begins to increase, until it reaches 100 mm^2 at 200 ms.

The effects of this simple parameter change are several. During the closure, oral pressure builds up behind the constriction, reducing the transglottal pressure, but introducing a pressure drop at the narrow constriction. Therefore, in the output Klatt parameters, we would see AV drop during the closure, while AF would rise, as shown in Fig. 3(b). Note that if one were controlling the Klatt parameters directly, both AF and AV could be increased simultaneously, contrary to what one expects based on acoustic models (assuming that subglottal pressure is constant). In HLsyn, the constraints among the source amplitudes prevent this anomaly. The increase in oral pressure creates an abducting force on the vocal folds, and results in an increase in the effective average glottal area during the closure, causing an increase in acoustic energy loss at the glottis. In terms of the KL parameters, this additional loss of energy means that the bandwidths of the formants will increase, particularly the first-formant

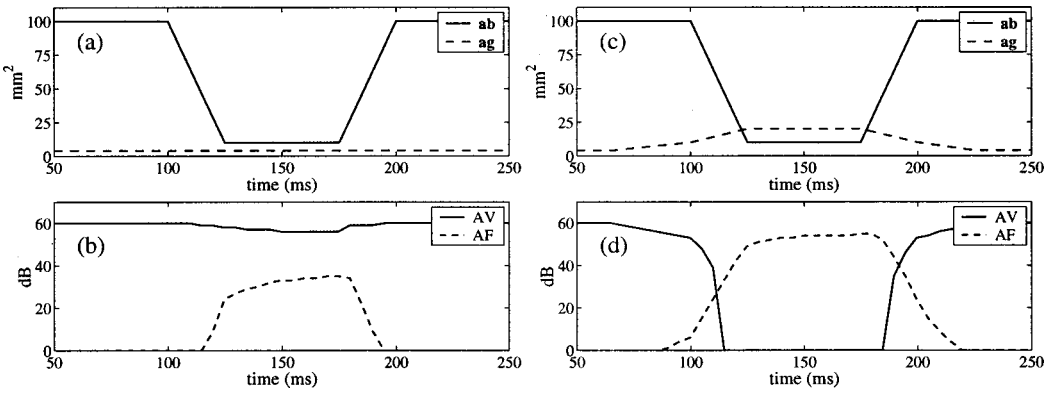


FIG. 3. Two examples of how variation of quasiarticulatory HLsyn parameters results in variation of acoustic Klatt-synthesizer parameters. (a) The variation in HLsyn parameter **ab** models the formation of an incomplete constriction formed by the tongue blade; parameter **ag** is held constant to model maintenance of the vocal folds in an adducted position. (b) The Klatt parameters AV and AF generated by HLsyn. (c) The parameter **ag** is increased during the constriction to model abduction of the vocal folds. (d) The output Klatt parameters AV and AF differ significantly from those in (b). (See the text for more details.)

bandwidth B1. Another effect of the increase in effective average glottal area is that formant frequencies, especially $F1$, are shifted. Thus, at least four KL parameters (AV, AF, $F1$, B1) related to both source and filter are affected by the variation in the HL parameter **ab**.

If, in addition to varying **ab**, **ag** was increased substantially during the closure, AV would decrease to zero[4] during the closure, while AF would be relatively high, as might be expected for a voiceless fricative. These effects are illustrated in Figs. 3(c) and (d).

The details of deriving time-varying KL parameters from HL parameters are the subject of a planned series of papers. In these papers we will describe in depth the parameters and mapping relations of the HLsyn system.[5] Because it is largely the differences in the derivation of source parameters that set HLsyn apart from traditional formant synthesis, we focus on acoustic sources in this paper, and derive the KL parameters AV, AF, AH, OQ, TL, $F0$, and DI. Given the interactions among sources and filters that are the basis of the system, however, the paper will necessarily include discussion of parameters related to the transfer function characteristics. Future papers will describe the derivation of KL filter parameters in more depth.

The paper is organized as follows. In the next section, the generation of acoustic sources in the vocal tract via formation and release of constrictions is described in some detail. The equations and methods with which KL source parameters are estimated from HL parameters are presented in Sec. III. Section IV presents examples of how various sources are generated in HLsyn. Finally, summary and discussion are presented in Secs. V and VI.

II. ACOUSTIC SOURCES

As discussed earlier, both periodic and noise sources can be generated in the vocal tract through the formation and release of constrictions, either at the glottis or somewhere along the supraglottal tract. Acoustically, the relevant parameters are amplitude and, in the case of periodic sources, spectral balance, open quotient, and fundamental frequency. In order to derive these acoustic parameters from articulatory parameters, it is necessary to understand how speakers control these sources. The control issues include the circumstances under which sources are induced, sustained, and inhibited, and the manner in which source characteristics such as amplitude are varied.

The properties of the quasiperiodic and noise components of the glottal source are determined by the configuration of the glottis, the stiffness or slackness of the vocal folds, and the pressure drop across the glottis (see, for example, Fant, 1982; Holmberg *et al.*, 1988; Titze, 1992). The glottal configuration can be adjusted in several ways, particularly in the degree of abduction or adduction of the vocal folds and of the arytenoid cartilages. There appear to be two main components to these abduction or adduction movements: control of the separation of the vocal folds at the vocal processes of the arytenoid cartilages, and control of the space between the arytenoid cartilages. Each of these two components can be varied somewhat independently.

During vowels and certain sonorant consonants, the configuration of the vocal tract above the glottis usually has very little influence on the acoustic source at the glottis. Under these circumstances, the amplitude of the glottal excitation increases with the subglottal pressure P_s (roughly proportional to $P_s^{1.5}$; see, for example, Ladefoged, 1962, and Isshiki, 1964), and the amplitude also depends on the amount of glottal abduction or adduction. There is an optimum positioning of the vocal folds that leads to a maximum amplitude for the quasiperiodic source (Verdolini *et al.*, 1998). Vibration of the vocal folds occurs only over a restricted range of glottal abduction or adduction. There are also transglottal pressure thresholds below which glottal vibration cannot be sustained or induced (Titze, 1988, 1992). These thresholds depend not only on the degree of separation of the vocal folds but also on the stiffness of the folds. Thus, the phonation threshold pressure is higher when a speaker increases the pitch of her voice by increasing the vocal-fold stiffness. The threshold pressure for initiating phonation is higher than that for sustaining phonation (Baer, 1975; Lucero, 1995). Typical phonation thresholds are in the range 3–5-cm H_2O for initiation of phonation and 2–3-cm H_2O for termination.

The waveform shape of the glottal pulses depends also on the glottal abduction or adduction, with the amount of downward tilt of the spectrum of the pulses being greater for

an abducted glottis and less for an adducted glottis. The presence of a fixed chink between the arytenoid cartilages also leads to an increased tilt in the spectrum of the quasiperiodic source (Hanson, 1995a, 1997). The frequency of this source is dependent on the stiffness of the vocal folds and is also dependent on the subglottal pressure, particularly when the vocal folds are relatively slack (Titze, 1989).

For certain combinations of transglottal pressure, vocal-fold tension, and degree of vocal-fold approximation, vibration of the folds may become irregular. Glottalization, which can occur if the folds are pressed together, is one example of such irregular vibration. Slifka (2000) has observed that irregular vibration in phrase-final position may be accompanied by falling subglottal pressure and rising glottal flow, suggesting that irregular vibration can also occur while the folds are moving apart, rather than being pressed together.

The noise component of the glottal source increases with the subglottal pressure (roughly proportional to $P_s^{1.5}$) and with the average cross-sectional area of the glottis A_g (proportional to $A_g^{0.5}$) (Stevens, 1971, 1998). The bandwidths of the formants, particularly the first formant, also increase with the degree of abduction of the vocal folds, due to associated increases in energy loss.

During obstruent consonants a constriction is formed in the vocal tract above the glottis. Consequently, when the glottis is not closed (and when the velopharyngeal port is not opened) there is an increased pressure in the vocal tract behind the constriction. The transglottal pressure, then, is reduced, and this reduced transglottal pressure tends to inhibit continued vocal-fold vibration. Different actions must be taken depending on whether vocal-fold vibration is to be maintained or is to be inhibited. For a voiced obstruent that occurs in intervocalic position, the vocal folds must be maintained in an adducted position, and the transglottal pressure must remain above the phonation threshold pressure if vocal-fold vibration is to be maintained. During the obstruent interval, the vocal folds may be slackened, and this action facilitates continued glottal vibration in spite of the reduced transglottal pressure. In the case of a voiceless obstruent in intervocalic position, the vocal folds are usually adjusted to an abducted position at the end of the preceding vowel, causing glottal vibration to cease when the transglottal pressure is reduced. The vocal folds are stiffened during the obstruent interval, thereby preventing vocal-fold vibration. The slackening or stiffening action for voiced and voiceless obstruents can carry over into the following vowel, causing a lowering or a raising of the fundamental frequency in the first few tens of milliseconds of the vowel.

Another means for inhibiting glottal vibration during the constricted interval for obstruent consonants is to further decrease the transglottal pressure by increasing the pressure in the mouth. This condition can be achieved by stiffening the walls of the vocal tract. Stiffening of the walls minimizes expansion of the vocal-tract volume when there is an increase in the supraglottal pressure and hence prevents significant flow of air through the glottis. On the other hand, glottal vibration can be facilitated by slackening the walls of the vocal tract, thus allowing passive expansion of the vocal-tract volume (Svirsky *et al.*, 1997). Continued airflow through the glottis also occurs if the speaker actively expands the vocal-tract volume. This active expansion can be achieved by lowering the larynx or by advancing the tongue root through contraction of the lower fibers of the genioglossus muscle (Bell-Berti, 1975).

In addition to the sources at the glottis, frication noise sources due to turbulent airflow can be generated in the vicinity of constrictions formed above the glottis by the lips, the tongue blade, or the tongue body. If the pressure drop across the constriction is ΔP and the cross-sectional area of the constriction is A_c, then the amplitude of the frication noise source near the constriction is roughly proportional to $\Delta P^{1.5} A_c^{0.5}$. In order to achieve maximum amplitude for the frication noise at the supraglottal constriction, it turns out that the cross-sectional area of this constriction should be about one-half that of the glottal opening (Stevens, 1998). The amplitude of the frication source is also somewhat dependent on the presence and orientation of obstacles (such as the lower incisors) against which the airflow impinges. In the case of a stop consonant, a brief frication source is generated near the consonant release, when the cross-sectional area of the consonantal constriction increases rapidly. Another type of source that occurs at the release of a stop or affricate consonant is a transient source. This source is a brief airflow pulse that is generated in the initial few tenths of a millisecond following the release. This source has been implemented in recent versions of KLsyn, and has been shown to yield improved naturalness for synthesized stop consonants and affricates (Massey, 1994). The current version of HLsyn does not include a transient source.

In summary, the generation of glottal sources in the human speech production system is a consequence of the positioning of the vocal folds and adjusting their stiffness in particular ways. When a pressure drop is created across the glottis, vocal-fold vibration may occur if certain conditions of stiffening and positioning of the folds are maintained. An aspiration noise source may also be generated in the vicinity of the glottis, with or without simultaneous glottal vibration. Likewise, a frication noise source is generated in the vicinity of a supraglottal constriction if there is a sufficient pressure drop across this constriction. In a quasiarticulatory synthesis system like HLsyn, then, the sources are not controlled directly, as they are in KLsyn, but rather they are determined through appropriate adjustments of parameters representing or controlling the subglottal pressure, the configuration of the glottis, the state of the vocal folds, the area of a supraglottal constriction, and, in the case of obstruent consonants, the vocal-tract volume and the stiffness of the vocal-tract walls. All of these actions have their counterpart in particular control parameters within HLsyn.

III. FROM HL TO KL PARAMETERS

In Sec. II, we described in general terms the factors that influence the properties of the quasiperiodic glottal source and the noise sources that are generated in the vicinity of vocal-tract constrictions. We turn now to a more detailed

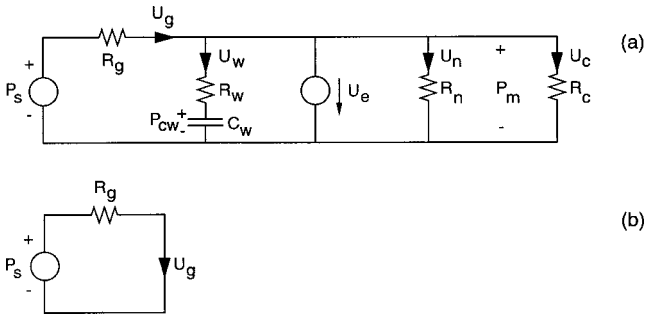


FIG. 4. (a) The low-frequency equivalent-circuit model that is the basis of HLsyn. (See the text for details.) (b) A simplified version of (a) holds for non-nasalized unconstricted vocal-tract configurations.

description of the calculation of these source parameters in the Klatt synthesizer when the HL parameters are specified.

As we have shown, the sound sources in speech production depend on the pressures and flows in the vocal tract, which in turn depend on vocal-tract constrictions and subglottal pressure. The basis of HLsyn is the low-frequency equivalent-circuit model of the vocal tract (Rothenberg, 1968; Stevens, 1993) that is shown in Fig. 4(a). In this circuit, the acoustic resistances R_g, R_n, and R_c are determined by the cross-sectional areas of the constrictions at the glottis (A_g), the velopharyngeal port (A_n), and the oral supraglottal airway (A_c), respectively. The glottal resistance R_g in the equivalent circuit consists of a viscous and a kinetic component. For most conditions of interest, the viscous component is small compared with the kinetic component, and can be neglected. The nasal and oral constriction orifices are approximated solely by kinetic resistances. Each kinetic resistance is of the form

$$R_{\text{kin}}=\frac{\rho U}{2A^2}, \tag{1}$$

where U is the volume velocity through the orifice, A is the cross-sectional area of the opening, and ρ is the density of air ($0.001\,14$ gm/cm^3).

The volume velocities U_g, U_n, and U_c through the constrictions are related to the pressure drop ΔP across the dynamic resistances according to the equation

$$U=A\sqrt{\frac{2\Delta P}{\rho}}, \tag{2}$$

where A is the cross-sectional area of the orifice.

The volume velocity U_w represents the rate of increase of the vocal-tract volume due to passive expansion or contraction of the vocal-tract walls. This expansion is a consequence of the increase or decrease of the intraoral pressure P_m. When P_m increases, for example, following the formation of a narrow supraglottal constriction, the compliant walls are pushed outward by an amount determined by the compliance C_w and resistance R_w of the walls; with the release of the constriction, P_m decreases and the walls spring back to their resting position. The source U_e, which has units of volume velocity, such as cm^3/s, represents active expansion or contraction of the vocal-tract volume by the talker. Positive values of U_w and U_e correspond to outward flow, or expansion of the oral cavity, while negative values correspond to contraction of the oral cavity.

Given the two sources P_s and U_e, the resistances of the constrictions, and the compliance and resistance of the vocal-tract walls, one can calculate the flows and the intraoral pressure P_m. For the production of most vowels and many sonorant consonants, there is not a narrow constriction in the oral cavity and the velopharyngeal port is closed, so that $R_c=0$ and $R_n=\infty$ in the equivalent circuit. In this case, the equivalent circuit reduces to the simplified version in Fig. 4(b), in which there is only one resistance, R_g, representing the glottal opening. In this case, $P_m=0$.

In HLsyn, the circuit shown in Fig. 4(a) has been implemented in software. The circuit variables P_s and U_e are specified directly by the input parameters **ps** and **ue**. The resistances R_c, R_n, and R_g are based on constriction sizes, some of which are specified directly by the user, and some of which are derived from the HL parameters. The resistance R_c is that associated with the narrowest of the supraglottal constrictions (excluding the one at the nasal port). The resistance of the walls R_w is a constant. The compliance of the walls C_w is derived from a modal value and the HL parameter **dc**. When values for the constriction sizes and C_w have been determined, the circuit is solved for the pressures and flows. Finally, the transglottal and intraoral pressures, together with the constriction sizes, are used to calculate the KL source parameters.

In the remainder of this section, we describe in detail the derivation and calculation of the circuit elements, pressures, and flows, and of the KL source parameters. We first present the derivation of the constriction sizes, as the resistances of the circuit depend on these, and the wall compliance. Next, the derivation of the pressures and flows, which depend on the resistances and wall compliance, are discussed. Finally, the calculation of KL parameters, based on the constriction sizes and the pressure drops at those constrictions, is given.

These calculations rely, to a large extent, on certain constants that can be specific to a gender, or to an individual. We refer to these constants as *speaker constants*. The speaker constants can be divided into several general categories. One group of constants is related to physical characteristics of an individual, such as vocal-tract dimensions and tissue characteristics. Default values for a subset of this group are given in Table IV. Other groups include acoustic constants such as neutral formant and bandwidth values (Table V); constants related to source characteristics (Table VI); threshold values, such as phonation threshold pressure (Table VII); and scale factors used during the calculation of KL parameters from pressures and constriction sizes (Table VIII). These speaker constants can be changed by the user of HLsyn. The examples given in this paper are based on the default values, except where specified.

In the following, we refer to several kinds of parameters. To help reduce confusion, we have used different typefaces for the different parameter types. As in the previous sections, HL parameters are written in **bold** and KL parameters in ALL CAPS. Parameters that are derived from HL parameters and then used to calculate KL parameters are referred to as

intermediate parameters, and are written with a sans serif font. Speaker constants are written in *slanted* (italic) type.

A. Constriction sizes

There are three types of constrictions that are of importance: those that occur at the glottis, one that occurs at the velopharyngeal port, and those that are formed with the lips, the tongue blade, or the tongue body. Strictly speaking, the constriction at the velopharyngeal port is a supraglottal constriction, but for the sake of convenience we will refer to it as the velopharyngeal constriction, and to those formed by the lips and tongue as supraglottal constrictions. It is the latter set of constrictions that determine the value of R_c in the circuit model of Fig. 4.

The constriction at the velopharyngeal port is specified by the HL parameter **an**. At the glottis, constrictions can be formed by adducting the membranous portions of the folds and the arytenoid cartilages. In HLsyn, the constriction at the posterior end of the vocal folds is controlled directly by the parameter **ap**. Because intraoral pressure can affect the degree of constriction at the membranous portion of the vocal folds, the parameter **ag** requires adjustment. This intermediate value of glottal area is referred to as agx.

Constrictions formed by the lips and tongue blade (except for liquid consonants) are directly controlled by the HL parameters **al** and **ab**, respectively. There are two other ways to form supraglottal constrictions. First, a constriction can be formed using the tongue body to create a narrowing in the oral cavity or in the pharynx. When there is not a localized constriction formed by the lips, tongue blade, or tongue tip, and the velopharyngeal port is not open, the first two or three natural frequencies of the vocal tract are rather directly related to the position of the tongue body.[6] Therefore, if the formants and the approximate length of the vocal tract are known, an estimate of the constriction formed by the tongue body can be obtained. In HLsyn, rather than specifying the area of this constriction directly, the parameter **f1** is used to estimate the size of the tongue body constriction, and this intermediate parameter is called acd. A constriction is also formed by the tongue blade during the production of liquid consonants. This constriction can become a source of frication noise when there is a relatively large airflow in the vicinity of a voiceless aspirated consonant, for which the glottis must be spread. The size of this constriction for a liquid consonant can also be derived from **f1**, and is referred to as acl.

In summary, the HL parameters **ap**, **al**, **ab**, and **an** refer to the cross-sectional areas of specific constrictions in the vocal tract. In HLsyn, these areas are not influenced by an interaction with other HL parameters or by pressures within the vocal tract. However, these parameters, together with **ag** and the formant frequencies, are used to derive five intermediate area parameters: acd, the cross-sectional area of the constriction formed by the tongue body; acl, the cross-sectional area of the constriction formed by the tongue blade when it is in a configuration for a liquid consonant; acx, the primary supraglottal constriction, which is the minimum value of **al**, **ab**, acl, and acd; agx, the cross-sectional area of the glottal opening at the membranous portion of the vocal folds; and agf, the total cross-sectional area of the glottal opening, including **ap**. (When the vocal folds are not vibrating, agf is the actual area of the glottis; when the vocal folds are vibrating, agf is the average glottal area over a cycle of vibration.) We now give a detailed description of the derivation of these intermediate areas.

1. Glottal areas, agx and agf

The HL parameter **ag** represents the intended cross-sectional area of the glottis, in the absence of intraoral pressure that is different from atmospheric. When intraoral pressure begins to build up, however, a force is applied on the vocal folds from above the glottis, which could increase the glottal area, depending on the compliance of the vocal folds. Therefore, for the purpose of calculating the pressures and flows in the vocal tract, **ag** must be modified to account for this change. The resulting parameter is called agx, and it is computed using the equation

$$\mathsf{agx} = (\mathbf{ag} + 100\mathsf{Pm} \times \mathsf{Cg} \times 2L_g) \ \ \mathrm{mm}^2, \tag{3}$$

where Pm is the intraoral pressure in dynes/cm^2 (the derivation of which is presented in Sec. III C), Cg is the compliance of the folds in cm^3/dyne, and L_g is the effective horizontal length of the vocal folds in centimeters (Stevens, 1998). (The factor of 100 in the second term on the right converts the additional area to mm^2.) The default values for L_g are given in Table IV. The compliance Cg is itself variable, depending on the parameter **dc**, and it is calculated according to the equation

$$\mathsf{Cg} = C_{gm}(1 + K_{Cg}\mathbf{dc}/100) \ \ \mathrm{cm}^3/\mathrm{dyne}, \tag{4}$$

where the speaker constant C_{gm} specifies a modal value of vocal-fold compliance about which Cg varies and K_{Cg} is a weighting factor; the default values of these constants are 5×10^{-6} cm^3/dyne and 0.34, respectively. Normally, **dc** varies in the range from -150 to 150 percent so that Cg varies from 2.5×10^{-6} to 7.6×10^{-6} cm^3/dyne, that is, about a factor of 3 (Svirsky *et al.*, 1997). Note that the HLsyn software does not allow Cg to be less than zero.

For example, if the oral pressure Pm is 4-cm H_2O and **dc** is zero, the effective glottal area at the membranous folds, agx, will be 3.9 mm^2 greater than the HL parameter **ag** for male speakers, and 2.7 mm^2 greater for females. Increasing the compliance of the folds by setting **dc** to 50% results in agx being 4.6 mm^2 greater than **ag** for males and 3.2 mm^2 greater for females. Note that if Pm is negative, agx is less than **ag**, that is, the folds are sucked together.

The intermediate parameter agf is simply the sum of agx and **ap**

$$\mathsf{agf} = \mathsf{agx} + \mathbf{ap}. \tag{5}$$

2. Size of the tongue body constriction, acd

The tongue body can form constrictions in the vicinity of the soft palate, the hard palate, or the pharyngeal wall. When it is a velar or a more fronted constriction, the first formant will be lower than some neutral value, while for a pharyngeal constriction, the first-formant frequency will be considerably higher than that neutral value. As stated above,

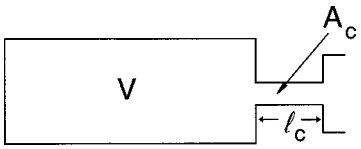


FIG. 5. Helmholtz resonator.

the cross-sectional area of the constriction, referred to as acd, is derived from the HL parameter **f1**. The derivation for a constriction in the oral cavity is based on an approximation of the vocal tract as a Helmholtz resonator. Such a resonator has a narrow neck and a relatively large, closed volume behind the neck, and is illustrated in Fig. 5. The volume behind the constriction is denoted V, and the length and area of the constriction are denoted l_c and A_c, respectively. The lowest natural frequency of this resonator is

$$f=\frac{c}{2\pi}\sqrt{\frac{A_c}{Vl_c}}\ \text{Hz}, \tag{6}$$

assuming the walls of the resonator to be hard. The walls of the vocal tract, however, have a finite acoustic mass. Consequently, when the areas of the constriction and of the glottal opening are zero, the natural frequency of the vocal tract is nonzero (Fant, 1972). We denote this minimum vocal-tract resonance as $f1_{\min}$, with a default value of 180 Hz. An equation for the first formant when the tongue-body constriction is in the velar region, including the effect of the walls, is then given by

$$\begin{aligned}\mathbf{f1}&=\sqrt{\left(\frac{c}{2\pi}\right)^2\frac{\text{acd}}{100V_{acd}L_{c\text{-}acd}}+f1_{\min}^2}\\&=f1_{\min}\sqrt{1+\left(\frac{c}{2\pi f1_{\min}}\right)^2\frac{\text{acd}}{100V_{acd}L_{c\text{-}acd}}}\ \text{Hz},\end{aligned} \tag{7}$$

where c is the velocity of sound in cm/s, the cross-sectional area A_c of the constriction is equal to acd in mm^2, the volume V is equal to V_{acd} in cm^3, and the constriction length l_c is equal to $L_{c\text{-}acd}$ in cm. Solving Eq. (7) for acd, we obtain

$$\text{acd}=100V_{acd}L_{c\text{-}acd}\left(\frac{2\pi f1_{\min}}{c}\right)^2\left[\left(\frac{\mathbf{f1}}{f1_{\min}}\right)^2-1\right]\ \text{mm}^2. \tag{8}$$

If $\mathbf{f1}<f1_{\min}$, acd is simply set to zero. Values for V_{acd} and $L_{c\text{-}acd}$ when the tongue-body constriction is in the velar region are specified by speaker constants; the default values are given in Table IV. Equations (7) and (8) are reasonable approximations for **f1** and acd as long as the dimensions of the vocal-tract resonator are small compared to the wavelength at frequency **f1**, that is, roughly $\mathbf{f1}<400$ Hz, or $\text{acd}<50$ mm^2. Under most circumstances in speech production, turbulence noise is only generated when acd is less than about 50 mm^2. Thus, any inaccuracies in estimating acd when $\text{acd}>50$ mm^2 will have essentially no influence on the source calculations to be described in Sec. III E. Note that when $\mathbf{f1}=f1_{\min}$, $\text{acd}=0$, that is, a complete velar closure is formed.

When a tongue-body constriction is formed in the pharyngeal region, the first-formant frequency becomes higher

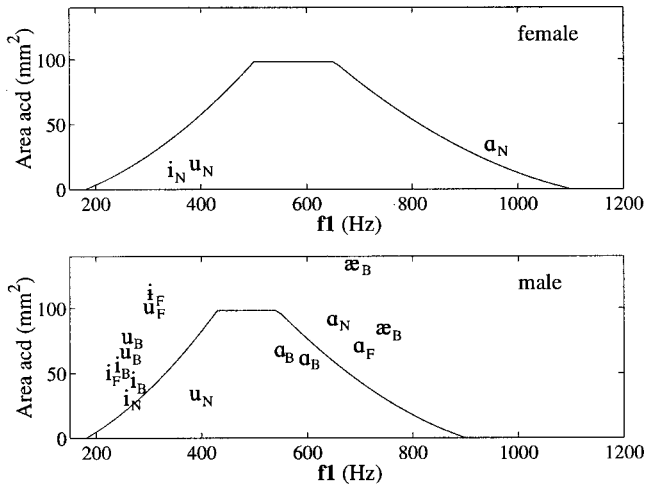


FIG. 6. The intermediate parameter acd, area of a constriction formed by the tongue dorsum, as a function of the HLsyn parameter **f1**, for adult females (top panel) and males (bottom panel). Superimposed on these curves are estimates of the minimum cross-sectional area of the vocal tract and the first-formant frequency for several vowels obtained from studies using x-ray or magnetic-resonance imaging: the subscript *F* indicates data from Fant (1960), the subscript *B* indicates data from Baer *et al.* (1991), and the subscript *N* indicates data from Narayanan *et al.* (1997). [Cross-sectional areas are presented graphically in Narayanan *et al.* (1997); actual values were obtained through communication with Narayanan.]

than its value for a vocal tract with a uniform cross-sectional area. We assume the following relation between acd and **f1** when **f1** is greater than a threshold frequency denoted $f1_{\text{phar}}$ (Table V):

$$\text{acd}=K_{\text{phar}}\left[\left(\frac{F1_{\max}-\mathbf{f1}+f1_{\min}}{f1_{\min}}\right)^2-1\right], \tag{9}$$

where $F1_{\max}$ is a speaker constant (Table VIII). Its default values are the maximum expected values of **f1**: 900 Hz for males and 1100 Hz for females. The constant K_{phar} is derived from appropriate speaker constants to ensure that acd will be 100 mm^2 when $\mathbf{f1}=f1_{\text{phar}}$. When $\mathbf{f1}=f1_{\max}$, $\text{acd}=0$. If the user sets **f1** to be greater than $F1_{\max}$, acd is also set to zero.

When **f1** is greater than the value for which $\text{acd}=100$ mm^2 in Eq. (7),[7] but less than the pharyngeal threshold frequency $f1_{\text{phar}}$, we assume that the tongue body does not form a narrow constriction at any point in the vocal tract and we therefore set acd to be 100 mm^2.

The complete equation for acd as a function of **f1** can then be summarized as

$$\text{acd}=\begin{cases}0, & \mathbf{f1}<f1_{\min}\\ \min[\text{Eq. (8)},\ 100], & f1_{\min}\leq\mathbf{f1}<f1_{\text{phar}}\\ \text{Eq. (9)}, & f1_{\text{phar}}\leq\mathbf{f1}\leq F1_{\max}\\ 0, & F1_{\max}<\mathbf{f1}.\end{cases} \tag{10}$$

Figure 6 shows plots of acd versus **f1**, based on the default values of the speaker constants for both females (top panel) and males (bottom panel). Superimposed on these curves are measurements of the first-formant frequency **f1** together with estimates of the minimum cross-sectional area acd of the vocal tract for some high and low vowels obtained from several studies with x-rays and with magnetic resonance imaging (Fant, 1960; Baer *et al.*, 1991; Narayanan

TABLE II. Typical formant frequencies for liquid consonants in American English produced by both males and females (Espy-Wilson, 1992; Stevens, 1998).

		f1 (Hz)	**f2** (Hz)	**f3** (Hz)
/l/	Female	450	1200	3000
	Male	400	1000	2800
/r/	Female	450	1200	1800
	Male	400	1100	1500

et al., 1997). There is a rough match between the estimated areas and **f1**, especially in the regions where acd is less than 50 mm^2, that is, cross-sectional areas for which turbulence noise is likely to be generated. It is noted, however, that when the area is small, estimates of cross-sectional area based on MRI or x-ray images may be subject to some error.

3. Constriction formed during a liquid consonant, acl

During the production of lateral and retroflex consonants, a relatively narrow constriction can be formed by the tongue blade. Although the constriction for a liquid consonant is narrower than that for a high vowel, it is usually not sufficient to create an increase in intraoral pressure of more than about 1-cm H_2O when the vocal folds are in a modal configuration, and therefore the only source is at the glottis. However, when the glottis is spread, as it might be during the production of a consonant cluster such as /kr/, the airflow can be sufficient to cause a pressure drop across the tongue-blade constriction, resulting in a frication noise source in the vicinity of that point in the oral cavity. The size of the constriction can be inferred from observations of turbulence noise that is generated in consonant clusters like /pr/ and /pl/. For these clusters, it is expected that the glottal opening at the /p/ release is about 30 mm^2. Because noise is usually generated at the lateral or retroflex constrictions immediately following the release of the aspirated consonant, it is expected that those constriction sizes are equal to or somewhat smaller than 30 mm^2. The constriction size for a high vowel is not usually less than about 30 mm^2 (Fant, 1960).

The pattern of formant frequencies for these liquid sounds usually lies outside the region that is occupied by vowels, glides, and other consonants. Typical values for the first three formants for lateral and retroflex configurations in American English are given in Table II (Espy-Wilson, 1992; Stevens, 1998). In HLsyn, when **f1**, **f2**, and **f3** lie near these frequency regions, it is assumed that there is a narrow constriction in the oral cavity due to the production of a liquid. The ranges used to determine the existence of this constriction are given in Table III. (Other languages and dialects of English may require adjustments. The user can change these ranges by setting the appropriate speaker constants in Table VII.)

We call the cross-sectional area of this constriction acl, and approximate it by the formula

$$\text{acl} = \left(\frac{\mathbf{f1}}{f1_{\text{liquid}}} \right)^2 \times K_{acl} \text{ mm}^2, \tag{11}$$

where $f1_{\text{liquid}}$ is a speaker constant based on the first-formant values reported in Table II, and K_{acl} is a scale factor having a default value of 25 mm^2. This equation is constructed to make acl somewhat sensitive to **f1** over the range of first-formant frequencies generally observed for liquid consonants. The square-law relation is selected in accord with the equation for the natural frequency for a Helmholtz resonator, which is proportional to the square root of the cross-sectional area of the neck of the resonator [Eq. (6)]. In general, acl will only influence the frication-source amplitude when the glottis is spread during a liquid, as in sequences like /pl/ or /tr/.

4. Supraglottal constriction, acx

As we have shown, there are four supraglottal constrictions that are considered in HLsyn: two, **al** and **ab**, are specified by HL parameters, while the other two, acl and acd, are derived from the HL parameters. In general, only one of these constrictions will be small enough to have a large pressure drop associated with it, and that constriction will be the source of noise generation. Hence, only one supraglottal constriction is included in the circuit model of Fig. 4. In HLsyn, after the intermediate parameters acl and acd are computed, a parameter acx is determined by taking the minimum of the four supraglottal constriction sizes

$$\text{acx} = \min(\text{acd}, \text{acl}, \mathbf{ab}, \mathbf{al}). \tag{12}$$

This parameter acx is then used to compute the resistance of the supraglottal constriction.

B. Compliance of the vocal-tract walls, Cw

The compliance of the walls Cw [C_w in Fig. 4(a)] varies about a modal value. This modal value is adjusted if the HL parameter **dc** is nonzero (as it might be during the production of obstruents), according to the equation

$$\text{Cw} = C_{wm}(1 + K_{Cw}\mathbf{dc}/100), \tag{13}$$

TABLE III. Formant ranges for which lateral and retroflex consonants are assumed.

Retroflex		Lateral	
Males	Females	Males	Females
350<**f1**<500	400<**f1**<550	350<**f1**<500	400<**f1**<550
f2<1400	**f2**<1600	**f2**<1300	**f2**<1400
f3<1800	**f3**<2000	**f3**>2700	**f3**>2900

where K_{Cw} is a weighting factor having a default value of 1.0. Default values of C_{wm} can be found in Table IV. The HLsyn software prevents Cw from becoming negative.

C. Pressures and flows

Having derived the glottal and supraglottal constriction sizes, and the compliance of the vocal-tract walls, we can now calculate the pressure drops across the constrictions and the airflows through them. By Kirchhoff's current law, we can write the following equation for the circuit of Fig. 4:

$$U_w = U_g - U_e - U_n - U_c . \tag{14}$$

Substituting Eq. (2) and the appropriate HLsyn parameters and intermediate parameters into Eq. (14), yields

$$\mathrm{Uw} = \left(\frac{\mathbf{ag} + \mathbf{ap}}{100} + 2\,\mathrm{Cg} L_g \mathrm{Pm} \right) \sqrt{\frac{2(980\mathbf{ps} - \mathrm{Pm})}{\rho}} - \mathbf{ue} - \frac{\mathbf{an} + \mathrm{acx}}{100} \sqrt{\frac{2\mathrm{Pm}}{\rho}}, \tag{15}$$

where Pm is the intraoral pressure in dynes/cm^2. The factor of 980 in the first term on the right-hand side of the equation converts cm H_2O to dynes/cm^2, and the factors of 100 in the first and third terms on the right convert mm^2 to cm^2. In this equation, everything is known but Pm and Uw. Additional equations can be written for the circuit of Fig. 4

$$\mathrm{Pm} = \mathrm{Pcw} + R_w \mathrm{Uw}, \tag{16}$$

$$\mathrm{Qw} = \mathrm{Cw}\,\mathrm{Pcw}, \tag{17}$$

$$\mathrm{Uw} = \dot{\mathrm{Q}}\mathrm{w}. \tag{18}$$

In Eq. (16), Pcw is the pressure drop across the wall compliance in dynes/cm^2 and R_w (the resistance of the walls) is a speaker constant, the default value of which is given in Table IV. Equation (17) gives the change in vocal-tract volume Qw due to the compliance of the vocal-tract walls, while Eq. (18) states that the rate of this change is the volume velocity of the yielding walls, Uw.

Equations (15)–(18) combine to form a system of four differential algebraic equations with four unknowns (Pm, Uw, Pcw, and Qw). Because this system is stiff (see, for example, Press *et al.*, 1986) for small values of Cw, an ''implicit'' scheme is used to approximate the solution, in which the mean flow Uw over a short time interval is approximated by the average of its values at the two ends of the interval. As a result, Eqs. (15)–(18) are replaced with the difference equations

$$\mathrm{Uw}(nT+T) = \left(\frac{\mathbf{ag} + \mathbf{ap}}{100} + 2\,\mathrm{Cg} L_g \mathrm{Pm}(nT+T) \right) \times \sqrt{\frac{2(980\mathbf{ps} - \mathrm{Pm}(nT+T))}{\rho}} - \mathbf{ue} - \frac{\mathbf{an} + \mathrm{acx}}{100} \sqrt{\frac{2\mathrm{Pm}(nT+T)}{\rho}}, \tag{19}$$

$$\mathrm{Pm}(nT+T) = \mathrm{Pcw}(nT+T) + R_w \mathrm{Uw}(nT+T), \tag{20}$$

$$\mathrm{Qw}(nT+T) = \mathrm{Cw}(nT+T)\,\mathrm{Pcw}(nT+T), \tag{21}$$

$$\mathrm{Qw}(nT+T) = \mathrm{Qw}(nT) + \frac{\mathrm{Uw}(nT) + \mathrm{Uw}(nT+T)}{2} \cdot T. \tag{22}$$

The unknowns $\mathrm{Uw}(nT+T)$, $\mathrm{Pcw}(nT+T)$, and $\mathrm{Qw}(nT+T)$ can be eliminated, leaving a single equation involving only $\mathrm{Pm}(nT+T)$ and known values ($\mathrm{Qw}(nT)$, $\mathrm{Uw}(nT)$, **ag**, etc.). Brent's algorithm (see, for example, Press *et al.*, 1986) is used to solve for $\mathrm{Pm}(nT+T)$, whereupon Eqs. (19)–(21) can be used to find $\mathrm{Uw}(nT+T)$, $\mathrm{Pcw}(nT+T)$, and $\mathrm{Qw}(nT+T)$, and the cycle can be repeated.

The time increment T can be set by the user. By default, $T = 0.1$ ms is used in order to capture the effects of the buildup of oral pressure when there is a rapid closure of a vocal-tract constriction, and the reduction of oral pressure when there is a rapid release of a constriction. Initial conditions corresponding to zero pressure in the oral cavity and no pressure drop across the capacitance are assumed at the onset of an utterance.

D. Examples of pressures and flows derived from HL parameters

As an illustration, we return to the example introduced in Sec. I and Fig. 3, in which the HL parameter **ab** was varied to create the effect of forming a narrow opening at the alveolar ridge. In the first part of the example, **ag** was held constant at its modal value, as it might be for a voiced obstruent, and in the second part, **ag** was increased during the constriction, as it might be for a voiceless obstruent. All other parameters were held constant at their default values. Therefore, during the time that the constriction is formed, the supraglottal constriction acx is equal to **ab**.

Let us first consider the example for which both **ab** and **ag** vary; the parameter tracks for **ab** and **ag** are given again in Figs. 7(a) and (b) for completeness. Figure 7(c) illustrates the resulting variation in oral pressure Pm, which increases to as much as 6.8-cm H_2O during the closure, well within the range that one might expect for adult speakers (Arkebauer *et al.*, 1967). Because of the increase in Pm, the glottal area agx is greater than **ag**. The parameter track for agx is superimposed on Fig. 7(b). Figure 7(d) shows the volume velocities through the glottal opening (Ug) and the supraglottal constriction (Uc). As one might expect, Ug initially increases as agx increases, but decreases somewhat as Pm builds up. At the release of the supraglottal constriction, Ug first increases because the transglottal pressure increases, but then decreases as agx is reduced. Likewise, Uc increases as agx increases, and then decreases due to the decrease in acx (**ab** in this case). At the release of acx, there is a rapid initial increase in Uc due to the pressure Pm across the constriction. However, Uc then falls off as Pm decreases and agx decreases. The intermediate parameter Uw is also shown in Fig. 7(d); it increases as Pm increases and pushes on the vocal-tract walls, and becomes negative when Pm decreases, representing the return of the walls to their resting position. Recall from Fig. 4 that $U_c = U_g - U_w$; when U_w is negative, the flow through the supraglottal constriction is greater than that through the glottis. Thus, as seen in Fig. 7(d), the yield-

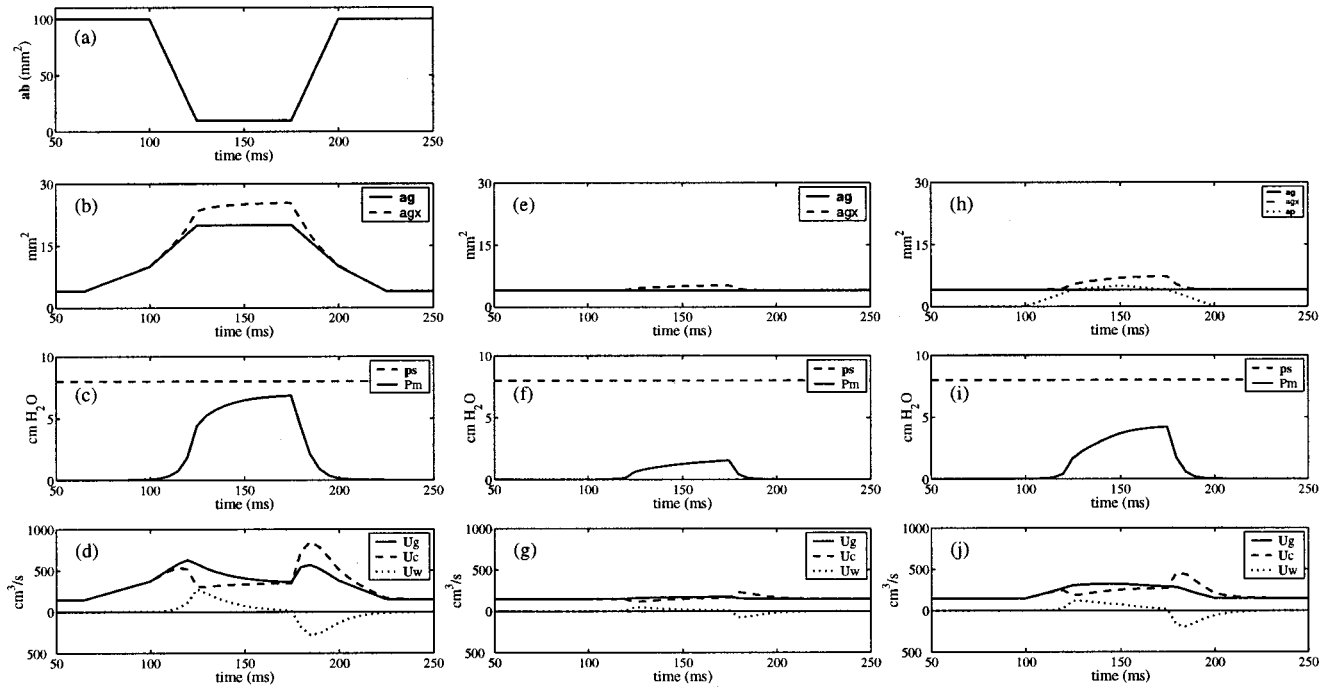


FIG. 7. Examples of the mapping of HLsyn parameters to intermediate pressure and flow parameters. Column 1: Panels (a)–(d) correspond to the example illustrated in Figs. 3(c) and (d). Column 2: Panels (e)–(g) correspond to the example illustrated in Figs. 3(a) and (b). Column 3: Panels (h)–(j) illustrate how use of the parameter **ap** results in variations in the pressure and flow variables more suitable for voiced obstruents. Row 1: Parameter **ab**, emulating the formation of a constriction by the tongue blade. Row 2: Parameter **ag**, the average area at the membranous portion of the vocal folds; intermediate parameter agx, the effective glottal area at the membranous folds; and **ap**, representing the area of a posterior glottal opening. Row 3: Oral pressure Pm and subglottal pressure **ps**. Row 4: Volume velocities Ug through the glottis, Uc through the supraglottal constriction, and Uw representing the expansion of the vocal-tract walls due to the increase in oral pressure. (See Sec. III D for details.)

ing walls make a significant contribution to the airflow Uc through the constriction at the time of the consonant release (Stevens, 2001).

Next, we consider the example for which **ab** varies but **ag** remains at its modal value [second column in Fig. 7, parts (e)–(g)]. In Fig. 7(f), we see that the oral pressure increases during the constriction, with a maximum value of about 1.5-cm H_2O, considerably less than the approximately 4-cm H_2O reported by Arkebauer *et al.* (1967) for voiced alveolar continuants. Consequently, the increase in agx and the decrease in transglottal pressure are negligible. In addition, variations in the volume velocities Uc and Ug and in the velocity of the vocal-tract walls Uw are minor. If, however, the parameter **ap** is increased to 5 mm^2 during the constriction, the situation improves, as illustrated in the third column of Fig. 7, parts (h)–(j). The oral pressure increases to a maximum of 4-cm H_2O during the constriction [Fig. 7(i)]. As a result, the increase in agx and the decrease in transglottal pressure are significant, although still less than in the unvoiced case. Likewise, Ug and Uc vary to a greater degree. In a later example (Sec. IV A), we show that use of the parameter **ap** is necessary to obtain values for the acoustic parameters AV and AF appropriate for voiced continuants, precisely due to the aerodynamic effects shown here.

Similar results would occur if the constriction were formed at the lips or by the tongue dorsum (using the parameters **al** or **f1**, respectively), or if the constriction corresponded to a liquid consonant (using the parameters **f1**, **f2**, or **f3**).

E. KL source parameter equations

We have described the calculation of the cross-sectional areas of the vocal-tract constrictions, as well as the pressure drops across those constrictions and the airflows through them, according to the HLsyn model. Through examples, we have demonstrated the application of the model to a narrowing along the supraglottal airway. Various glottal configurations, in combination with that narrowing, result in quite different aerodynamic conditions in the vocal tract. The output of the model seems to agree with experimental data reported in the literature (e.g., Arkebauer *et al.*, 1967). Once the aerodynamic quantities are determined, appropriate values of the Klatt acoustic parameters can be calculated. In this section we describe the derivation of the Klatt parameters AV, AF, AH, OQ, TL, and DI. In addition, determination of the fundamental frequency ($F0$) by adjustment of the HL parameter **f0** is presented. The mapping relations between the HL and intermediate parameters, and between the intermediate and KL parameters are based on our knowledge and assumptions about the generation of acoustic sources, as described in Sec. II.

The pressure in the oral cavity, Pm, is the principal intermediate parameter of the model. This oral-cavity pressure, together with the orifice areas and the subglottal pressure **ps**, are the intermediate parameters that determine the KL parameters controlling the amplitude of the voicing (AV), frication (AF), and aspiration (AH) sources, as well as the shape of the voicing source (OQ and TL). Under certain circumstances, irregularities of a diplophonic nature can be

added to the voicing source; the magnitude of these irregularities (DI) depends mainly on the glottal area.

1. Source amplitudes: AV, AF, and AH

The amplitude of the voicing source AV is calculated from the pressure drop across the glottis and the parameter agx representing the effective glottal opening between the membranous portions of the vocal folds. When the parameter agx is in the range for which voicing can occur, the amplitude of voicing is assumed to be proportional to (**ps** − Pm)$^{3/2}$. This three-halves power relation between AV and the transglottal pressure is based in part on theoretical analysis (Stevens, 1998) and in part on experimental data (Ladefoged, 1962; Isshiki, 1964; Bouhuys *et al.*, 1968). The glottal area for a condition that we define as *modal* is specified by a speaker constant, *agm*. Its default settings are 4 mm^2 for adult males and 3 mm^2 for adult females. The modal condition is optimal in the sense that it gives rise to the maximum amplitude of voicing for a given transglottal pressure. When the actual glottal area agx is greater or less than this optimal glottal area, it is assumed that the amplitude of voicing decreases. Voicing can occur when $agx_{\min} \leqslant \text{agx} < agm + agx_{\max}$, where $agx_{\min}$ and $agx_{\max}$ are speaker constants (Table VII). When agx falls within this range, the equation governing the amplitude AV of the voicing source (in dB) is

$$AV = 20\log_{10}[(\mathbf{ps} - \text{Pm})^{3/2}] + K_v + K_{dAV} \times \frac{[\text{agx} - agm]}{100}, \tag{23}$$

where K_v and K_{dAV} are scale factors that adjust the value of AV to an appropriate range for the Klatt synthesizer,[8] and the factor of 100 in the third term converts mm^2 to cm^2. By default, K_v is 33 dB for both males and females. The value of K_{dAV} depends on agx

$$K_{dAV} = \begin{cases} K_{dAVagm-}, & agx_{\min} \leqslant \text{agx} < agm \\ -K_{dAV1agm+}, & agm \leqslant \text{agx} < agm + agx_{\text{mid}} \\ -K_{dAV2agm+}, & \\ \quad agm + agx_{\text{mid}} \leqslant \text{agx} < agm + agx_{\max}, & \end{cases} \tag{24}$$

where $K_{dAVagm-}$, $K_{dAV1agm+}$, and $K_{dAV2agm+}$ govern the slope of the AV curve in three separate regions of agx. Default values of these three speaker constants are given in Table VIII. When the pressure in the mouth, Pm, is zero, the first term in Eq. (23) is simply $30\log_{10}$**ps**. Equation (23) is illustrated in Fig. 8 for the case in which the subglottal pressure **ps** is set to its modal value *psm*, the oral pressure Pm is zero, and the relevant speaker constants are set to their default values. For agx greater than *agm*, the amplitude of voicing decreases slowly at first, but more rapidly above a certain point. By default, this midpoint, $agm + agx_{\text{mid}}$, is 9 mm^2 for females and 11 mm^2 for males. For agx greater than $agm + agx_{\max}$, the amplitude of voicing is zero; the default value of this cutoff area is 12 mm^2 for females and 15 mm^2 for males. The cutoff area is based on data reported by Hirose and Niimi (1987), from which one can infer that for male speakers, vibration will cease once the glottal area is roughly 15 mm^2. For females, the cutoff area is somewhat less. When the vocal folds are tightly pressed together, that is, $\text{agx} < agm$, the amplitude of voicing is also decreased. Below a certain point, the folds will not vibrate at all (Stevens, 1998). Thus, the amplitude of voicing is set to zero for agx below a certain value, $agx_{\min}$. The default value of $agx_{\min}$ is 1 mm^2 for both males and females.

A finite amount of transglottal pressure P_T is necessary to initiate and maintain vocal-fold vibration. Such a phonation threshold pressure (Titze, 1992) has been implemented in HLsyn. The amplitude of voicing AV is set to zero when

$$P_T = \mathbf{ps} - P_m < P_{thr\text{-}m} - K_{Pthr\text{-}dc}\mathbf{dc}, \tag{25}$$

where $P_{thr\text{-}m}$ has the default value of 3.5-cm H_2O and the scale factor $K_{Pthr\text{-}dc}$ has the default value 0.03-cm H_2O per percent. According to this equation, either decreasing subglottal pressure or increasing oral pressure can cut off voicing. Note that this cessation of voicing occurs regardless of the value of agx, although in practice the phonation threshold pressure should be somewhat sensitive to agx. In addition, variation of parameter **dc** changes the phonation threshold pressure; decreasing **dc** raises the threshold, because the increased tension makes the vocal folds more difficult to vibrate, and increasing **dc** has the opposite effect. For example, increasing **dc** to 50% decreases the phonation threshold pressure by 1.5-cm H_2O; likewise, decreasing **dc** to −50% increases the threshold pressure by that amount. Although the phonation threshold pressure is usually higher for initiating phonation than for maintaining it (Baer, 1975; Lucero, 1995), at present only one threshold has been incorporated into HLsyn.

The amplitude of aspiration (AH) also depends on the pressure in the oral cavity both directly (in terms of the pres-

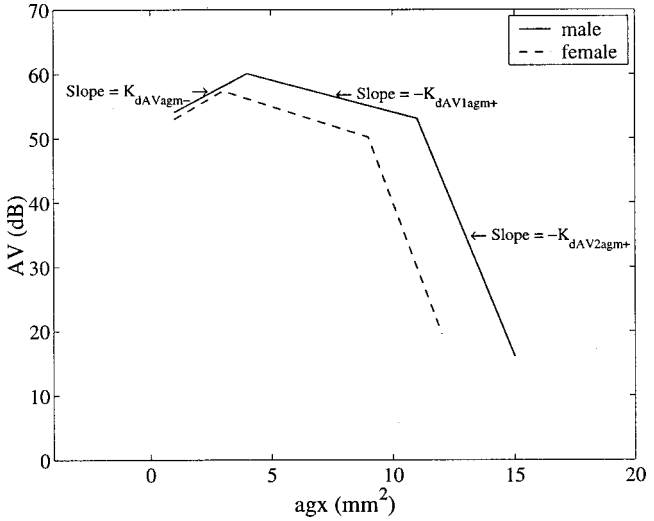


FIG. 8. Plot of Eq. (23), the dependence of the acoustic parameter AV (amplitude of voicing) on the HLsyn intermediate parameter agx (effective average area of the glottis) for the case in which the subglottal pressure **ps** is set to its modal value *psm*, the oral pressure Pm is zero, and the relevant speaker constants are set to their default values. When $\text{agx} = agm$, the modal glottal area, AV is at a maximum. If agx falls below *agm*, AV decreases. If agx increases above *agm*, AV decreases at a shallow rate while $\text{agx} \leqslant agm + agx_{\text{mid}}$, and at a rapid rate for $\text{agx} > agm + agx_{\text{mid}}$. When $\text{agx} < agx_{\min}$ or $\text{agx} > agm + agx_{\max}$, AV is zero.

sure difference across the glottis) and indirectly (in terms of the actual glottal area, which in turn depends on the oral-cavity pressure in the case of obstruent consonants)

$$\mathrm{AH} = 20\log_{10}\left[(\mathbf{ps} - P_m)^{3/2} \times \frac{\mathsf{agf}^{1/2}}{10}\right] + K_{AH}, \tag{26}$$

where the scale factor K_{AH} adjusts the level of AH to an appropriate range for the Klatt synthesizer. The default value of this speaker constant is 27 dB (Table VIII); it was chosen such that if the effective glottal area is the modal value *agm* during the production of a vowel, AH will be about 20 dB less than AV. The justification for the variation of AH with transglottal pressure and glottal area is based on experimental data and theoretical analysis of turbulence noise generation due to airflow through a narrow constriction (Shadle, 1985; Stevens, 1993).

The amplitude of frication (AF) is calculated from the oral-cavity pressure P_m, and the area acx of the oral-cavity constriction by

$$\mathrm{AF} = 20\log_{10}\left[P_m^{3/2} \times \frac{\mathsf{acx}^{1/2}}{10}\right] + K_{AF}, \tag{27}$$

where K_{AF} is a scale factor used to adjust AF so that the frication noise is in the appropriate range for the Klatt synthesizer.

2. Glottal waveform shape parameters: OQ and TL

When the glottal area agx is different from the modal area *agm*, and if conditions are appropriate for continued vocal-fold vibration, the glottal pulses undergo some change. For example, if agx is greater than *agm*, there are increases in the open quotient OQ and the tilt parameter TL. The equation for the open quotient is

$$\mathrm{OQ} = OQm + (\mathsf{agx} - agm) \times K_{OQ}, \tag{28}$$

where *OQm* is the modal open quotient for a speaker, and K_{OQ} is a scale factor. The default value of *OQm* is 50 percent for males and 65 percent for females. (For default values of K_{OQ}, see Table VIII.) The minimum and maximum values of OQ over the range of agx where phonation is possible turn out to be about 40 and 75 percent, respectively, given the default values of the various parameters (cf. Oliveira, 1997). More generally, OQ is constrained to fall between $OQ_{\min}$ and $OQ_{\max}$, speaker constants which have default values of 0 and 99 percent, respectively, for both males and females.

The modal value of the additional source spectral tilt TL is defined as *TLm*. Adjustments are made to this modal value under three circumstances: the effective glottal area agx is not equal to the modal value *agm*; the parameters acx and **an** are smaller than some critical value; or, **ap** is nonzero. Each of these factors is included in the equation for TL

$$\begin{aligned}\mathrm{TL} = {} & TLm + K_{TL}(\mathsf{agx} - agm) \\ & + K_{TL}\max(0,\ [acx_{TL} - \max(\mathsf{acx}, \mathbf{an})]) \\ & + \max(0,\ 20\log_{10}(3000 \times 2\pi T_{TL})),\end{aligned} \tag{29}$$

where K_{TL} is a scale factor, and acx_{TL} and T_{TL} are described below. The second term in the equation increases TL when the effective glottal area agx is greater than the modal area *agm*. If agx is less than *agm*, TL is reduced. The third term in the equation represents the effect of a narrow supraglottal constriction. The speaker constant acx_{TL} is the critical value below which acx or **an** affect TL; its default value is 20 mm^2. If both acx and **an** are less than acx_{TL}, TL is increased. If either acx or **an** is greater than acx_{TL}, this term is simply set to zero.

The fourth term in Eq. (29) increases TL when the HL parameter **ap** is nonzero. The presence of a glottal chink prevents the flow derivative from changing abruptly when the vocal folds close, leading to a loss of high-frequency energy (Hanson, 1995a, 1997). We model the flow in the vicinity of closure as an exponential, with a time constant of $T_{TL} = M/R$, where M and R are the total acoustic mass and resistance of the trachea, the glottis, and the vocal-tract airways. Because the resistances of the tracheal and vocal-tract airways are relatively small, the overall resistance can be approximated by the glottal resistance $R_g = R_{gv} + R_{gk}$, where R_{gv} and R_{gk} are the viscous and kinetic resistances, respectively. Consequently, $T_{TL} = M/R_g$, where

$$M = \rho\left(\frac{L_t}{A_t} + \frac{L_v}{A_v} + 100\frac{L_{vg}}{\mathbf{ap}}\right), \tag{30}$$

$$R_g = \frac{12\mu L_{vg} L_{hp}^2}{(\mathbf{ap}/100)^3} + \frac{\sqrt{2\rho \times 980|\mathbf{ps} - P_m|}}{\mathbf{ap}/100}. \tag{31}$$

The speaker constants L_t, A_t, L_v, and A_v are the length and cross-sectional areas of the trachea and the vocal tract, respectively, L_{vg} is the vertical length of the glottis, L_{hp} is the horizontal length of the posterior glottal chink, ρ is the density of air, and μ is the viscosity of air in the glottis (0.000 194 dyne-s/cm^4). The additional tilt at 3000 Hz introduced by such an exponential is $20\log_{10}[3000 \times 2\pi T_{TL}]$ (the KL parameter TL is defined as the additional decrease in source spectrum amplitude at 3000 Hz). If this equation results in a negative number, the additional tilt due to **ap** is simply set to zero. The default values for the speaker constants in Eqs. (30) and (31) are given in Table IV.

The parameter TL is constrained to fall between $TL_{\min}$ and $TL_{\max}$, speaker constants which have default values of 0 and 41 dB, respectively, for both male and female speakers.

3. Fundamental frequency, F0

There are several sources of variation in the fundamental frequency of the voice source. The primary source is the adjustment of vocal-fold tension with the intention of varying intonation. The subglottal pressure also influences $F0$, particularly at lower values of $F0$. These changes are correlates of the prosodic structure of an utterance. Adjustments to subglottal pressure can also be made on a more global level, with the intention of increasing or decreasing the overall intensity of an utterance. All else being equal, these changes in subglottal pressure will also affect $F0$. Finally, segmental properties can influence the fundamental frequency. Vowel

height is positively correlated with fundamental frequency, and vocal-fold tension may be adjusted for segmental purposes (see Sec. II).

As described in Sec. I B, the HL parameter **f0** only represents the nonsegmental variations of $F0$ due to changes in vocal-fold tension. This input contour of **f0** is then adjusted to account for changes in subglottal pressure, segmental changes in vocal-fold compliance, and the intrinsic pitch of vowels. In this section we describe the mapping relations that bring about these perturbations.

Vowel height is derived from the parameter **f1**. A speaker constant, f_{neutral}, is defined as the $F1$ border between high and low vowels. By default, f_{neutral} is 590 Hz for females and 500 Hz for males. When $\mathbf{f1} < f_{\text{neutral}}$, the tongue body is assumed to be high. The degree of vowel height depends on how far **f1** deviates from f_{neutral}. $F0$ is incremented by the amount

$$\Delta F0_{\text{height}} = \mathbf{f0} \times K_{dF0\text{-height}}(f_{\text{neutral}} - \max(\mathbf{f1}, f1_{\text{min-}F0})), \tag{32}$$

where $f1_{\text{min-}F0}$ places a lower bound on the value of the first formant frequency that can be used in this equation (250 Hz by default). The constant $K_{dF0\text{-height}}$ is a scale factor (Table VIII). As an example, if **f1** is 300 Hz, $F0$ will be increased by about 10% for males and about 13% for females, relative to the value it would have for a low vowel.

Changes in the parameter **ps** affect the KL parameter $F0$ according to the equation

$$\Delta F0_{ps} = K_{dF0\text{-}ps} \times (\mathbf{ps} - \mathrm{Pm} - psm), \tag{33}$$

where *psm* is the modal subglottal pressure and $K_{dF0\text{-}ps}$ is a speaker constant with a default value of 3 Hz/cm H_2O. The default value of *psm* is 8-cm H_2O for males and 6.5-cm H_2O for females. If the transglottal pressure, for example, is higher than the modal value of subglottal pressure, the fundamental frequency increases. Likewise, $F0$ decreases when transglottal pressure falls below modal subglottal pressure. For a vowel produced with prominence, a speaker might increase subglottal pressure by 2-cm H_2O, resulting in an increase in $F0$ of 6 Hz. Atkinson (1978) and Titze (1989) have suggested that fundamental frequency is probably more sensitive to transglottal pressure at the low end of a speaker's pitch range, but we have not yet implemented this effect in HLsyn.

An additive adjustment is also made to $F0$ based on changes in the compliance of the vocal folds, **dc**

$$\Delta F0_{dc} = -K_{dF0\text{-}dc}\mathbf{dc}, \tag{34}$$

where the scale factor $K_{dF0\text{-}dc}$ has a default value of 0.3 Hz/percent. If the compliance of the vocal folds increases, the fundamental frequency decreases. A decrease in vocal-fold compliance of 50% would increase $F0$ by 15 Hz.

The equation for $F0$, including all three factors given in Eqs. (32)–(34), is then

$$F0 = \begin{cases} \mathbf{f0} + \Delta F0_{\text{height}} + \Delta F0_{ps} + \Delta F0_{dc}, & AV > 0 \\ 0, \quad AV = 0. \end{cases} \tag{35}$$

The glottal source is controlled such that when AV becomes greater than zero, vocal-fold vibration is initiated at the beginning of a glottal pulse.

4. Degree of diplophonia, DI

Under certain circumstances, vocal-fold vibration can become irregular (see Sec. II). The Klatt synthesizer can introduce irregularities in the glottal waveform train by delaying the onset of alternating glottal pulses, as well as attenuating the amplitudes of these delayed pulses. The degree of delay and attenuation is specified by the parameter DI. The case in which irregular vibration occurs as the result of pressed vocal folds is implemented in HLsyn. That is, when

$$agx_{\text{min-}DI} < \mathrm{agx} < agm, \tag{36}$$

where $agx_{\text{min-}DI}$ is a speaker constant having a default value of 1 mm^2, the parameter DI is introduced. Its value depends on the degree to which the glottal cross-sectional area agx deviates from the modal value, *agm*

$$\mathrm{DI} = K_{DI} \times \frac{\left(1 - \dfrac{\mathrm{agx}}{agm}\right)}{\dfrac{\mathrm{agx}}{agm}} = K_{DI} \times \left(\frac{agm - \mathrm{agx}}{\mathrm{agx}}\right), \tag{37}$$

where K_{DI} is a scale factor that can be adjusted depending on the degree of diplophonia desired for a particular voice. Although irregular vibration can occur under circumstances other than when the vocal folds are pressed together, these have not yet been implemented in HLsyn.

IV. EXAMPLES

In this section we present several examples of synthesis using HLsyn.[9] These examples were chosen to complement the discussion of acoustic sources in Sec. II, illustrating how HLsyn can be used to emulate various types of acoustic sources that occur during the production of speech. The examples are summarized here.

(1) Fricative consonants, and how an appropriate balance between amplitude of voicing and frication can be achieved for voiced fricatives.
(2) Stop consonants, and the maintenance of voicing during closure of voiced stops.
(3) Production of frication during liquids and glides that occur in the vicinity of voiceless stops.
(4) Voice-source characteristics and their variation due to segmental environment.
(5) Fundamental frequency, and the effects of vowel height, adjacent consonants, and subglottal-pressure variations.

A. Fricative consonants

We extend the example given in Sec. III D by synthesizing the words ''lacy'' and ''lazy,'' which contain a voiceless

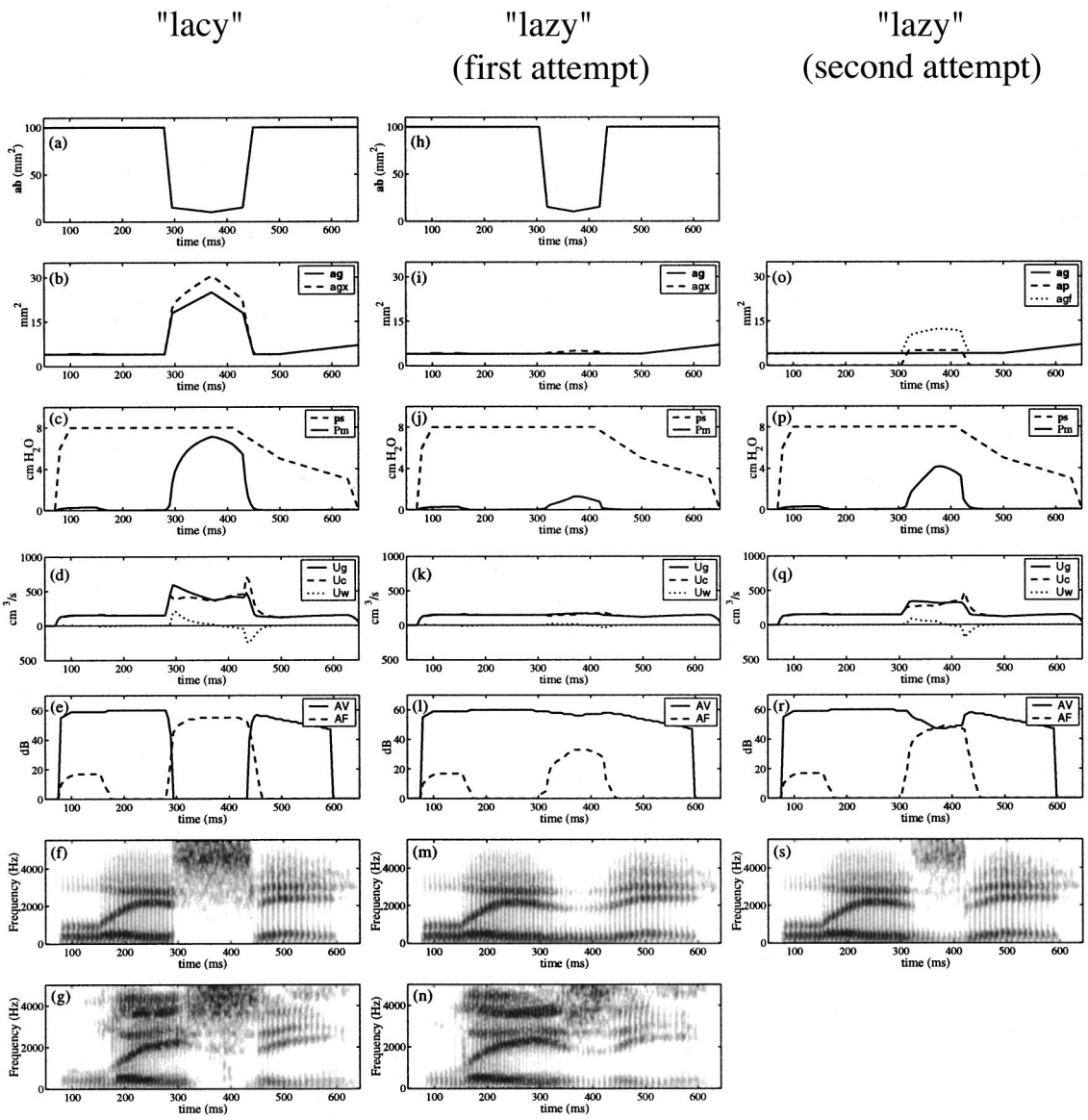


FIG. 9. Examples of the synthesis of fricatives. Column 1: Panels (a)–(g) illustrate synthesis of a voiceless fricative, /s/. Column 2: Panels (h)–(n) illustrate synthesis of a voiced fricative, /z/. Column 3: A second attempt to synthesize the word ''lazy,'' illustrating how inclusion of the parameter **ap** (area of a posterior glottal opening) improves synthesis of a voiced fricative. Row 1: Parameter **ab**, emulating the formation and release of a constriction by the tongue blade. Row 2: Parameter **ag**, the average area at the membranous portion of the vocal folds; intermediate parameter agx, the effective glottal area at the membranous folds; **ap**, representing the area of a posterior glottal opening; and intermediate parameter agf, the total glottal area. Row 3: Oral pressure Pm and subglottal pressure **ps**. Row 4: Volume velocities Ug through the glottis, Uc through the supraglottal constriction, and Uw representing the expansion of the vocal-tract walls due to the increase in oral pressure. Row 5: Acoustic parameters AV (amplitude of voicing) and AF (amplitude of frication). Row 6: Spectrograms of the synthesized utterances. Row 7: Spectrograms of natural renditions of ''lacy'' and ''lazy'' (recorded by the second author). (See Sec. IV A for details.)

and a voiced fricative, respectively. The first column in Fig. 9 shows the results of an attempt to synthesize the word lacy. In Figs. 9(a) and (b), parameters **ab** and **ag** are varied in a way similar to that in Figs. 7(a) and (b). It is noted that the minimum value of **ab** is about one-half of the maximum value of agx, thereby maximizing the amplitude of frication noise, as discussed in Sec. II. The tracks for the intermediate parameters agx, Pm, Ug, Uc, and Uw are given in Figs. 9(b)–(d), and are similar to those in Figs. 7(b)–(d). Below these, in Fig. 9(e), are the tracks for the Klatt acoustic parameters AV and AF, the amplitudes of voicing and frication, respectively. As desired, parameter AV drops to zero during the /s/, and AF rises to 55 dB, nearly equal to the value of AV during the sonorant portions of the utterance. A spectrogram of the utterance is included in Fig. 9(f), while Fig. 9(g) is a spectrogram of the word lacy spoken by the second author. Comparison of these two spectrograms shows that in the synthesized version, the amplitude of frication during /s/ relative to the amplitude of voicing during the sonorants is appropriate.

The second column in Fig. 9 shows the results of a first attempt to synthesize the word lazy. For the segment /z/, parameter **ab** is varied as it was for lacy, except the duration of the constriction is shorter [compare to (a)]. Figures 9(i)–(k) correspond to Figs. 7(e)–(g); as in the earlier example, the intermediate parameters agx, Pm, Ug, Uc, and Uw dur-

ing the fricative show only minor changes. Consequently, as shown in Fig. 9(l), the acoustic parameter AV falls by only 4 dB, and parameter AF rises to only 33 dB, that is, 23 dB less than AV. Spectrograms of the synthesized and natural renditions of lazy are included in Figs. 9(m) and (n), respectively. In the synthesized utterance, the amplitude of voicing in the fricative is too strong, while the amplitude of frication is too weak. Informal listening verifies the impression that the source amplitudes are out of balance during the fricative; the synthesized word sounds more like ['lei ði] than ['lei zi].

Finally, in the third column of Fig. 9 we show that inclusion of the parameter **ap** results in more appropriate acoustics for the segment /z/. Parameter **ab** is varied as it was for the first attempt to synthesize /z/ [part (h)] and is not repeated. Figures 9(o)–(q) correspond to Figs. 7(h)–(j). As in that example, although the glottal area **ag** is held constant throughout the constriction, an increase in the area between the arytenoid cartilages means that Pm will increase to a value more appropriate for a voiced fricative; consequently, the total glottal area agf increases, relative to the previous attempt [compare agf in (o) to agx in (i)], and the flow through the supraglottal constriction, Uc, also increases. These changes result in a reduction in the acoustic parameter AV, and an increase in parameter AF [Fig. 9(r)]. Note that AF is less than it was for the voiceless fricative in column 1 because Pm has to be low enough to allow voicing to continue. The resulting spectrogram in Fig. 9(s) shows that inclusion of the parameter **ap** for the synthesis of a voiced fricative is necessary to obtain appropriate acoustic output [compare to the natural utterance in (n)]. Informal listening verifies that this approach produces better-quality voiced fricatives.

B. Stop consonants

Production of stop consonants in English (and in many other languages) requires careful balance and timing of several acoustic sources: voicing, frication, and aspiration. For a voiced stop consonant, the vocal folds should continue to vibrate for some time after the formation of the supraglottal constriction, while for a voiceless stop, vibration should cease immediately. At the release of the constriction in both voiced and voiceless stops, a burst of frication noise should be produced. Following the release, voicing should begin rather quickly for voiced stops, while voice onset should be delayed for voiceless aspirated stops; during this delay of about 40–60 ms, aspiration noise should be produced. Although the timing of the acoustic sources is somewhat complex, we show that such timing can be generated with relatively simple control of HLsyn parameters representing glottal and supraglottal constrictions.

In the first column of Fig. 10 [panels (a)–(h)], we illustrate the production of an unvoiced aspirated stop consonant using HLsyn. The utterance ''a Kaiser'' has been synthesized using parameters appropriate for a female speaker. Certain of the speaker constants were set to have nondefault values: the model values of the first two formant bandwidths $B1_m$ and $B2_m$ were set to 120 Hz (Hanson, 1997). In addition, because the presence of a posterior glottal opening is believed to be normal for female speakers, the HL parameter **ap** was set to 1 mm^2 throughout most of the utterance, increasing only for the voiced fricative /z/.

Variations of the glottal constriction and the dorsal supraglottal constriction are illustrated in Figs. 10(a) and (b). Because /k/ is formed by raising the tongue body, the parameter **f1** is reduced to 180 Hz between 120–215 ms to create the supraglottal constriction. The intermediate parameter acd, the area of the constriction at the dorsum, is derived from **f1**, and is 0 mm^2 when **f1**=180 Hz. The HL parameter **ag** begins to increase from its modal value of 3 mm^2 just before closure. It peaks at 25 mm^2 just after the release of the constriction, and falls back to its modal value over about 80 ms. The intermediate parameter agx is derived from **ag**, and is slightly greater than **ag** when the intraoral pressure increases.

Figure 10(c) illustrates parameter **dc**, which is used to gain more control over the onset of voicing. Recall that if transglottal pressure is above phonation threshold pressure, vocal-fold vibration will occur when the intermediate parameter agx is below a certain threshold; for females, the default is 12 mm^2. However, it may not be desirable to have voicing start as soon as agx reaches 12 mm^2; Ní Chasaide and Gobl (1993) examined transitions from labial stop consonants to vowels in five languages, and found that only in German did the vocal folds seem to be significantly more abducted at the onset of voicing in vowels following voiceless consonants. That is, in the other four languages (including English), the folds appeared to be quite well adducted before they began to vibrate, even when the preceding consonant was not voiced and the vocal folds had, presumably, been spread. One way to achieve this effect is to increase the phonation threshold pressure by decreasing the compliance of the vocal folds, as shown earlier in Eq. (25). Therefore, for our example, at the formation of the supraglottal constriction, parameter **dc** is decreased, reaching a minimum of −75% shortly after formation of the constriction. It then rises, reaching zero after 150 ms. The result of this variation of **dc** will be discussed shortly.

In Fig. 10(d) we show the HL parameter **ps**, along with the intermediate parameter Pm. When the closure is formed at the dorsum, the intraoral pressure Pm rises almost immediately to be equal to the subglottal pressure and then falls just as quickly back to zero when the constriction is released. The airflow through the glottis and the supraglottal constriction are illustrated in Fig. 10(e). Both flows increase slightly as the vocal folds spread, but Uc quickly drops to zero at the formation of the closure. The glottal flow Ug also drops off, reaching zero when the intraoral pressure approaches **ps**. When the closure is released, air can move through the vocal tract and both Uc and Ug increase. The vocal folds close following the release, and therefore both Ug and Uc decrease.

The resulting acoustic parameters are shown in Fig. 10(f). The amplitude of voicing AV is quickly cut off as the supraglottal constriction is formed and intraoral pressure Pm builds up. At release of the constriction, frication noise is generated at the velar constriction. Its amplitude AF rapidly falls off when Pm drops to atmospheric. Although transglot-

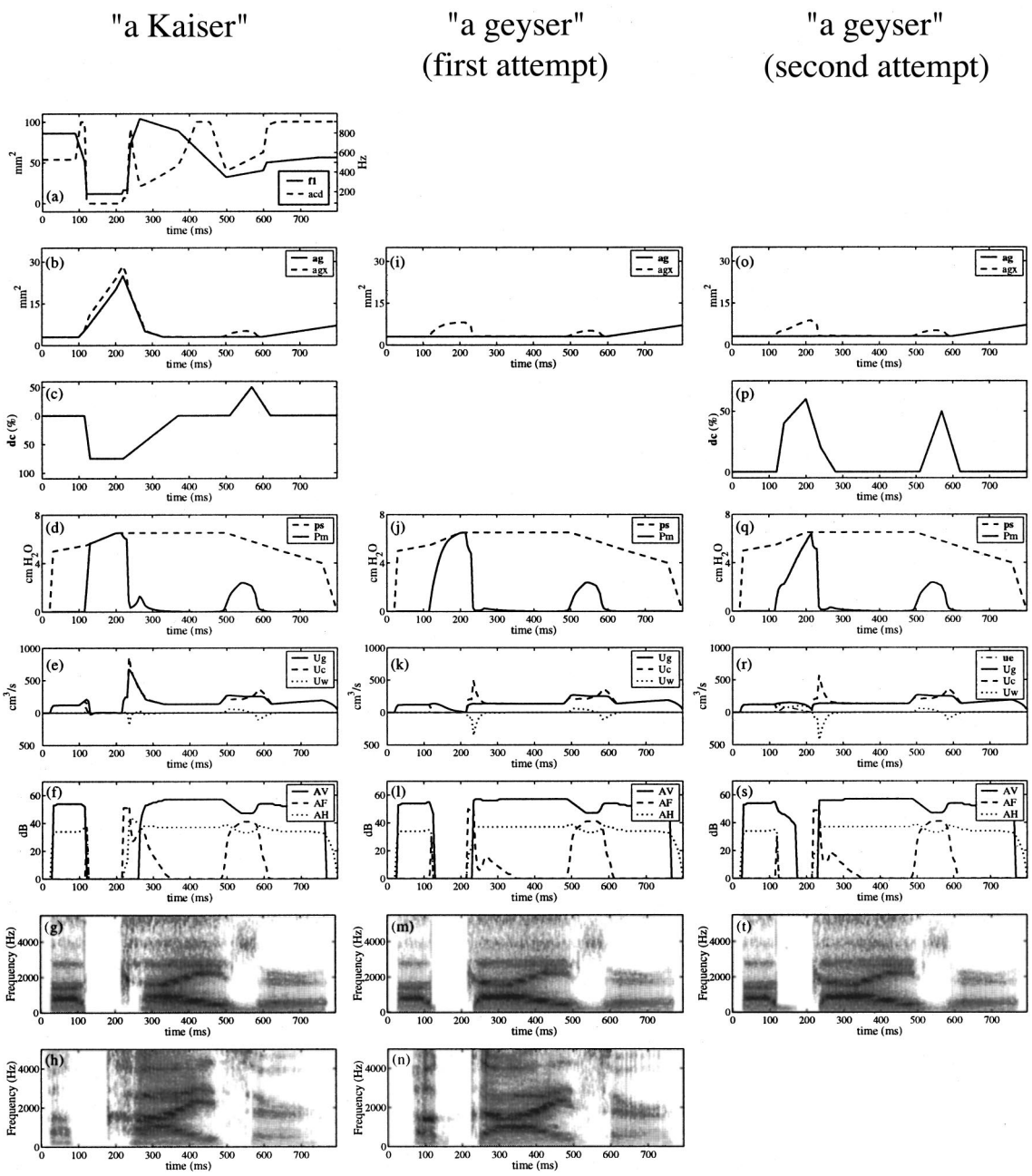


FIG. 10. Examples of the synthesis of stop consonants. Column 1: Panels (a)–(h) illustrate synthesis of /k/. Column 2: A first attempt to synthesize the phrase ''a geyser,'' illustrating synthesis of a voiced stop. Column 3: A second attempt to synthesize the utterance ''a geyser,'' illustrating how inclusion of the parameters **ue** and **dc** improves synthesis of a voiced stop. Row 1: Parameter **f1**, which is reduced to 180 Hz between 120–215 ms, to indicate that a supraglottal closure has been formed by the tongue body, and the intermediate parameter acd is reduced to 0 mm^2. Row 2: Parameter **ag**, the average area at the membranous portion of the vocal folds, and intermediate parameter agx, the effective glottal area at the membranous folds. Row 3: Parameter **dc**, change in compliances of the vocal folds and walls. Row 4: Oral pressure Pm and subglottal pressure **ps**. Row 5: Volume velocities Ug through the glottis, Uc through the supraglottal constriction, Uw representing the expansion of the vocal-tract walls due to the increase in oral pressure, and **ue** representing active expansion of the vocal-tract volume. Row 6: Acoustic parameters AV (amplitude of voicing), AF (amplitude of frication), and AH (amplitude of aspiration). Row 7: Spectrograms of the synthesized utterances. Row 8: Spectrograms of natural renditions of ''a Kaiser'' and ''a geyser'' (produced by the first author). (See Sec. IV B for details.)

tal pressure is nearly equal to subglottal pressure at this point, the vocal folds do not begin to vibrate for two reasons: first, the folds do not adduct as quickly as the oral pressure falls, and second, the phonation threshold pressure is raised due to the reduction in the compliance of the folds [recall panel (c)]. Until both the intermediate parameter agx is less than 12 mm^2 and the vocal-fold compliance is increased sufficiently, the folds do not vibrate. The parameter agx reaches 12 mm^2 at about 264 ms (40 ms after release), but the phonation threshold pressure is too high for voicing to begin at that time. It is not until about 280 ms, when agx ≈ 5.5 mm^2, that voicing can begin. In the meantime, the flow of air past the adducting folds results in the generation of aspiration noise (see parameter AH in the panel). Finally, Fig. 10(g) shows a spectrogram of the synthesized speech; it can be compared with a spectrogram of a natural rendition of the utterance in Fig. 10(h) (produced by the first author).

The time course of the parameter **f1** has a small hesita-

tion in it at the release of the stop consonant: **f1** increases rapidly for about 5 ms, flattens out for about 10 ms, and then rises rapidly again. This plateau in **f1** is reflected in the area of constriction acd, and represents vocal-tract wall effects on the articulator trajectory at release. These effects are the result of the buildup of oral pressure displacing the soft tissues of the vocal tract during the stop closure; at release, the tissues spring back to their normal positions, altering the time course of the release relative to what it would be if there were no pressure buildup or if the walls were rigid (Stevens, 1998, 2001). Inclusion of these wall effects on the articulator trajectory at consonant release is necessary to get stop bursts with appropriate amplitudes and durations (Hanson and Stevens, 2000).

An initial approximation to the synthesis of a voiced stop is illustrated in the second column of Fig. 10. The utterance ''a geyser'' was synthesized by changing only two of the HLsyn parameters in the previous example for an unvoiced stop: first, the parameter **ag** is not increased during the stop closure, but rather is held constant at its modal value of 3 mm^2 [panel (i)], and second, the parameter **dc** is not varied during the stop consonant. The parameter **f1** (and therefore the intermediate parameter acd) varies exactly as it did for the previous example, and is not repeated.

In Fig. 10(j) are illustrated the subglottal pressure **ps** and the oral pressure Pm. In comparison with the voiceless stop, the differences are that the oral pressure does not rise as rapidly following formation of the supraglottal constriction, and it falls somewhat more quickly following the release. Figure 10(k) shows the airflow through the glottis (Ug) and the supraglottal constriction (Uc). Compared with Fig. 10(e), the main differences are that flow through the glottis continues well into the stop closure, and following the release of the velar constriction, Uc and Ug do not increase as much as they did for the voiceless stop.

The resulting acoustic-source parameters are shown in Fig. 10(l). The most obvious differences with the voiceless stop occur at the release. First, because Pm falls off more rapidly, the amplitude of frication (AF) is about 5 dB less for the voiced stop. Second, because the vocal folds are not abducted at the release, no significant aspiration noise (AH) is generated, and the amplitude of voicing (AV) becomes nonzero within about 10 ms. Another, more minor, difference occurs at the closure: because the oral pressure rises somewhat more slowly than it did for the voiceless stop, voicing continues for about 10 ms into the closure.

Figure 10(m) is a spectrogram of the synthesized rendition of ''a geyser.'' Comparing it to the spectrogram in Fig. 10(g), one can see the short voice bar at the beginning of the stop closure, reduced amplitude of frication at the release, and the short VOT, typical of a voiced stop in American English. A spectrogram of a natural rendition of the utterance is shown in Fig. 10(n). Unlike the synthesized version, this natural utterance has a significant voice bar at the beginning of the stop closure.

In order to produce significant voicing during the stop closure, we have made a second attempt to synthesize a voiced stop consonant (third column of Fig. 10) by introducing two changes. First, the compliance of the vocal folds is increased at the stop closure by increasing the HLsyn parameter **dc** [panel (p)], thereby lowering the phonation threshold pressure. Second, the parameter **ue** is also increased at the stop closure, reaching a maximum of about 75 cm^3/s and then falling to zero at the release [panel (r)]. Recall that **ue** accounts for active expansion of the volume of the vocal tract. Comparing the flows illustrated in Fig. 10(r) with those in panel (k), we see that Ug continues throughout the stop closure interval. As shown in Fig. 10(q), the oral pressure Pm rises much more slowly than before.

The acoustic-source parameters are shown in Fig. 10(s), where it can be seen that the amplitude of voicing is not cut off until shortly before the release. The voice bar is clearly seen in the spectrogram in Fig. 10(t).

In different languages and other English dialects, implementation of the voicing contrast might require control parameters different from those shown in Fig. 10 for American English. For example, in French the voiceless stop is not aspirated, so that a different timing of the parameter **ag** is used, but the use of **dc** probably remains about the same. The voiced stop is more heavily voiced than in English, a result that could be achieved with greater involvement of the parameter **ue** in Fig. 10(r).

C. Frication during liquid consonants in the vicinity of voiceless stops

When a voiceless aspirated stop consonant is followed by a high vowel or a sonorant consonant, the frication noise generated at the stop release is followed not only by aspiration noise but by a second phase of frication noise generated near the constriction for the sonorant. This second phase of frication occurs because the combination of a narrow supraglottal constriction with spread vocal folds prevents oral pressure from falling as rapidly as it normally would following a stop release. The resulting pressure drop across the oral constriction generates turbulence noise which primarily excites the cavity in front of the constriction. Thus, the vocal tract during this time is excited by a mix of frication and aspiration noise. This type of noise occurs in initial consonant sequences like ''twin,'' ''clay,'' and ''pure.''

In this section we describe synthesis of the words ''pray'' and ''bray.'' Panels in the first column of Fig. 11 illustrate synthesis of ''pray.'' As discussed in Sec. III A 3, a liquid consonant is indicated in HLsyn when the first three formants fall into certain ranges. In Fig. 11(a) it can be seen that until about 200 ms, parameters **f1**–**f3** fall into ranges for a liquid consonant. The area of the corresponding constriction acl is derived from the HL parameter **f1** using Eq. (11), which gives a minimum value of acl of 25 mm^2, assuming default values of the speaker constants. In the early part of this utterance, there is also a constriction formed at the lips; panel (b) shows that this constriction, **al**, is zero until 90 ms, when the constriction is released. The supraglottal constriction acx is then determined according to Eq. (12) to be

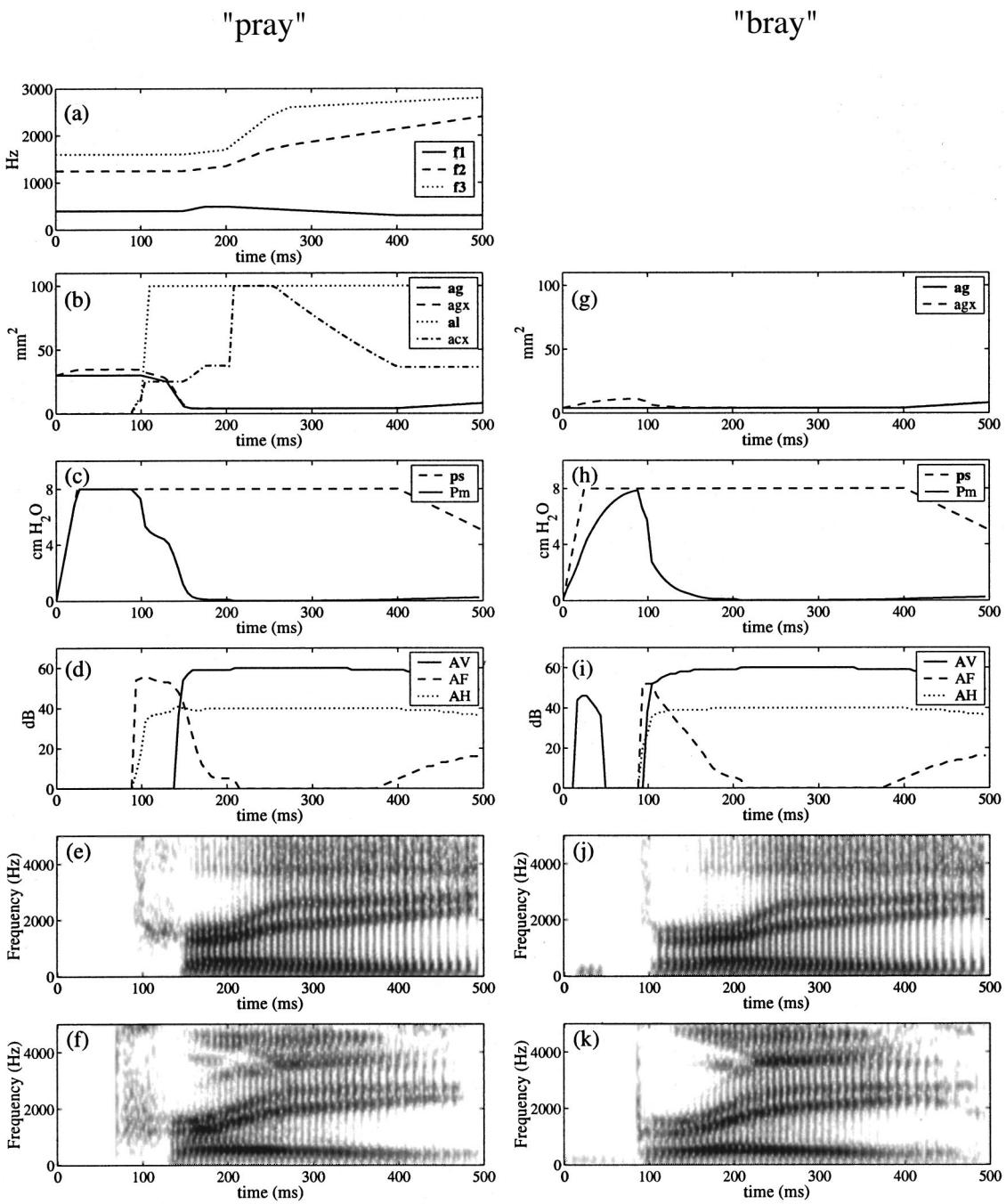


FIG. 11. Examples of the synthesis of liquid consonants in the vicinity of stop consonants. Synthesis parameters are for the words ''pray'' (column 1) and ''bray'' (column 2). Row 1: HL parameters representing the first three natural frequencies of the vocal tract, **f1**, **f2**, and **f3**. Row 2: Cross-sectional area at the lips, **al**; intermediate parameter acx, based on the narrowest supraglottal constriction; cross-sectional area at the membranous portion of the folds, **ag**; and effective area of the glottis agx. Row 3: Oral pressure Pm and subglottal pressure **ps**. Row 4: Acoustic parameters AV (amplitude of voicing), AF (amplitude of frication), and AH (amplitude of aspiration). Row 5: Spectrograms of the synthesized utterances. Row 6: Spectrograms of natural renditions of ''pray'' and ''bray'' (produced by the second author). (See Sec. IV C for details.)

$$\text{acx} = \begin{cases} \textbf{al}, & 0 \leqslant t \leqslant 100 \\ \text{acl [Eq. (11)]}, & 100 < t \leqslant 200 \\ \text{acd [Eq. (8)]}, & 200 < t, \end{cases} \tag{38}$$

where t is the time in milliseconds. The resulting acx contour is given in Fig. 11(b). Also shown in panel (b) is the HL parameter **ag**, which is varied to model adduction that begins at the stop-consonant release and finishes about 50 ms later, and the intermediate parameter agx. Not shown is parameter **dc**, which has a contour similar to the voiceless stop consonant described in Sec. IV B and Fig. 10.

Figure 11(c) illustrates the oral pressure Pm that results from the parameters shown in panels (a) and (b). During the stop closure it is equal to the subglottal pressure **ps**. Following the release it begins to fall rapidly, but then levels off at 4-cm H_2O because acx≅**ag**. Once **ag** falls below acx, at about 135 ms, Pm again falls rapidly.

Figure 11(d) shows the resulting KL source-amplitude

contours. Unlike the voiceless aspirated stop illustrated in Fig. 10, the parameter AF rises rapidly following the stop release and remains close to this level until voicing begins; the parameter AF is greater than AH, and is the dominant noise parameter. Panel (e) shows the spectrogram of the synthesized utterance. The frication during the burst is broadband, but during the remainder of the VOT only the third formant is excited by frication. The higher formants are only weakly excited by aspiration noise. Panel (f) is a spectrogram of the word ''pray'' produced by the second author.

The word ''bray'' (second column of Fig. 11) was synthesized by simply changing two HLsyn parameters: **ag** was set to 4 mm^2 throughout the stop consonant [panel (g)] and **dc** was set to vary in a manner appropriate for a voiced stop (see Sec. IV B). As a result, we see in Fig. 11(h) that at the stop release oral pressure Pm falls off much more rapidly than it does for pray. Consequently, in Fig. 11(i) voicing starts shortly after the release. Figure 11(j) is a spectrogram of the synthesized utterance, while Fig. 11(d) is a spectrogram of a natural rendition produced by the second author.

D. Voice-source shape characteristics

As described in Sec. III E 2, the KL parameters that affect the shape of the glottal waveform, OQ and TL, may be affected in certain segmental environments. Both parameters increase during voiced segments for which (1) the membranous or cartilaginous folds are spread or (2) a narrow supraglottal constriction exists. In this section we show how OQ and TL vary during the utterance ''a Kaiser,'' the synthesis of which was described in Sec. IV B and Fig. 10. Recall that for that example the HL parameter **ap** had a minimum value of 1 mm^2. Although not discussed in that example, the parameter **ap** was increased to 5 mm^2 during the segment /z/ for reasons discussed in Sec. IV A in connection with the synthesis of lazy. In addition, parameter **ag** was increased from its modal value to about 7 mm^2 during the final syllable [Fig. 10(b)] to emulate spreading of the folds which may occur utterance finally (Slifka, 2000, described in Sec. II).

In panel (a) of Fig. 12 we include the spectrogram of the synthesized version of a Kaiser for reference, while panels (b) and (c) illustrate the variation of OQ and TL, respectively. These two parameters are calculated at all times during an utterance, although, of course, they have an influence on the waveform only during parts of the utterance when there is glottal vibration. The parameter OQ increases near and during the segments /k/ and /z/ because the glottal opening at the membranous portion of the vocal folds, agx, is larger than the modal value of the glottal opening, *agm*. During /k/ the vocal folds are not vibrating, but because they begin to spread before they cease to vibrate at the consonant closure (near 120 ms), and then begin to vibrate before they are fully adducted following the consonant release (270 to 330 ms), OQ is increased relative to its modal value during the neighboring vowels. It also increases through the constricted interval for /z/ due to the effect of the raised oral pressure. The parameter TL is increased relative to its modal value throughout the utterance, but especially so near and during the two obstruent segments. During the vowel segments it is about 5 dB above its modal value because the HL

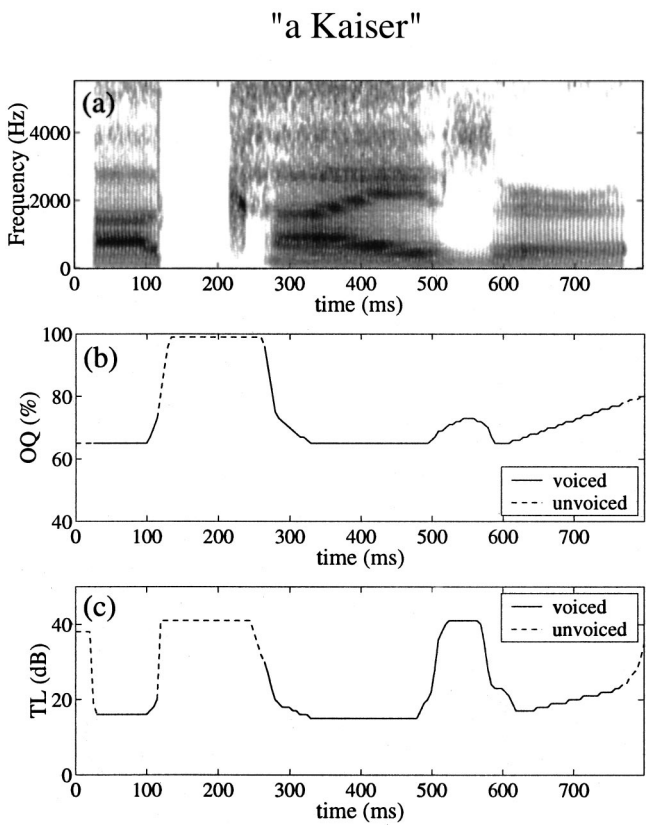


FIG. 12. An example of the effect of segmental environment on the glottal-waveform shape parameters, OQ and TL. (a) Spectrogram of the synthesized utterance ''a Kaiser,'' described earlier in Sec. IV B and Fig. 10. (b) Klatt parameter OQ (open quotient). (c) Klatt parameter TL (additional spectral tilt at 3000 Hz). (See Sec. IV D for details.)

parameter **ap** has a minimum value of 1 mm^2. It increases even more close to and throughout the segment /k/ largely because the vocal folds are spread. Increases in TL near and during the segment /z/ are primarily due to the narrow supraglottal constriction and the increase in **ap** to 5 mm^2, but there is also a small increase due to the passive spreading of the vocal folds. Finally, both OQ and TL increase during the final segment, due to the gradual increase in parameter **ag**.

E. Perturbations to fundamental frequency

In our final example we show how adjustments are made to the input parameter **f0** to account for changes in transglottal pressure, variation of vocal-fold compliance, and intrinsic pitch of vowels. Figure 13(a) shows a spectrogram of a synthesized rendition of the utterance ''He sees the ball.'' The dashed line in Fig. 13(b) illustrates the parameter **f0**, which represents variations in the intonation contour derived from some prosodic representation. It is constant at 100 Hz, except during the vowel /i/ in ''sees,'' which is meant to carry a high pitch accent. Overlayed on **f0** with a solid line is the KL parameter $F0$, that is, the intonation contour that is actually produced. $F0$ is greater than **f0** during ''he'' and ''sees'' because high vowels have increased intrinsic pitch. Subglottal pressure [Fig. 13(c)] drops somewhat during the reduced vowel in ''the,'' resulting in a slight drop in $F0$ relative to **f0**. During the closure of the segment /b/, there are two contributions to the decrease in $F0$: first, transglottal pressure falls

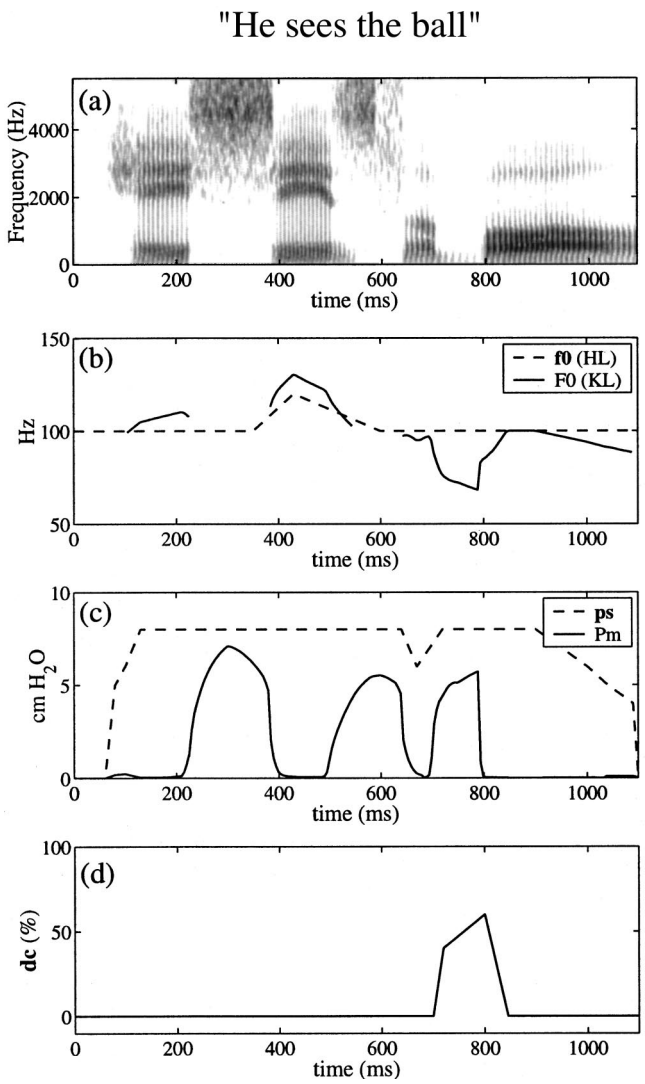


FIG. 13. An example of how the HL parameter **f0** is adjusted to account for variations in transglottal pressure, changes in vocal-fold compliance, and vowel intrinsic pitch. (a) Spectrogram of the synthesized utterance "He sees the ball." (b) The HL parameter **f0** is constant for most of the utterance, except for a prominence on the word "sees." The acoustic parameter $F0$ shows variations more typical of a natural utterance. (c) Subglottal pressure **ps** and oral pressure Pm. (d) Changes in the compliances of the vocal folds and the vocal-tract walls. (See Sec. IV E for details.)

due to the increase in oral pressure, and vocal-fold compliance [Fig. 13(d)] increases to prolong vocal-fold vibration. Finally, subglottal pressure begins to fall about halfway through the final vowel, resulting in a decrease in $F0$ relative to **f0**. The $F0$ contour looks more natural than the input **f0** parameter track. We emphasize that these perturbations are due to mappings that are part of HLsyn, and therefore the intonation rules in a text-to-speech system based on HLsyn need only be concerned with changes in pitch specified by the **f0** parameter.

V. SUMMARY: SOURCE SYNTHESIS WITH HLsyn

We summarize first the parameters that control the glottal source during times when there is essentially no pressure buildup above the glottis. During the sonorant intervals, particularly the vowels, manipulation of the glottal source has two functions in English. One is to mark the beginnings and endings of phrases, and the other is to create particular degrees of prominence to individual syllables in a phrase. The amplitude of the glottal source during these time intervals in the proposed synthesis model is determined primarily by two articulatory parameters: the subglottal pressure **ps** and the degree of abduction or adduction of the vocal folds, specified by **ag**. These parameters are manipulated by a speaker at the beginning and end of a phrase and also create prominences of syllables within a phrase or produce syllables with reduced prominence. Glottal abduction or adduction results not only in a change in source amplitude but also a modification of the waveform shape of the glottal source. The waveform modification due to glottal abduction tends to weaken the high frequencies in the source spectrum, whereas greater adduction leads to enhanced high-frequency spectrum energy. The former often occurs in vowels near the end of a phrase and in reduced vowels, whereas the latter often occurs in vowels that are produced with greater prominence. There may also be influences on glottal abduction/adduction in portions of vowels that are adjacent to voiced and voiceless obstruent consonants. The fundamental frequency of glottal vibration also plays a role in marking phrase boundaries and prominences. It is determined by several physiological parameters: (1) a basic **f0** parameter that is the result of an intentional change in vocal-fold stiffness for enhancing or reducing the prominence of a syllable or for marking a phrase boundary; (2) the subglottal pressure; (3) intrinsic effects of vowel height; and (4) the voicing feature for an adjacent obstruent consonant.

During the obstruent intervals when there is a pressure increase above the glottis, acoustic sources can be generated both at the glottis and in the vicinity of a supraglottal constriction, depending on whether the consonant is voiced or voiceless. In English, the goal in producing a voiceless consonant is to avoid overlap between the glottal vibration source and the frication noise source near the supraglottal constriction. For a voiced consonant, on the other hand, such an overlap in sources is the goal, and there should be an appropriate balance in intensity between the two kinds of sources. During these time intervals when there is a narrowing of the airway above the glottis, the acoustic sources are controlled by parameters that adjust the intraoral pressure and hence the pressure across the glottis, as well as parameters that change the stiffness of the vocal folds. The intraoral pressure is achieved by creating an appropriate balance between the glottal opening (parameters **ag** and **ap**) and the supraglottal constriction (the minimum of several constriction parameters including **al**, **ab**, and ranges of formant frequencies that lead to a narrow supraglottal constriction, particularly for velars, palatals, and liquids). The parameters **dc** and **ue**, which influence movements of the vocal-tract walls, also play a role in setting the intraoral pressure, as does the subglottal pressure **ps**. Whether or not the vocal folds vibrate during the consonantal interval is a function of the transglottal pressure, the glottal configuration (given by **ag**), and the incremental vocal-fold stiffness (given by **dc**). The transglottal pressure is determined by the subglottal pressure **ps**, the glottal opening **ag**+**ap**, the area of the supraglottal constriction, and the parameters **dc** and **ue** that

control the movements of the vocal-tract walls. The acoustic effect of some of these parameters often spreads beyond the obstruent interval into an adjacent vowel.

VI. TOWARD RULES FOR SYNTHESIS

This review of source parameters and their origin in the underlying articulatory control of speech can provide some insight into the links between the three levels of the speech production process: the acoustic patterns of speech, the articulatory, laryngeal, and respiratory gestures that give rise to these patterns, and the discrete phonological representation in terms of which utterances are planned. The variability in the acoustic pattern that arises from a given phonological representation in the speech planning stage is well known, and provides a formidable obstacle to the design of the text-to-speech systems that produce natural and intelligible speech. This variability is also a major impediment to the development of systems for automatic speech recognition. Viewed in terms of the HLsyn-based articulatory pattern interposed between the planning stage and the sound, however, the variability problem becomes much less severe. The present discussion of source parameters provides some support for this view.

In the production of running speech, the time-varying gestures represented by the HL source parameters during sonorant and obstruent intervals often overlap across segments (Browman and Goldstein, 1990), and there may be an interaction between requirements of the segments and the requirements of the prosodic units. The overlap or interaction is often difficult to specify when sources are described in purely acoustic terms, but can be described more naturally in terms of HL parameters. For example, in a sequence like ''his farm,'' the voiced–voiceless obstruent sequence often shows weakened glottal vibration during much of the time interval when the tongue blade is in an alveolar position for /z/. This coarticulation occurs naturally if one takes into account the influence of the abducting glottal gesture for /f/ on the preceding adducting glottal gesture for /z/. Or, it often happens that a phrase-final voiced obstruent consonant shows minimal glottal vibration during the obstruent interval. This is a naturally occurring consequence of the phrase-final reduction in subglottal pressure, leading to a transglottal pressure that can fall below the threshold of phonation. These and many other examples illustrate how physiologically oriented HL parameters for synthesis of sources in speech can be derived from the linguistic planning stage. Partial overlap or interaction between these parameters is very common in running speech and can be implemented with relatively simple rules. The acoustic consequences of this gestural overlap can be highly variable, so that a synthesis method based solely on acoustic parameters or stored acoustic patterns is quite complex.

ACKNOWLEDGMENTS

Development of HLsyn was supported by NIH Grants No. NS-27407-01 and No. MH52358. Preparation of the manuscript was supported in part by NIH Grant No. DC04331. Programming support was provided by Corine A. Bickley, David R. Williams, Eric Carlson, and Robert E. Beaudoin. Special thanks are due to Robert E. Beaudoin for development of the scheme to solve the circuit model for the pressures and flows. Shrikanth Narayanan generously found time to round up and share with us vocal-tract area function data presented in Fig. 6. We gratefully acknowledge the comments and suggestions of Yi Xu and three other reviewers on an earlier version of the manuscript.

APPENDIX: TABLES OF SPEAKER CONSTANTS

The tables in this Appendix contain the speaker constants which can be set by users of the HLsyn system. We note that the units given in Tables IV–VIII are as they must be specified in the system. In the interest of clarity, the units used in the paper sometimes vary from those in the tables. For example, the parameter $K_{dF0\text{-}dc}$ is specified in Table VIII in dHz/percent, but in the paper we use the units Hz/percent.

TABLE IV. Speaker constants: Physical characteristics.

Constant	Description	Female	Male
L_v	Length of vocal cavity	15 cm	17 cm
A_v	Cross-sectional area of oral vocal cavity, neutral setting	3 cm^2	3.5 cm^2
L_t	Length of trachea	11 cm	12 cm
A_t	Cross-sectional area of trachea	2 cm^2	2.5 cm^2
L_g	Effective horizontal length of glottis	0.7 cm	1 cm
L_{hp}	Horizontal length of posterior glottal opening	0.2 cm	0.3 cm
L_{vg}	Vertical length of glottis	0.3 cm	0.4 cm
$L_{c\text{-}acd}$	Length of velar constriction	3.5 cm	4 cm
V_{acd}	Volume of cavity behind velar constriction	40 cm^3	50 cm^3
C_{wm}	Modal compliance of vocal-tract walls	0.001 cm^5/dyne	0.001 cm^5/dyne
R_w	Resistance of vocal-tract walls	10 dyne-s/cm^5	10 dyne-s/cm^5
C_{gm}	Modal compliance of the vocal folds	5×10^{-6} cm^3/dyne	5×10^{-6} cm^3/dyne

TABLE V. Speaker constants: Acoustic characteristics.

Constant	Description	Female (Hz)	Male (Hz)
f_{neutral}	$F1$ border between high and low vowels	590	500
$f1_{\text{min}}$	Lowest Helmholtz frequency of vocal tract	180	180
$F1_{\text{max}}$	Upper threshold on first resonance of vocal tract	1100	900
$f1_{\text{liquid}}$	First resonance for liquid consonants	450	400
$f1_{\text{phar}}$	Lowest first-formant resonance when constriction is in pharyngeal cavity	650	540
$B1_m$	Modal bandwidth of first formant	80	80
$B2_m$	Modal bandwidth of second formant	90	90

TABLE VI. Speaker constants: Source characteristics.

Constant	Description	Female	Male
agm	Modal value of **ag**	3 mm^2	4 mm^2
TLm	Modal source spectral tilt	10 dB	5 dB
OQm	Modal open quotient	65%	50%
psm	Modal subglottal pressure	6.5-cm H_2O	8-cm H_2O

TABLE VII. Speaker constants: Thresholds and breakpoints.

Constant	Description	Female	Male
acx_{TL}	Value of acx above which the supraglottal constriction does not affect TL	20 mm^2	20 mm^2
AF_{min}	Threshold below which frication does not occur	35 dB	35 dB
$agx_{\text{min-}DI}$	Minimum value of agx for which parameter DI>0	1 mm^2	1 mm^2
agx_{max}	Maximum value of agx for which voicing occurs (AV>0) is $agm+agx_{\text{max}}$	9 mm^2	11 mm^2
agx_{mid}	Downward slope of AV vs agx increases when agx$=agm+agx_{\text{mid}}$	6 mm^2	7 mm^2
agx_{min}	Minimum value of agx for which voicing occurs (AV>0)	1 mm^2	1 mm^2
$f1_{\text{min-}F0}$	Lowest first formant used in the calculation of intrinsic pitch	250 Hz	250 Hz
$F1_{\text{min-liq}}$	Minimum **f1** for which a liquid consonant can be assumed	400 Hz	350 Hz
$F1_{\text{max-liq}}$	Maximum **f1** for which a liquid consonant can be assumed	550 Hz	500 Hz
$F2_{\text{max-lateral}}$	Maximum **f2** for which a lateral consonant can be assumed	1400 Hz	1300 Hz
$F3_{\text{min-lateral}}$	Minimum **f3** for which a lateral consonant can be assumed	2900 Hz	2700 Hz
$F2_{\text{max-retro}}$	Maximum **f2** for which a retroflex consonant can be assumed	1600 Hz	1400 Hz
$F3_{\text{max-retro}}$	Maximum **f3** for which a retroflex consonant can be assumed	2000 Hz	1800 Hz
OQ_{max}	Maximum value of OQ	99%	99%
OQ_{min}	Minimum value of OQ	0%	0%
$P_{thr\text{-}m}$	Modal phonation threshold pressure	3.5-cm H_2O	3.5-cm H_2O
TL_{max}	Maximum value of TL	41 dB	41 dB
TL_{min}	Minimum value of TL	0 dB	0 dB

TABLE VIII. Speaker constants: Scale factors.

Constant	Description	Female	Male
K_v	Scale factor for computation of amplitude of voicing, AV	33 dB	33 dB
$K_{dAVagm-}$	Slope of AV vs agx curve for agx$<agm$	220 dB/cm^2	200 dB/cm^2
$K_{dAV1agm+}$	Slope for first portion of AV vs agx curve for agx$>agm$	120 dB/cm^2	100 dB/cm^2
$K_{dAV2agm+}$	Slope of second portion of AV vs agx curve for agx$>agm$	420 dB/cm^2	400 dB/cm^2
K_{AH}	Scale factor for AH computation	27 dB	27 dB
K_{AF}	Scale factor for AF computation	40 dB	40 dB
K_{OQ}	Scales effect of glottal opening on open quotient, OQ	3.96%/mm^2	3.3%/mm^2
K_{TL}	Scales effect of glottal opening, nasal port, and supraglottal constriction on the source spectral tilt, TL	1.8 dB/mm^2	1.5 dB/mm^2
K_{DI}	Scale factor controlling degree of diplophonia when agx$<agm$	15	15
$K_{dF0\text{-}dc}$	Scales effect of parameter **dc** on fundamental frequency $F0$	3 dHz/percent	3 dHz/percent
$K_{dF0\text{-}ps}$	Scales effect of parameter **ps** on fundamental frequency $F0$	30 dHz/cm H_2O	30 dHz/cm H_2O
$K_{dF0\text{-height}}$	Scale factor for effect of vowel height on fundamental frequency $F0$	0.46 ms	0.5 ms
K_{acl}	Scale factor for computing the cross-sectional area of a liquid constriction from **f1**	25 mm^2	25 mm^2
$K_{Pthr\text{-}dc}$	Scales effect of parameter **dc** on phonation threshold pressure	0.03 cm H_2O/percent	0.03 cm H_2O/percent
K_{Cw}	Scales effect of **dc** on the compliance of the vocal-tract walls, C_w	1 (dimensionless)	1 (dimensionless)
K_{Cg}	Scales effect of **dc** on the compliance of the vocal folds, C_g	0.34 (dimensionless)	0.34 (dimensionless)

[1]In using the synthesizer it is convenient to specify such parameters in units that are commonly used in the literature. In particular, cross-sectional areas of constrictions are expressed in mm^2 and subglottal pressure is expressed in cm H_2O. The remaining parameters are expressed in cgs units. In formulas to be discussed later, constants may be introduced to convert areas and pressures to cgs units.

[2]When the tongue body forms a narrow constriction in the oral cavity, the first formant is relatively low, compared to its value for a more neutral position. Likewise, when a narrow constriction is formed at the pharynx, the first formant becomes relatively high. Thus, the constriction size can be estimated from **f1**, and does not need to be specified separately. (See Sec. III A 2.) When a more localized and short constriction is formed with the lips or the tongue blade, the cross-sectional area is generally not well predicted by **f1**. The parameters **al** and **ab**, representing these constriction areas, are specified separately. When these parameters are introduced to synthesize labial or tongue-blade consonants, a correction to **f1** is made depending on the value of the **al** or **ab** parameter. Thus, the original **f1** parameter is, in effect, the natural frequency of the vocal tract without local constrictions. These relations between **f1** and vocal-tract constrictions are described in more detail in a future paper.

[3]The parameter trajectories for this and later examples are simplified to be sequences of straight lines. For high-quality synthesis, trajectories should probably be more smooth.

[4]Note that 0 dB on a plot of AV or AF (and later AH) actually means a source of zero amplitude.

[5]The paper is based on our current working version of HLsyn, which is constantly being updated and may vary somewhat from the commercial version.

[6]This statement is based on perturbation theory that relates small changes in the cross-sectional area of a tube to changes in the natural frequencies of the tube. When the perturbations in the tube extend over a length of 3 or more centimeters, such as perturbations of the kind produced by moving the tongue body, only the first two (and possibly three) natural frequencies are influenced, and higher natural frequencies remain relatively unchanged. It follows, then, that if the first and second formant frequencies are known, and if there are no local constrictions with a length of less than 3 centimeters, then the tongue-body position can be predicted from $F1$ and $F2$. In particular, the degree of constriction in the vocal tract can be predicted from $F1$. In English, lip rounding usually accompanies nonlow back vowels, and is not adjusted independently of tongue-body position.

[7]The choice of $\mathsf{acd} = 100\ mm^2$ is somewhat arbitrary. The point is that this value of acd is well above the size for which frication noise is generated.

[8]The speaker constants K_v, K_{dAV}, K_{AF}, and K_{AH} are set so that the amplitude of voicing (AV) is about 60 dB during a vowel, the amplitude of frication (AF) is about 60 dB during a fricative, and the amplitude of aspiration (AH) is about 40 dB during vowels and aspirated stops.

[9]Parameter and audio files corresponding to these examples are accessible through the worldwide web at ⟨http://speech.mit.edu/publications⟩.

Arkebauer, H. J., Hixon, T. J., and Hardy, J. C. (**1967**). ''Peak intraoral air pressures during speech,'' J. Speech Hear. Res. **10**, 196–208.

Atkinson, J. E. (**1978**). ''Correlation analysis of the physiological factors controlling fundamental voice frequency,'' J. Acoust. Soc. Am. **63**, 211–222.

Baer, T. (**1975**). ''Investigation of phonation using excised larynges,'' Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Baer, T., Gore, G. 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.

Bell-Berti, F. (**1975**). ''Control of pharyngeal cavity size for English voiced and voiceless stops,'' J. Acoust. Soc. Am. **57**, 456–461.

Bickley, C. A., Stevens, K. N., and Williams, D. R. (**1997**). ''A framework for synthesis of segments based on pseudoarticulatory parameters,'' in *Progress in Speech Synthesis*, edited by J. P. H. van Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg (Springer, New York).

Bouhuys, A., Mead, J., Proctor, D. F., and Stevens, K. N. (**1968**). ''Pressure-flow events during singing,'' in *Sound Production in Man* (New York Academy of Sciences, New York).

Browman, C. P., and Goldstein, L. (**1990**). ''Tiers in articulatory phonology, with some implications for casual speech,'' in *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*, edited by J. Kingston and M. E. Beckman (Cambridge University Press, Cambridge).

Espy-Wilson, C. Y. (**1992**). ''Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English,'' J. Acoust. Soc. Am. **92**, 736–757.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Fant, G. (**1972**). ''Vocal tract wall effects, losses, and resonance bandwidths,'' Speech Trans. Lab. Q. Prog. Stat. Rep. 2–3, Royal Institute of Technology, Stockholm, pp. 28–52.

Fant, G. (**1982**). ''Preliminaries to analysis of the human voice source,'' Speech Trans. Lab. Q. Prog. Stat. Rep. 4, Royal Institute of Technology, Stockholm, pp. 1–27.

Hanson, H. M. (**1995a**). ''Glottal characteristics of female speakers,'' Ph.D. thesis, Harvard University, Cambridge, MA.

Hanson, H. M. (**1995b**). ''Synthesis of female speech using the Klatt synthesizer,'' Speech Communication Group Working Papers 10, Research Laboratory of Electronics, MIT, pp. 84–103.

Hanson, H. M. (**1997**). ''Glottal characteristics of female speakers: Acoustic correlates,'' J. Acoust. Soc. Am. **101**, 466–481.

Hanson, H. M., and Stevens, K. N. (**2000**). ''Modeling stop-consonant releases for synthesis,'' J. Acoust. Soc. Am. **107**, 2907.

Hanson, H. M., Stevens, K. N., and Beaudoin, R. E. (**1997**). ''New parameters and mapping relations for the HLsyn speech synthesizer,'' J. Acoust. Soc. Am. **102**, 3163.

Hirose, H., and Niimi, S. (**1987**). ''The relationship between glottal opening and the transglottal pressure differences during consonant production,'' in *Laryngeal Function in Phonation and Respiration*, edited by T. Baer, C. Sasaki, and K. Harris (Little, Brown, Boston).

Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). ''Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice,'' J. Acoust. Soc. Am. **84**, 511–529; **85**, 1787(E).

Holmes, J. N. (**1983**). ''Formant synthesizers: Cascade or parallel?'' Speech Commun. **2**, 251–273.

Isshiki, N. (**1964**). ''Regulatory mechanism of voice intensity variation,'' J. Speech Hear. Res. **7**, 17–29.

Klatt, D., and Klatt, L. (**1990**). ''Analysis, synthesis, and perception of voice quality variations among female and male talkers,'' J. Acoust. Soc. Am. **87**, 820–857.

Ladefoged, P. (**1962**). ''Subglottal activity during speech,'' in *Proceedings of the Fourth International Congress of Phonetic Sciences* (Mouton, The Hague), pp. 73–91.

Lin, Q., and Fant, G. (**1990**). ''A new algorithm for speech synthesis based on vocal tract modeling,'' Speech Trans. Lab. Q. Prog. Stat. Rep. 2–3, Royal Institute of Technology, Stockholm, pp. 45–52.

Lucero, J. C. (**1995**). ''The minimum lung pressure to sustain vocal fold oscillation,'' J. Acoust. Soc. Am. **98**, 779–784.

Massey, N. S. (**1994**). ''Transients at stop-consonant releases,'' Master's thesis, Massachusetts Institute of Technology, Cambridge, MA.

Narayanan, S., Alwan, A., and Song, Y. **(1997)**. ''New results in vowel production: MRI, EPG, and acoustic data,'' in Proceedings of 1997 European Speech Processing Conference **2**, pp. 1007–1009.

Ní Chasaide, A., and Gobl, C. (**1993**). ''Contextual variation of the vowel voice source as a function of adjacent consonants,'' Lang. Speech **36**, 303–330.

Oliveira, L. C. (**1997**). ''Text-to-speech synthesis with dynamic control of source parameters,'' in *Progress in Speech Synthesis*, edited by J. P. H. van Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg (Springer, New York).

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (**1986**). *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, Cambridge).

Rothenberg, M. (**1968**). *The Breath Stream Dynamics of Simple-Released-Plosive Production.* Number 6 in Bibliotheca Phonetica (Karger, Basel).

Sensimetrics Corp. (**1992**). ''SenSyn™.''

Shadle, C. **(1985)**. ''The acoustics of fricative consonants,'' Technical Report 506, Research Laboratory of Electronics, MIT, Cambridge, MA.

Slifka, J. **(2000)**. ''Respiratory constraints at prosodic boundaries in speech,'' Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Stevens, K. N. (**1971**). ''Airflow and turbulence noise for fricative and stop consonants: Static considerations,'' J. Acoust. Soc. Am. **50**, 1180–1192.

Stevens, K. N. (**1993**). ''Models for the production and acoustics of stop consonants,'' Speech Commun. **13**, 367–375.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Stevens, K. N. (**2001**). ''The properties of the vocal-tract walls help to shape several phonetic distinctions in language,'' in *To Honour Eli Fischer-Jørgensen: Festschrift on the Occasion of Her 90th Birthday, February 11th, 2001*, edited by N. Grønnum and J. Rischel (Reitzel, Copenhagen).
Stevens, K. N., and Bickley, C. A. (**1991**). ''Constraints among parameters simplify control of Klatt formant synthesizer,'' J. Phonetics **19**, 161–174.
Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., and Wilhelms-Tricarico, R. (**1997**). ''Tongue surface displacement during bilabial stops,'' J. Acoust. Soc. Am. **102**, 562–571.
Titze, I. R. (**1988**). ''The physics of small-amplitude oscillation of the vocal folds,'' J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R. (**1989**). ''On the relation between subglottal pressure and fundamental frequency in phonation,'' J. Acoust. Soc. Am. **85**, 901–906.
Titze, I. R. (**1992**). ''Phonation threshold pressure: A missing link in glottal aerodynamics,'' J. Acoust. Soc. Am. **91**, 2926–2935.
Verdolini, K., Druker, D. G., Palmer, P. M., and Samawi, H. (**1998**). ''Laryngeal adduction in resonant voice,'' J. Voice **12**, 315–327.

Effect of overpressure and pulse repetition frequency on cavitation in shock wave lithotripsy

Oleg A. Sapozhnikov and Vera A. Khokhlova
Department of Acoustics, Physics Faculty, M. V. Lomonosov Moscow State University, Moscow 119899, Russia

Michael R. Bailey[a)]
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, College of Ocean and Fishery Sciences, University of Washington, 1013 Northeast 40th Street, Seattle, Washington 98105

James C. Williams, Jr. and James A. McAteer
Department of Anatomy and Cell Biology, MS 5055, Indiana University School of Medicine, 635 Barnhill Drive, Indianapolis, Indiana 46202-5120

Robin O. Cleveland
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

Lawrence A. Crum
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, College of Ocean and Fishery Sciences, University of Washington, 1013 Northeast 40th Street, Seattle, Washington 98105

(Received 8 February 2002; revised 15 June 2002; accepted 20 June 2002)

Cavitation appears to contribute to tissue injury in lithotripsy. Reports have shown that increasing pulse repetition frequency [(PRF) 0.5–100 Hz] increases tissue damage and increasing static pressure (1–3 bar) reduces cell damage without decreasing stone comminution. Our hypothesis is that overpressure or slow PRF causes unstabilized bubbles produced by one shock pulse to dissolve before they nucleate cavitation by subsequent shock pulses. The effects of PRF and overpressure on bubble dynamics and lifetimes were studied experimentally with passive cavitation detection, high-speed photography, and *B*-mode ultrasound and theoretically. Overpressure significantly reduced calculated (100–2 s) and measured (55–0.5 s) bubble lifetimes. At 1.5 bar static pressure, a dense bubble cluster was measured with clinically high PRF (2–3 Hz) and a sparse cluster with clinically low PRF (0.5–1 Hz), indicating bubble lifetimes of 0.5–1 s, consistent with calculations. In contrast to cavitation in water, high-speed photography showed that overpressure did not suppress cavitation of bubbles stabilized on a cracked surface. These results suggest that a judicious use of overpressure and PRF in lithotripsy could reduce cavitation damage of tissue while maintaining cavitation comminution of stones.
[DOI: 10.1121/1.1500754]

PACS numbers: 43.80.Gx [FD]

I. INTRODUCTION

Shock wave lithotripsy (SWL) has proven to be highly successful for the treatment of urinary stones. Indeed, even with the emergence of endourological approaches for stone removal, the majority of calyceal and upper ureteral stones are treated by SWL.[1] Although it is accepted that SWL works very well to break stones, there is increased awareness that shock waves cause collateral damage to the kidney.[2]

Concern over adverse effects has stimulated interest in finding ways to improve lithotripsy. A logical first step toward this goal has been the effort to determine the physical mechanisms of shock-wave action. This work, primarily involving *in vitro* studies, has shown that cavitation plays an important role in stone fragmentation.[3–5] Likewise, cavitation is strongly implicated in causing the hemorrhagic lesion that characterizes SW-induced damage to the kidney.[6–10] Thus, cavitation appears to be involved both in stone comminution and tissue damage.

Several strategies have been proposed to control cavitation in lithotripsy including alteration of the shock pulse to suppress bubble growth and protect tissue,[11] and the use of dual sequential pulses timed to increase the force of bubble collapse and, thereby, enhance stone breakage.[12–16] These strategies are potentially very useful, but they require that physical modifications be made to the lithotripter.

Several studies have explored alternative methods to influence cavitation in SWL. One approach has been to use increased static pressure (overpressure) to minimize cavitation at the focal point of the lithotripter. Delius[17] found that relatively low overpressures had a dramatic protective effect on isolated cells. Maximal cell protection was achieved by overpressure of only ≈1 bar. Overpressure also reduced SW damage to stones, but the effect on cells was much more pronounced. This opens the possibility that if low overpressure could be applied during SWL, tissue-damaging cavita-

a)Electronic mail: bailey@apl.washington.edu

tion in the vasculature might be suppressed without inhibiting stone fragmentation.

Another means to regulate cavitation and one that is available to the urologist at the time of treatment is to change the pulse-repetition frequency (PRF). Increasing the SW delivery rate generates more cavitation bubbles.[18] It has recently been shown that stone fragmentation *in vitro* and *in vivo* is significantly improved by slowing the SW rate.[19] It has also been demonstrated that delivery of SWs at very fast rate increases kidney injury in experimental animals.[20]

These findings suggest that the combined use of overpressure and manipulation of SW rate may result in reduced tissue damage without compromising stone comminution. However, it is difficult to assess the potential value of these strategies or to know how best to apply them because the mechanisms of the overpressure effect and the rate effect are not known. Therefore, we undertook a study involving numerical calculations, ultrasound imaging, and high-speed photography to characterize the effect of low overpressure and SW rate on cavitation bubble dynamics at the focus of an electrohydraulic lithotripter. We tested the hypothesis that increased static pressure and decreased PRF act to reduce cavitation activity by allowing bubbles to dissolve between shock pulses. Bubbles that reside in close contact with the surface of a stone may be stabilized by that interaction and are more resistant to dissolution.[21–23] Thus under conditions of overpressure or low PRF bubbles free within the surrounding medium dissolve between pulses while bubbles that are stabilized within cracks and crevices persist and continue to play a role in stone comminution.

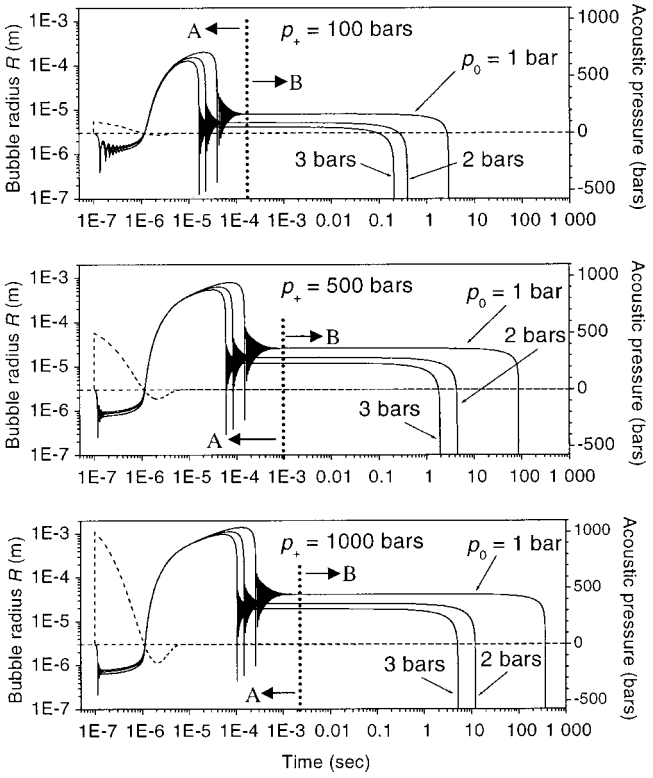


FIG. 1. Calculated $R(t)$ curves (solid lines) for different amplitude shock waves $p_+ = 100$, 500, and 1000 bar and varying static pressure $p_0 = 1$, 2, and 3 bar. Initial bubble radius was $R_0 = 3\ \mu$m. The driving lithotripter wave forms are shown as dashed lines. Time intervals A and B, divided by dotted lines, distinguish different stages in bubble dynamics. The maximum bubble radius and the bubble lifetime increased with p_+ and decreased with p_0.

II. THEORY

A. Basic equations

Our numerical model is the synthesis of a bubble dynamics model created by Church[24] based on the Gilmore equation[25] and a bubble dissolution model described by Epstein and Plesset.[26] As shown in Fig. 1, our model traces the radius of a bubble $R(t)$ over time as the lithotripter pulse sets the bubble into a dramatic growth and collapse followed by a slow dissolution of the bubble. The lifetime of a bubble (t_L^*, the asterisk denotes a calculated value) was defined as the time between the shock wave arrival until the bubble dissolution. The model was used to calculate the lifetime t_L^* as a function of overpressure, initial bubble radius, and shock wave amplitude.

1. Lithotripter pulse model

In the calculations, it was supposed that a small bubble of radius R_0 exists in the medium. A lithotripter pulse arrives at $t=0$ and initiates the bubble dynamics. To represent the wave form radiated by an electrohydraulic lithotripter source, a theoretical temporal profile $p'(t)$ was defined by Church[24] as $p'(t) = p_+ \Theta(t) 2e^{-\alpha t} \cos(\omega t + \pi/3)$, where p_+ is the peak positive pressure of the pulse, $\Theta(t)$ is Heaviside step function, and $\alpha = 9.1 \times 10^5\ \mathrm{s}^{-1}$ and $\omega/2\pi = 8.33 \times 10^4\ \mathrm{s}^{-1}$ are characteristic decay constants. The pulse, modeled after measured wave forms, is a short ($\approx 1\ \mu$s) spike followed by a longer ($\approx 4\ \mu$s) pressure trough, so that the total area of the wave form is equal to zero. The amplitude of the pulse is characterized by the peak positive p_+ pressure which has a fixed ratio with the peak negative p_- pressure $p_+/p_- \approx 6.25$.

2. Bubble dynamics model

In the lithotripsy literature it is common to use a cavitation model based on the behavior of a single spherical bubble.[24] Although this model is somewhat idealistic (in reality a cluster of bubbles exists, and the bubbles are not spherical), it adequately describes many features of the cavitation phenomena.[11,27–29] In this work, the model of a single spherical cavitation bubble (initial radius R_0 between 1 and 30 μm) was employed to study the effect of static pressure ($p_0 = 1-4$ bar) on cavitation in lithotripsy. Note that p_0 is the absolute static pressure and therefore the overpressure in these calculations varied from 0 to 3 bar. Note that pressures are expressed in bar not MPa for convenient comparison to standard atmospheric pressure (1 bar). Fundamental assumptions of the model are as follows: a single bubble exists in an infinite liquid medium; the bubble remains spherical at all times; the interior of the bubble (e.g., temperature, pressure, gas concentration) is spatially uniform; the radius of the bubble is much less than the characteristic length of the acoustic excitation at the corresponding time; body forces (e.g., gravity) have negligible effect on the bubble pulsation; the gas within the bubble behaves as an ideal gas; gas con-

tent in the bubble is changing because of gas diffusion into and out of the surrounding liquid; and vapor pressure in the gas is constant.

The bubble radius $R(t)$ is described by the Gilmore equation,[25]

$$\left(1-\frac{\dot{R}}{C}\right)R\ddot{R}+\frac{3}{2}\left(1-\frac{\dot{R}}{3C}\right)\dot{R}^2 = \left(1+\frac{\dot{R}}{C}\right)H+\left(1-\frac{\dot{R}}{C}\right)\frac{RdH}{Cdt}, \tag{1}$$

where a dot signifies a time derivative, C is the sound speed in the liquid at the bubble wall, and H is the difference between the specific enthalpy in water at the bubble wall relative to the specific enthalpy in the liquid far from the bubble,

$$H=\int_{p_\infty}^{p(R)}\frac{dp}{\rho}. \tag{2}$$

Here $p(R)$ is the pressure in the liquid at the bubble wall and $p_\infty=p_0+p'$ is the pressure, static p_0 plus acoustic p', in liquid far from the bubble. Pressure p and density ρ in the liquid are related by the Tait equation,

$$p=A(\rho/\rho_0)^\Gamma-B, \tag{3}$$

where $A=C_0^2\rho_0/p_0\Gamma$, $B=A-p_0$, Γ is an empirical constant, p_0 is the static pressure in the liquid. The parameters of water were used for further numerical simulations: $\Gamma=7$, the equilibrium density $\rho_0=1000\ \mathrm{kg/m^3}$, the sound speed $C_0=1485$ m/s. From Eqs. (2) and (3), the enthalpy can be expressed in terms of pressure:

$$H=\frac{A^{1/\Gamma}}{\rho_0}\frac{\Gamma}{\Gamma-1}\{[p(R)+B]^{(\Gamma-1)/\Gamma}-[p_\infty+B]^{(\Gamma-1)/\Gamma}\}. \tag{4}$$

The sound speed at the bubble wall, C, can be found from the enthalpy:

$$C^2=C_0^2+(\Gamma-1)H. \tag{5}$$

The pressure in the liquid at the bubble wall is given by

$$p(R)=p_g-\frac{2\sigma}{R}-\frac{4\mu\dot{R}}{R}, \tag{6}$$

where p_g is the pressure of the gas within the bubble, $\sigma=0.0725$ N/m is the surface tension, and $\mu=0.001$ kg/m/s is the coefficient of shear viscosity of water. The gas pressure p_g is proportional to the quantity of gas in the bubble and the gas temperature. Following the method described by Church,[24] we model the complicated bubble dynamics at each temporal step as two elementary separate substeps. At each step, we model the change of bubble radius due to the gas diffusion isothermally and the change of bubble radius due to pressure and inertia adiabatically. The calculation proceeds as follows: first, the bubble radius was modeled as changing isothermally from its current value R to a so-called equilibrium radius R_{0n}, following the change of the gas in the bubble to its current number of moles n; and second, the bubble radius changes adiabatically from the R_{0n} to the new radius R with no change in n.

The first process of bubble radius change due to diffusion takes place slowly under constant static pressure, so that gas pressure is $p_0+2\sigma/R_{0n}$. The equilibrium radius can be found from the equation of the isothermal process in an ideal gas, written for the initial and current parameters of the bubble:

$$\frac{(p_0+2\sigma/R_0)R_0^3}{n_0}=\frac{(p_0+2\sigma/R_{0n})R_{0n}^3}{n}. \tag{7}$$

The solution of this cubic equation can be expressed in the following form:

$$R_{0n}=(a+\sqrt{a^2-b})^{1/3}+(a-\sqrt{a^2-b})^{1/3}-\frac{2\sigma}{3p_0}, \qquad a=\frac{n}{n_0}\frac{R_0^3}{2}\left(1+\frac{2\sigma}{p_0R_0}\right)-\left(\frac{2\sigma}{3p_0}\right)^3, \quad b=\left(\frac{2\sigma}{3p_0}\right)^6. \tag{8}$$

In the second adiabatic process, the equation for R is $p_gR^{3\gamma}=(p_0+2\sigma/R_{0n})R_{0n}^{3\gamma}$, where γ is the adiabatic exponent of the gas. Using Eq. (7), the gas pressure can be written as follows:[24]

$$p_g=\left(p_0+\frac{2\sigma}{R_0}\right)\frac{n}{n_0}\left(\frac{R_0}{R}\right)^{3\gamma}\left(\frac{R_{0n}}{R_0}\right)^{3(\gamma-1)}. \tag{9}$$

The gas diffusion in the liquid is described by the diffusion equation

$$\frac{\partial c}{\partial t}+u\frac{\partial c}{\partial r}=D\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial c}{\partial r}\right), \tag{10}$$

where c is gas concentration (the number of moles of gas per unit volume), r is the radial coordinate, u is the radial component of the liquid velocity, and D is the diffusion constant of the gas in the liquid. The constant $D=2.42\times10^{-9}\ \mathrm{m^2/s}$ for air dissolved in water at 20 °C. The liquid can be considered as incompressible most of the time, which gives the following expression for the radial velocity: $u=R^2\dot{R}/r^2$. At the initial time, the concentration of dissolved gas is uniform and equal to c_i. Usually this value is less than the saturated gas concentration c_0 at the initial static pressure p_0. The relative initial concentration can be characterized by the ratio $f=c_i/c_0$. The value of c_0 is related to the static pressure p_0 by Henry's law $c_0=k_Hp_0$, where k_H is a constant. Equation (10) for the dissolved gas concentration $c(r,t)$ has to be solved at $t>0$ and $r>R$. The initial condition is $c(r,0)=c_i$, the boundary conditions are $c(\infty,t)=c_i$ and $c(R,t)=c_s$, where c_s is saturated gas concentration in liquid at the bubble interface. According to Henry's law, $c_s=k_Hp_g$, or, neglecting the dependence of the Henry's constant k_H on temperature,

$$c_s=c_0p_g/p_0. \tag{11}$$

We are interested in calculating the bubble lifetime t_L^*, which includes excitation of the bubble by a lithotripter pulse, consequent expansion, collapse, rebounds, and finally the bubble dissolution due to gas diffusion from the bubble interior to the liquid. The direct numerical solution of Eq. (10) is computationally intensive; therefore appropriate approximations are desired to simplify it. These approxima-

tions have already been used in papers on cavitation and are different for different stages in the bubble cycle.[24,26,30] We use the approximation by Eller and Flynn,[30] as Church[24] did, for region A in Fig. 1 where the bubble radius is rapidly changing, and the approximation by Epstein and Plesset[26] for region B where the static bubble slowly dissolves. The application and appropriateness of the models are discussed in the following with reference to the radius versus time plots in Fig. 1. Church[24] described in more detail the application of the Eller and Flynn model to lithotripsy.

During the time when the bubble is growing and collapsing (region A in Fig. 1), the gas concentration in the liquid varies from the saturated value c_s to the uniform value c_i in a very thin spherical layer near the gas–water interface, because the interface is moving. The thickness of this layer can be estimated from Eq. (10) as $\delta=\sqrt{Dt_*}$, where t_* is a characteristic time. During the first $t_1\approx 3\ \mu$s, the transition layer thickness is on the order of $\delta_1\approx 0.1\ \mu$m, which is much less than the bubble radius at this time, $R_1\approx 3-200\ \mu$m (see Fig. 1). For hundreds of microseconds, the layer continues to grow ($\delta_2\approx 1\ \mu$m), but so does the bubble ($R_2\approx 100-1000\ \mu$m). The use of the inequality $\delta/R\ll 1$ in Eq. (10), as was shown by Eller and Flynn[30] in their first-order approximation, results in the following expression for the number of gas moles in the bubble:

$$n=n_0-4\sqrt{\pi D}\int_0^{\tau}\frac{F(\tau')}{\sqrt{\tau-\tau'}}d\tau', \tag{12}$$

where $\tau=\int_0^t R^4(t')dt'$ and $F=c_s-c_i$.

Bubble growth is followed by collapse and several ''rebounds,'' subsequent growth and collapse cycles. During each of the collapses the radius of the bubble becomes small enough that the assumption $\delta/R\ll 1$ breaks down. However, each collapse stage is short, a small bubble has a small surface area through which diffusion can take place, and so there is little change in the value of n. Therefore we continue to use Eq. (12) throughout the collapse and rebound region to ensure we capture the diffusion at the times when the bubble is large enough—the error induced in the calculation of n will be negligible.[24]

Then, after about $t_m\approx 1$ ms (region B in Fig. 1), the bubble has stopped pulsating. Its equilibrium radius $R_m=R(t_m)$ is larger than the initial bubble radius R_0 because of gas diffusion into the bubble during bubble growth. Now, surface tension, static pressure, and a gas concentration gradient drive gas diffusion from the bubble to the liquid, leading to the bubble dissolution ($n=0$). Eller and Flynn's approximation is not valid during this process. On the other hand, the bubble radius during dissolution varies relatively slowly. Therefore, it is possible to set $u=0$ in Eq. (10). The resulting diffusion equation has an analytical solution in the form of the convolution of the corresponding Green's function with the initial distribution of the dissolved gas concentration. At time $t=t_m$, the diffusion layer thickness δ_m is of the order of several micrometers, so it is still smaller than the corresponding bubble radius ($R_m\geqslant 10\ \mu$m).[24] Therefore, one can use an approximation $\delta_m\to 0$ to simplify the initial condition at $t=t_m: c(r,t_m)=c_i$. The boundary conditions at $t>t_m$ are the same as at $t<t_m: c(\infty,t)=c_i$ and $c(R,t)=c_s$. Then, for the number of gas moles in the bubble,[24] one has

$$\frac{dn}{dt}=-4\pi R^2 D(c_s-c_i)\left(\frac{1}{R}+\frac{1}{\sqrt{\pi D(t-t_m)}}\right). \tag{13}$$

Note that at small time $t-t_m$, Eq. (13) gives the same expression for dn/dt as Eq. (12) at small t. However for larger time, these equations differ substantially. In Eq. (13) it is not assumed that the diffusion layer is thin.

During the slow changes of the bubble size, the gas temperature can be considered as a constant value equal to the liquid temperature T_0. The isothermal gas process is governed by

$$p_g\tfrac{4}{3}\pi R^3=nGT_0, \tag{14}$$

where $G=8.31$ m^3 Pa/mol K is a universal gas constant. The bubble radius no longer needs to be modeled by the dynamic equations [Eqs. (1)–(9)], instead one can use the quasistatic equation:

$$p_g=p_0+\frac{2\sigma}{R}. \tag{15}$$

Equations (13)–(15) give the following equation for the bubble radius:

$$\frac{dR}{dt}=-K\frac{1-f+\dfrac{2\sigma}{p_0R}}{1+\dfrac{4\sigma}{3p_0R}}\left(\frac{1}{R}+\frac{1}{\sqrt{\pi D(t-t_m)}}\right), \tag{16}$$

where $K=GT_0Dk_H^{-1}$ and $f=c_i/c_0$. The initial condition is $R=R_m$ at $t=t_m$. Note that the choice of the value of t_m is somewhat arbitrary: t_m should be larger than the time when the bubble rebounding finishes, but much smaller than the lifetime t_L. In the case of bubble behavior in the lithotripter field, the choice $t_m=1$ ms is reasonable. The right-hand side of Eq. (16) has singularities at $t\to t_m$ and when $R\to 0$. These singularities can be avoided by calculating the value of $\xi=\sqrt{t-t_m}$ as a function of R, instead of considering R as a function of t. The equation for $\xi=\xi(R)$ follows from Eq. (16):

$$\frac{d\xi}{dR}=-\frac{\sqrt{\pi D}}{2K}\frac{R\left(R+\dfrac{4\sigma}{3p_0}\right)}{(R+\xi\sqrt{\pi D})\left[R(1-f)+\dfrac{2\sigma}{p_0}\right]}. \tag{17}$$

The function $\xi(R)$ is to be calculated when the bubble radius changes from $R=R_m$ to $R=0$ to complete dissolution of the bubble. The initial condition is $\xi(R=R_m)=0$.

Therefore, Eq. (1) was solved numerically to calculate rapid bubble motion, and Eq. (17) was then solved to calculate the slow bubble dissolution. To solve for fast bubble motion, Eq. (1) was written as two first-order differential equations for bubble wall radius R and velocity dR/dt. A fifth-order Runge–Kutta algorithm with adaptive step-size control was used on a nonuniform time grid $\Delta t(t)$ that followed the dynamics of the bubble.[31] Equation (12) was integrated simultaneously on the same time grid and yielded a

calculation of gas diffusion during rapid motion. The number of moles n, the derivative dn/dt, equilibrium bubble radius R_{0n}, and gas pressure in the bubble p_g were calculated at each time point t for use in solving Eq. (1) at the next time step. The values of p_g and R_{0n} were obtained explicitly using Eqs. (8) and (9) from the known values of R and n. The derivative dn/dt was calculated using backward finite difference with the solutions for n at the current and previous time points. However, the solution for the number of moles n, Eq. (12), is implicit, as it depends on not-yet-known values of p_g and R_{0n}. An iterative procedure was employed at each time point to calculate $n \rightarrow R_{0n} \rightarrow p_g \rightarrow n \rightarrow \cdots$.[24] Initial values of R_{0n} and p_g for iteration were obtained from the number of moles, $n + dn/dt\,\Delta t$, extrapolated linearly from the previous time point. Linear extrapolation for n was also used for the Runge–Kutta evaluations within the full time step. Run times for each $R(t)$ curve were less than 10 min on a Compaq XP1000 workstation. During the second slow stage of bubble dissolution, Eq. (17) was solved numerically by fourth-order Runge–Kutta algorithm. Parameter values chosen were for water at laboratory conditions: $\sigma = 0.0725$ N/m, $\mu = 0.001$ N s/m^2, $\rho_0 = 1000$ kg/m^3, $C_0 = 1485$ m/s, $T_0 = 20\,^\circ$C, $\gamma = 1.4$, and $k_H = 0.782 \times 10^{-5}$ mol/m/N.

The value of initial gas concentration in water is an important parameter in studying the effect of overpressure. Two limiting cases can be considered. In one limiting case, the water has been given sufficient time (and mixing) at overpressure to become saturated: absolute gas concentration c_i increases proportionally with p_0, such that the relative gas concentration $f = 1$ for all p_0. In the other limiting case, the ambient pressure changes quickly so that c_i is constant, and f varies inversely proportionally with p_0. Our experiments closely modeled the latter case because our system had very small gas/water interfaces through which air could diffuse. Therefore, all simulations were performed assuming a fast change of the pressure p_0. Initial gas concentration c_i in water was taken equal to the saturation value at atmospheric conditions, $c_i = c_0(p_0 = 1\text{ bar}) = 0.7899$ mol/m^3, which corresponds to $f = 1$ for $p_0 = 1$ bar and $f \propto 1/p_0$. The effect of variable f has been explored by Epstein and Plesset,[26] and as might be expected, the lower the concentration of air in the water the faster the bubble dissolves.

B. Theoretical results

Calculated $R(t)$ curves describing the bubble life are shown in Fig. 1 for a bubble with initial radius 3 μm. Calculations are shown for shock waves with peak acoustic pressures $p_+ = 100$, 500, and 1000 bar and for absolute static pressure $p_0 = 1$, 2, and 3 bar. The bubble radius curves are solid, and each dashed line represents the acoustic pressure of the lithotripter pulse driving the bubble. All axes except acoustic pressure are plotted in logarithmic scale. It is seen that the initial positive-pressure spike of the lithotripter pulse results in a rapid constriction of the bubble to a radius of less than 1 μm (the first collapse). Then, the negative-pressure phase of the lithotripter pulse initiates the growth of the bubble. The inertia imparted to the liquid surrounding the bubble is sufficiently large that the bubble continues to grow after the lithotripter pulse has passed. Growth continues for a relatively long period (in excess of 100 μs). Eventually, the static pressure in the fluid forces the bubble to collapse again. We denote the time of this inertial collapse as t_c^*. At about 1 ms, after a series of rebounds (subsequent smaller growth and collapse cycles), the bubble attains a stable radius referred to by Church as the ''time-varying equilibrium radius.''[24] The term equilibrium radius is used here, too, although the radius is changing slowly as the bubble dissolves. The equilibrium radius is larger than the initial radius, because during the expansion phase of the bubble, the gas concentration within the bubble is low, and gas from the surrounding liquid diffuses into the expanded bubble. However, the nonpulsating bubble slowly dissolves, the bubble radius tends to zero, and finally, the bubble disappears. This time t_L^* defines the bubble lifetime.

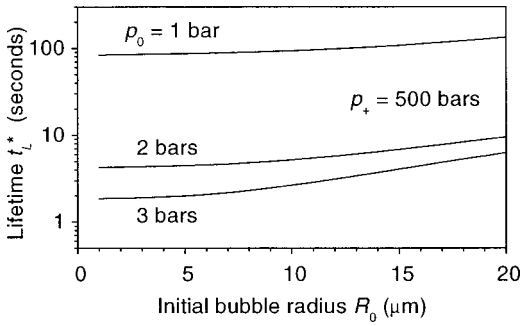


FIG. 2. Calculated bubble lifetime t_L^* for $p_+ = 500$ bar as a function of initial bubble radius R_0 for $p_0 = 1$, 2, and 3 bar. Lifetime, which starts with arrival of the lithotripter pulse and ends with bubble dissolution, is fairly insensitive to R_0 especially in the range 1–10 μm.

For a given shock amplitude one can observe that the static pressure has a negligible effect on bubble dynamics during the first 10 μs. The strong negative acoustic pressure of the lithotripter pulse drives the bubble at this period. The peak positive ($p_+ = 100$–1000 bar) and negative ($p_- = 16$–160 bar) pressure amplitudes in the lithotripter pulse are much greater than p_0 (1–3 bar).

After the lithotripter pulse has passed, the static pressure becomes the main factor driving the inertial behavior of the bubble. As a result, the increase in p_0 gives rise to a shortening of the collapse time t_C^*, a decrease in maximum bubble radius, and an increase in minimum radius. The equilibrium radius decreases and the lifetime t_L^* shortens for higher p_0.

The lifetime t_L^* shows how fast cavitation bubbles dissolve—these bubbles are potential nucleation seeds for cavitation during a subsequent shock wave. The lithotripter pulse repetition period ($T = 1/\text{PRF}$) is of the order of $T = 1$ s. If the lifetime exceeds this value, the cavitation bubbles do not have enough time to dissolve between successive pulses, and the cavitation should be very pronounced. If, however, $t_L^* < T$, then the cavitation seeds dissolve before the next shock wave arrives, and one would expect that cavitation may be diminished. It is instructive, therefore, to study how the lifetime t_L^* depends on various parameters, in particular, on initial bubble radius, R_0, static pressure, p_0, and the lithotripter pulse peak pressure, p_+.

Figure 2 shows curves for lifetime *versus* initial bubble

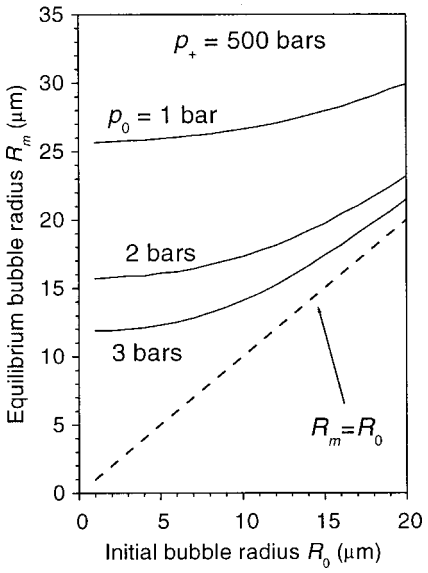


FIG. 3. Equilibrium bubble radius R_m at 1 ms following growth and collapse, as a function of initial bubble radius R_0 for $p_+=500$ bar and $p_0=1$, 2, and 3 bar. Radius R_m is fairly insensitive to R_0 especially in the range 1–10 μm.

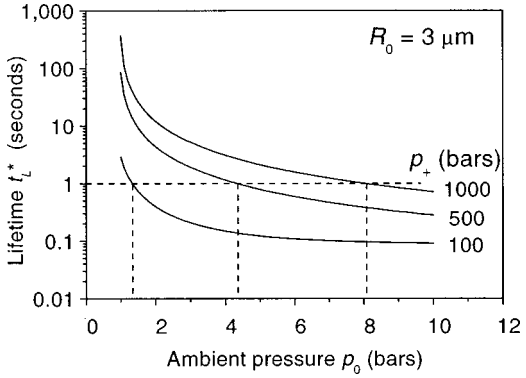


FIG. 4. Calculated lifetime t_L^* vs static pressure p_0 for initial bubble radius $R_0=3$ μm and $p_+=100$, 500, and 1000 bar. Larger negative pressure excursions p_-, corresponding to increased p_+, cause longer and larger bubble growth, which creates increased gas influx, larger R_m, and longer lifetime. The lifetime decreased dramatically with increased static pressure p_0 and with decreased lithotripter pulse amplitude p_+. The horizontal dashed line marks the pulse repetition period of 1 s, which corresponds to a typical clinical rate of 1 Hz. The vertical dashed lines indicate the corresponding static pressure for the different positive peak pressures.

radius at different static pressures p_0. The lithotripter pulse peak pressure equals $p_+=500$ bar. One can see from these curves that lifetime is fairly insensitive to the initial bubble size, especially if $R_0<10$ μm. This insensitivity is in agreement with the results of Church,[24] who calculated that the $R(t)$ curves for a bubble in a lithotripter are practically identical for different R_0. As a result, the amount of the gas that diffuses into the bubble during the relatively long period of its inertial behavior is also insensitive to the value of R_0. Note that the amount of diffused gas exceeds the initial amount of gas in the bubble by a factor of $10–10^4$. Therefore, gas diffusion makes the equilibrium bubble radius after the inertial collapse and rebounds, R_m, fairly independent of R_0. The corresponding curves for R_m vs R_0 calculated at $t=1$ ms are shown in Fig. 3. For small values of R_0, the slopes of the curves are nearly zero; in other words, many values of R_0 produce roughly the same R_m. For larger values of R_0 the amount of diffused gas becomes smaller than the initial amount of gas in the bubble, and R_m asymptotically approaches R_0. For higher static pressures the influence of the initial bubble radius starts at smaller values of R_0.

Figure 4 shows the lifetime t_L^* *versus* static pressure p_0 for different lithotripter pulse amplitudes p_+. Since t_L^* depends only weakly on initial bubble radius, R_0 in all calculations was 3 μm. At $p_0=1$ bar, t_L^* ranges from 3 to 400 s. Larger p_-, corresponding to increased p_+, caused the bubble to grow larger and last longer before it collapsed. The larger, longer growth causes more gas infusion into the bubble, which creates a larger bubble R_m after growth and collapse. The larger bubble subsequently takes longer to dissolve. It is seen that lifetimes quickly decrease with increase of static pressure. At $p_0=4$ bar, the range of t_L^* values is reduced to 0.2–3 s, more than one order of magnitude less than for $p_0=1$ bar. Absolute values of t_L^* are subject to initial gas saturation levels f and the accuracy of the lithotripter pulse model; however modest overpressures two orders of magnitude less than p_+ or p_- strongly reduce t_L^*. Figure 5 shows that for peak pressures p_+ (ordinate) less than 320 bar and static pressure $p_0=3$ bar, t_L^* is less than 1 s. Therefore for a standard clinical exposure (320 bar, 1 Hz),[32] increasing the static pressure to 3 bar means bubbles created by one pulse dissolve before the next pulse arrives.

III. EXPERIMENTS

A. Materials and methods

1. Lithotripters

A Dornier HM-3 electrohydraulic lithotripter (Dornier Medical Systems, GmbH) was used for the pipette experiments shown in Fig. 10. All other experiments were conducted in a research lithotripter patterned after the Dornier HM-3 lithotripter.[32] Refurbished Dornier electrodes (Service

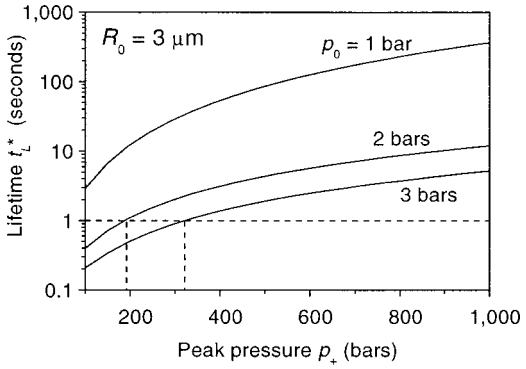


FIG. 5. Calculated lifetime t_L^* vs lithotripter pulse amplitude p_+ for $p_0=1$, 2, and 3 bar, $R_0=3$ μm. Lifetimes are higher for stronger acoustic driving pressures, because the bubble grows larger. Static pressure accelerates dissolution primarily by increasing the gas concentration gradient across the gas–liquid interface. At the clinical lithotripsy conditions $p_+=320$ bar and PRF=1 Hz, static pressure $p_0=3$ bar (bubble lifetime $t_L^*<1$ s) causes bubbles to dissolve between pulses. The horizontal dashed line marks the pulse repetition period of 1 s, which corresponds to a typical clinical rate of 1 Hz. The vertical dashed lines indicate the corresponding positive peak pressures for the different static pressures.

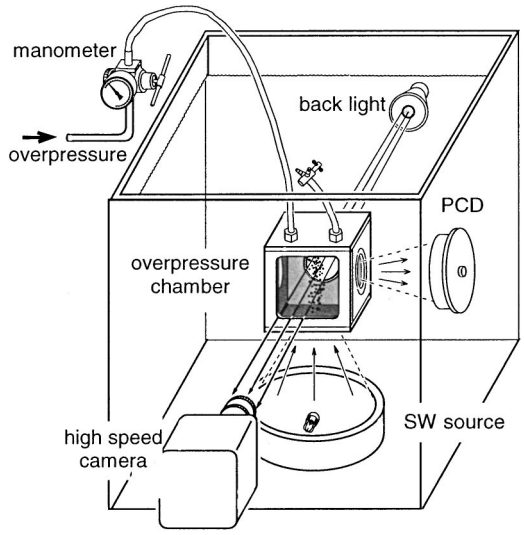


FIG. 6. Experimental setup with the pressure cube. A high-speed camera recorded backlit images of the bubbles through a glass window. A thin mylar window in the cube enabled detection of acoustic emissions from collapsing bubbles with a passive cavitation detector (PCD)—Ref. 33. Lithotripter pulses passed through a polyphenylene oxide (PPO) plate that was acoustically well-matched to water (Ref. 32).

Trends, Kennesaw, GA) were used. Dimensions of the brass ellipsoidal reflector were half major axis $a=13.8$ cm and half minor axis $b=7.75$ cm. The reflector was 1.3 cm short of being a semiellipsoid. Charging potentials 18, 20, and 24 kV were used, and the capacitance was 80 nF. Water was de-ionized and NaCl was added to reach a conductivity of 600 μS/cm. Water was degassed to 4 ppm of oxygen. Water temperature was 37 °C in the clinical bath and 22 °C in the research bath. Peak positive pressures at 18 kV charging potential were 350 ± 80 bar (mean$\pm$standard error for $N=10$ measurements) measured with a Reference Shock Wave Hydrophone (Sonic Industries, Hatboro, PA). Higher charging potential produced larger positive and negative pressures. Cleveland *et al.*[32] described in more detail the design and acoustic output of the research lithotripter.

Lithotripter pulses were triggered at several pulse repetition frequencies (PRFs). The slowest PRF, referred to as the slow rate, was single pulses triggered minutes apart. Clinically relevant rates of 0.5, 1, 2, and 3 Hz were also investigated. Measurements were made on the last of ten pulses triggered at these rates.

2. Containers and pressure chambers

Measurements were made of cavitation in water in pressure chambers placed in the lithotripter. Three types of containers were used: a polyethylene terephthalate (PETE) soda bottle (''pressure bottle'')[33] with a flange, 4 mL, low-density polyethylene pipette bulbs (Sigma Chemical Company, St. Louis, MO), and a plastic chamber (''pressure cube'') with glass and mylar ports (2.5 cm diameter) for optical and acoustic windows.[23] The length, diameter, and wall thickness of the approximately cylindrical pressure bottle were 130, 80, and 0.3 mm and of the pipette bulb were 42, 15, and 0.5 mm. The experimental setup with the pressure cube is shown in Fig. 6. A 12.7-mm-thick polyphenylene oxide (PPO) bottom plate on the chamber transmitted the lithotripter pulses with negligible change of wave form or amplitude at the focus.[34] The top plate was made of the same material, and negligible reflection was measured. Compressed air (1–3 bar gauge) was used to pressurize each chamber. Pressure was elevated for 1 min before measurements were made. The water–air interface was small to minimize gas diffusion into the water and was contained in the 0.25 in. (6.4 mm) inner diameter of the pressure tubing. Hose barbs, quick disconnect pressure fittings, and plastic wire ties were used to connect segments of tubing and pipettes. The water in the chambers was taken from the lithotripter water bath.

3. High-speed photography

Cavitation in the water tank was filmed with a Kodak Ektapro 4540 high-speed digital camera (Eastman Kodak Company, Rochester, NY) at a rate of 40 500 frames/s (i.e., each frame was exposed for 25 μs). We used a 24–120 mm lens with and without a 2×-magnifying lens. The focal area was backlit with a 1000 W lamp and was filmed through the acrylic wall of the tank. The sensor had 64×64 pixels, and images were presented in 256 grayscale levels. The 10 or 20 mm square images (30 mm depth of focus) captured the cluster of bubbles growing and collapsing along the lithotripter axis which ran from lower right to upper left in the image. Images were stored on VHS videotape. Video images were digitized from videotape and analyzed with NIH Image (Bethesda, MD) on a Macintosh G3 computer. The images in Fig. 12 were acquired with an Imacon 200 high speed camera (DRS Hadland, Cupertino, CA) and stored digitally. A 1000 J flash lamp provided lighting from the front, and exposure times were 100 ns/frame. Interframe spacing was 50 μs, and the camera resolution was 1280 by 1024 pixels. In Fig. 12, two images are shown selected from many 15-image movie sequences.

4. Passive cavitation detection

A passive acoustic cavitation detector (PCD) recorded the shock waves emitted during bubble collapse.[35,36] The concave piezoceramic sensing element had a radius of curvature of 200 mm, an aperture diameter of 100 mm, and a resonance frequency of 1.08 MHz. The acoustic axes of the lithotripter and the PCD were perpendicular and confocal. The output signal was high-pass filtered (300 kHz) to remove noise created by excitation of a radial resonance of the sensing element. Signals were demodulated to remove 1 MHz ringing by the PCD. The PCD has been calibrated[35] for cavitation bubble collapses, and a filtered PCD signal of 1 V corresponds to an acoustic emission from a bubble with amplitude 9.5 MPa at 10 mm from the center of the cavitation bubble.

The PCD signal typically consisted of two spikes. The first spike occurred when the lithotripter pulse first interacted with any bubbles initially present in the focal region. The spike resulted from nonlinear scattering from bubbles (i.e., a combination of simple scattering and sound generated by the bubble motion). The bubbles then grew and collapsed. A sec-

ond spike was emitted when the collapse halted. The time between the spikes associated with the growth and collapse is referred to as the collapse time t_C.

Both the PCD and the high-speed camera were simultaneously triggered by a photodetector[32] that detected the spark discharge that initiated the lithotripter pulse. The discharge occurred at time $t=0$. The lithotripter pulse arrived at the focal point $F2$ after 180 μs because it traveled the length of the major axis $2a=276$ mm at a sound speed $C \approx 1.5$ mm/μs. The high-speed camera first recorded bubbles in the frame (each frame was 25 μs in duration) at $t=180$ μs. The PCD had an extra delay of the acoustic travel time from $F2$ to the PCD element (200 mm/1.5 mm/μs $=133$ μs). In PCD measurements presented in this paper, the 133 μs delay has been subtracted from the time axis for easier comparison to camera images.

5. Diagnostic ultrasound

B-mode ultrasound images were obtained using a clinical Bruel and Kjaer 3535 diagnostic ultrasound scanner and a 7.5 MHz curved scan head (model 8545). The frame rate was 34 frames/s, and the displayed mechanical index [MI = derated peak negative pressure (MPa) over the square root of the frequency (MHz)] was 0.8. The scan head was mounted rigidly in the water such that the lithotripter axis was at the 5 cm depth of the image where the scanner was focused. Images were recorded to videotape and later transferred to a Macintosh G3 for analysis. NIH Image software was used to quantify the length of time that echogenic regions on the image persisted. Histograms of pixel-grayscale level were measured for a sequence of image frames. Times were determined by the number of consecutive frames containing pixels within the region of interest brighter than a threshold value.

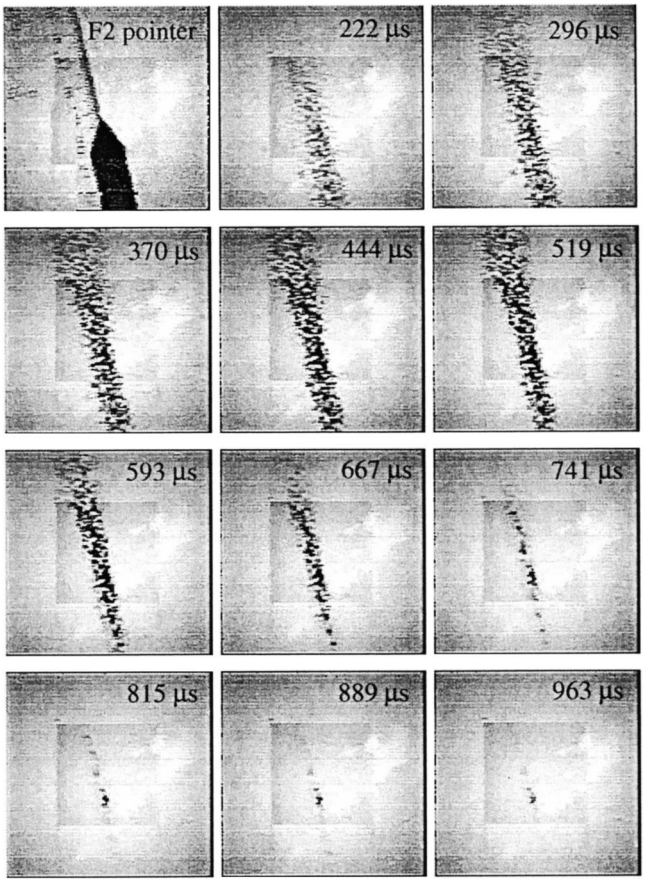


FIG. 7. A time sequence of high-speed photographs of cavitation produced by the lithotripter. The pointer in the first frame (upper left) marks the focal point of the lithotripter and shows the direction of acoustic propagation along the lithotripter axis. Each frame shows a 7.25 by 7.25 cm field of view. A cylindrical cluster ($\approx 1.5\times > 8$ cm) of bubbles formed and collapsed to a 1.5 mm thin line in about 500 μs. Lithotripter charging potential was 24 kV, $p_0=1$ bar, and PRF=1 Hz.

B. Experimental results

1. Free field

The lithotripter created a cylindrical cluster of cavitation bubbles.[11,36,37] Figure 7 shows the cluster in a sequence of high-speed frames. In the first frame, a pointer indicates the geometric focus $F2$ of the lithotripter and the direction of pulse propagation along the lithotripter axis. The remaining frames show the result of the last lithotripter pulse in a sequence of ten lithotripter pulses triggered at 1 Hz and 24 kV. Static pressure p_0 was 1 bar. At 222 μs (42 μs after the shock wave arrived at the focus) a cluster of bubbles was starting to grow along the shock wave axis. The bubbles grew in size, and the cluster reached dimensions of 1.5 cm in diameter by greater than 8 cm in length (frames at 444 and 519 μs). The cluster then collapsed to a narrow line of bubbles (frame at 741 μs) after more than 500 μs. The last two frames showed ''rebound'' or secondary growth and collapse of the cluster.[35]

2. Overpressure

Overpressure dramatically suppressed bubbles as seen in Fig. 8. Both images show the pressure bottle[33] positioned in the lithotripter field. In the top frame, pressure in the bottle and in the surrounding water was 1 bar, and the field was very similar to that in Fig. 7 without the bottle. In the bottom frame, static pressure in the bottle had been increased to 4 bar, and no bubbles were seen in the bottle.

Acoustic emissions from the bubbles were observed to correlate with collapses of the bubble cluster. Figure 9 shows video images of the cluster and simultaneous PCD measurements recorded in the pressure cube. At $p_0=1$ bar, a large dense cluster grows and collapses in $t_C \approx 350$ μs. A strong spike was recorded by the PCD at 180 μs when bubbles were first seen in the images and again ≈ 350 μs later when the bubbles collapsed. Timing agreement was excellent. At $p_0=1.5$ bar, very sparse and short-lived bubbles appeared. Very little signal was detected by the PCD, likely because few bubbles existed at the PCD focus. When a two-spike signal was detected, t_C was less than 200 μs. This reduction in t_C with overpressure was in good agreement with calculations and measurements reported by Cleveland *et al.*[22] and our calculations in Fig. 1. An increase in static pressure of only 0.5 bar significantly reduced the number of bubbles seen, the measured collapse times t_C of the bubbles, and the signal detected by PCD.

An ultrasound scanner and a pressurized pipette bulb were used to investigate the reduction in bubble density by

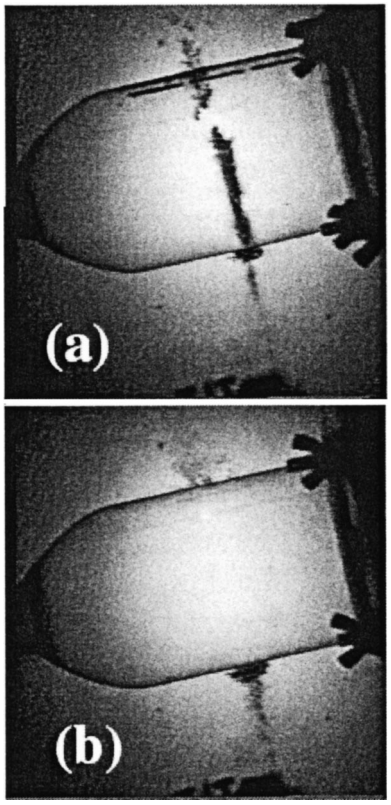


FIG. 8. High-speed images in and around a polyethylene terephthalate (PETE) bottle. Pressure in the bottle was $p_0=1$ bar (a) and 4 bar (b); outside the bottle $p_0=1$ bar in both cases. In the pressurized bottle, no bubbles were seen. Good acoustical transmission is evidenced by the presence of a cavitation cluster on both sides of the bottle. Lithotripter charging potential was 18 kV and PRF=1 Hz.

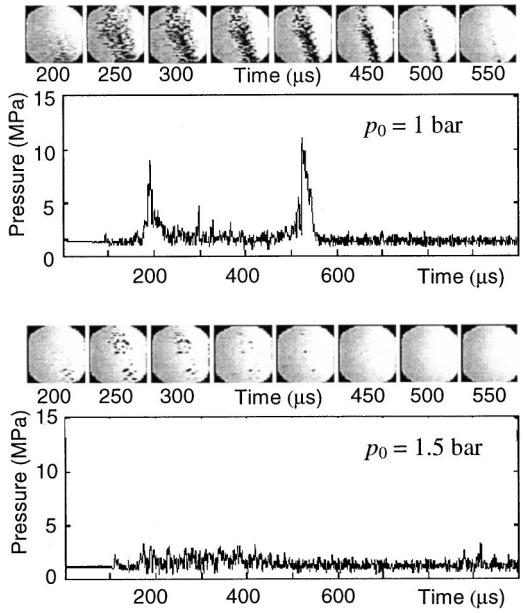


FIG. 9. Simultaneous high-speed camera images and passive cavitation detection for $p_0=1$ bar and 1.5 bar, charging potential 18 kV, and PRF $=1$ Hz. At $p_0=1$ bar, a dense cluster of bubbles grew and bubbles collapsed. The PCD spikes bracketed the growth and collapse; $t_C\approx 350$ μs. At $p_0=1.5$ bar, few bubbles were seen and little signal was detected. Bubbles and signal lasted less than 200 μs. The overpressure of only 0.5 bar dramatically reduced the number and the lifetime of the cavitation bubbles.

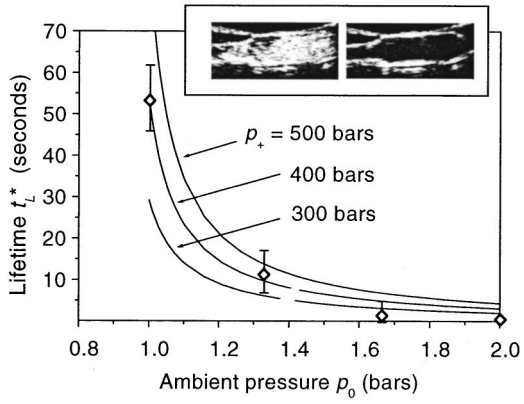


FIG. 10. Comparison of calculations and measurements of the bubble lifetime vs p_0. Measured lifetime t_L (diamonds) was determined from the presence of hyperechogenicity in *B*-mode-images of the pipette. Calculated lifetime t_L^* (solid line) is shown for shock wave amplitudes that cover the variation in the shock waves generated by this lithotripter.

overpressure. The scanner imaged from the top of the bulb and the lithotripter pulse propagated from the right in the inset images in Fig. 10.[22] Pipette bulbs were filled with nondegassed water, and single lithotripter pulses (18 kV) were fired. The ultrasound scanner was too slow to detect the t_C, but it displayed bright echoes likely from residual bubbles,[22,38,39] which were too small to be seen in the high-speed camera images after bubble collapse. The inset in Fig. 10 shows ultrasound images at $p_0=1$ bar (left) and $p_0=2$ bar (right), 1 s after pulse arrival. The echogenic cloud that filled the pipette at atmospheric pressure ($p_0=1$ bar) was subsequently seen to dissipate by $t=60$ s. At $p_0=2$ bar, no bright echoes were detected in ultrasound images of the pipette. This result indicates that overpressure either suppressed bubble growth altogether, or the lifetime of the bubbles was too short to be recorded.

Figure 10 shows a plot of bubble lifetime versus p_0. Figure 10 compares the calculated lifetime t_L^* for selected shock wave amplitudes (solid lines) with the measured lifetime t_L of hyperecho in the pipette (diamonds). An initial radius of $R_0=3$ μm was used in the calculations, but recall that the calculations were not very sensitive to initial bubble size. Peak pressures of the pulses were $p_+=384\pm 62$ bar and $p_-=100\pm 14$ bar (mean$\pm$standard error for $N=10$ measurements)[32] but could not be recorded simultaneously with the measurement. Curves were calculated for three values of p_+ to cover the range of p_+ produced by interspark variation. Lifetime t_L measured on *B*-mode ultrasound image was the time between the interference caused by spark discharge and hyperecho dissipation and was quantified on digitized video frames as described in Sec. III A. Both measured t_L and calculated t_L^* decay quickly with increasing static pressure and are in excellent qualitative agreement with each other. The dissolution time in the experiment was slightly less than in the calculations for large overpressure.

3. Pulse repetition frequency

High-speed camera images were collected for $p_0=1$, 1.5, and 2 bar, and PRF=0.5, 1, 2, and 3 Hz. Figure 11 shows the images recorded at one half of the time for the

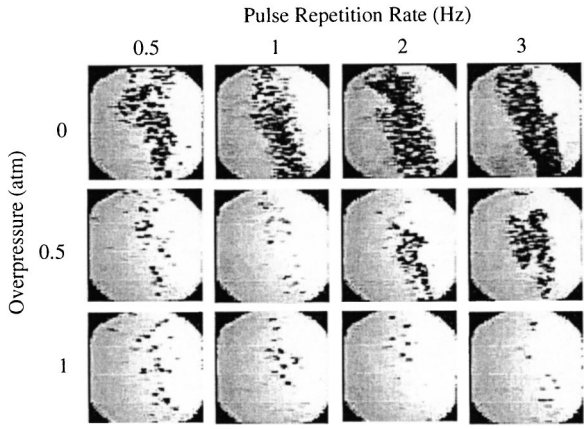


FIG. 11. High speed images of the peak bubble cloud as a function of overpressure and pulse repetition frequency (PRF). At $p_0=1$ bar (top), bubbles did not dissolve between pulses and a dense bubble cluster is apparent for all the PRFs getting more dense with increasing PRF. At $p_0=2$ bar (bottom), bubbles did dissolve between pulses and a sparse bubble cluster is apparent for all the clinical PRFs. At $p_0=1.5$ bar (center), 2 and 3 Hz were fast enough rates to create a dense cluster, but 0.5 and 1 Hz were too slow. Hence, the lifetime t_L at $p_0=1.5$ bar was greater than $(1\text{ Hz})^{-1}=1$ s but less than $(2\text{ Hz})^{-1}=0.5$ s. These images show that increased static pressure and PRF can be used together to suppress cavitation.

second collapse $t_C/2$. Charging potential was 18 kV. At zero overpressure ($p_0=1$ bar), all four rates produced a dense cluster of bubbles in water. By visual inspection, cluster density appeared to increase with increased PRF. At 0.5 bar overpressure ($p_0=1.5$ bar), bubbles were sparse at PRF $=0.5$ and 1 Hz, but dense at 2 and 3 Hz. At overpressures of 1 bar and above, the field was sparse for all the tested PRFs. This result indicates that bubble lifetime was greater than 2 s at atmospheric pressure, was in the range of 0.5–1.0 s at 1.5 bar, and was less than 0.33 s at 2 bar. These values agree well with the measured values in Fig. 10 and are slightly lower than the calculated values. The similarity indicates that the ultrasound is able to detect bubbles that cause cavitation seen with the camera, and that p_0 and PRF can be used together to suppress cavitation.

The denser clouds at higher PRF had smaller bubbles and longer t_C. In Fig. 12, a cavitation cluster created at 2 Hz (top) is compared to that created at 3 Hz (bottom). Maximum on-axis bubble diameter was 1.8 ± 0.2 mm at 1 Hz, 1.5 ± 0.3 mm at 2 Hz, and 0.8 ± 0.2 mm at 3 Hz. The largest bubble was selected from six image sequences at each frequency. The t_C was 225 ± 35 μs at 1 Hz, 336 ± 45 μs at 2 Hz, and 430 ± 40 μs at 3 Hz (six samples each). Measurement of t_C was by PCD signal recorded simultaneously with camera imaging. Collapse times determined from camera images agreed to within the interframe resolution of 50 μs.

4. Stabilized bubbles

The appearance of a few bubbles following a single lithotripter pulse or at $p_0=2$ bar and possible differences between liquids warrant investigation of stabilizing effects. Apfel,[21] Atchley and Prosperetti,[40] and Crum[41] among others have developed a ''crack and crevice model'' for how bubbles trapped within solid particles are stabilized against overpressure, surface tension, and dissolution. Bubble stabilization particularly that by cracks in a solid is pertinent to why overpressure suppresses damage to suspended cells but has little effect on kidney stone comminution.[17,22,23,42]

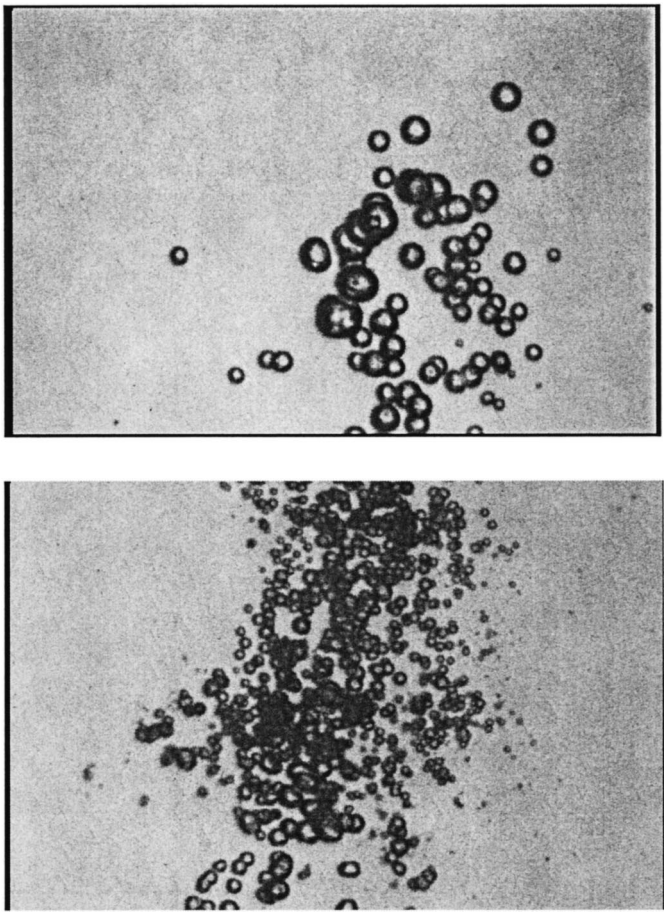


FIG. 12. Cavitation cluster created at 2 Hz (top) and 3 Hz (bottom). The width of the image is 29 mm. The denser cloud, created at higher PRF, contained smaller bubbles. The time to bubble collapse was also longer.

Stabilization of bubbles by cracks was investigated with a glass microscope slide before and after it cracked. The PETE bottle pressure chamber[33] was used, and the slide fit across the diameter of the bottle as shown in Fig. 13. The transparency of the glass enabled back lighting and visualization of the cluster of bubbles on both sides of the slide. Figure 13 shows camera images of the cluster produced at $p_0=1$ bar (top) and at $p_0=4$ bar (bottom) before (left) and after (right) a crack formed in the glass slide. Charging potential was 18 kV, and PRF was 1 Hz. At $p_0=1$ bar, a dense cluster of bubbles was seen in both cases. The cluster was slightly bulbous at the interface with the glass. The crack can be seen in the center of the bulge. At $p_0=4$ bar, no bubbles were seen on the crack-free glass (left) or in the free-field surrounding the cracked glass. But at the crack, bubbles can be seen. These data indicate that overpressure suppressed cavitation in the free field but not at the crack in this solid object. It is important to note that the crack shown in this image was caused by shock wave treatment while the slide was immersed in the water within the lithotripter tub. Thus, the crack could not have been filled *a priori* with gas. It appears, therefore, that the crack may have served to attract, then stabilize cavitation nuclei.

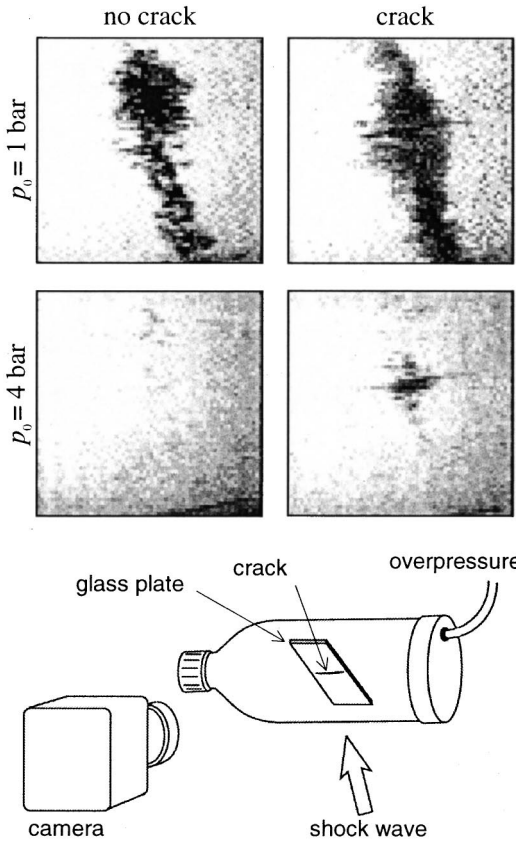


FIG. 13. Images (3.8 cm per side) of the bubble cluster produced at $p_0 = 1$ bar (top) and at $p_0 = 4$ bar (bottom) before (left) and after (right) a crack formed in the glass slide. At 4 bar the cavitation was suppressed in the free-field but was persistent at the crack. Charging potential was 18 kV, and PRF was 1 Hz.

IV. CONCLUSIONS

It is now well established that a typical clinical exposure of 2000 lithotripter pulses causes collateral damage, primarily vascular trauma, to the kidney.[2] The precise physical acoustic mechanisms responsible for this injury have yet to be determined, but there is good evidence to suggest that cavitation is involved. During SWL, cavitation has been detected *in vivo* by ultrasonography[38] and single transducer PCD.[7] In recent studies simultaneous dual-PCD and *B*-mode ultrasound followed by dual-source high intensity focused ultrasound to mark the interrogation site, were used to definitively localize cavitation bubble activity to the renal parenchyma.[43] It has also been shown that renal injury in pigs is dramatically reduced when shock waves are administered using a pressure-release insert that suppresses cavitation.[10] Taken together, these data strongly support the idea that lithotripter shock waves produce cavitation *in vivo*, and that bubble activity plays a role in tissue damage.

Cavitation is also strongly implicated as a mechanism involved in stone fragmentation in SWL. *In vitro* studies have shown that damage to stones is dramatically reduced when the stones are treated while immersed in viscous media[5] or when shock waves are administered at overpressure sufficient to eliminate cavitation.[41]

If, as such studies suggest, cavitation contributes both to stone fragmentation and to tissue damage, strategies to control bubble activity could be very useful, especially if it were possible to find conditions that inhibit kidney trauma without interfering with stone comminution. Our observations help to explain how overpressure and PRF influence cavitation. We show that cavitation is dramatically affected by the presence of nuclei from previous shock waves, and that increased overpressure and reduced PRF allow those cavitation bubbles that are not stabilized by interactions with solid surfaces to dissolve between pulses.

Numerically, the time history of a bubble subjected to a lithotripter pulse was modeled from a synthesis of models by Church[24] and by Epstein and Plesset.[26] Rapid growth and collapse of the bubble was calculated with a numerical solution[24] of the Gilmore equation.[25] In this time of rapid motion of the air–water interface gas diffusion was approximated by Eller and Flynn's model[30] in which the thickness of the diffusion layer of the gas in the fluid is very small. At longer time, the thickness of the diffusion layer of gas could not be neglected, but the bubble radial velocity could. Then the equation by Epstein and Plesset[26] was used. The calculated lifetime t_L^* of a bubble was found to be fairly insensitive to initial bubble radius R_0 in the range 1–10 μm, because gas influx was much greater than initial gas constant of the bubble. Lifetime t_L^* was sensitive to static pressure p_0 and the peak acoustic driving pressure p_+, because in our wave form increased p_+ meant increased peak negative pressure p_-. Other factors are gas concentration in solution f and the diffusion coefficient D that depends on temperature and type of gas and liquid. These factors vary among *in vitro* and *in vivo* samples, and the effect of tissue constraint on our model has been described by Zhong *et al.*[9] Calculated bubble lifetime t_L^* and the duration of the growth and collapse cycle t_C^* are shortened by increasing static pressure p_0.

Lifetimes t_L for bubbles created by lithotripter pulses were measured by *B*-mode ultrasound and by high-speed camera. Lithotripter pulses created hyperechogenic regions in water within pipette bulbs. The disappearance of the hyperechogenic region was measured versus p_0 and compared very well to the calculated t_L^*. The lifetime of hyperechogenic regions at $p_0 = 1$ bar has previously been shown to be comparable *in vivo* and *in vitro*;[38,43] although in some *in vivo* locations, hyperecho is washed out of the region of interest by blood perfusion before bubbles dissolve.[33] High-speed camera images indicated that a dense cavitation cluster appeared when the time between shock waves (1/PRF) was less than the bubble lifetime (t_L). Measured t_L and calculated t_L^* were found to agree very well and both showed a sharp decay with increased static pressure p_0. However, measured lifetimes were reduced more strongly by overpressure than was predicted by calculated t_L^*. Our model assumed symmetric pulsation of a single bubble and did not account for bubble fragmentation. Thus the presence of a cluster of small bubbles was neglected. The equilibrium bubble radius may have been smaller because of asymmetry in collapse and splintering of one bubble into many daughter bubbles. Measurements with *B*-mode ultrasound showed that the hyperechoic region was brighter, larger, and denser with subsequent lithotripter pulses and with increasing PRF. Larger bubbles are more likely to collapse nonspherically

and break into smaller bubbles, which then grow. An increase in the number of bubbles after each shock wave would contribute to the brighter, larger, denser cluster. The very presence of several other bubbles may reduce both the gas diffusion into each bubble, as several bubbles compete for the gas dissolved in the fluid, and the pressure and inertia driving the growth and collapse of each bubble, as the cluster itself changes the compressibility of the fluid bubble mixture. Our measurements with the high-speed camera indicated that clusters were denser, bubbles were smaller, and collapse times longer at higher PRF. The measurements suggest that reduction of gas diffusion into each bubble, reduction of fluid outward motion, and reduction of pressure driving bubble collapse were the dominant effects of increased cluster density.

High-speed camera images showed that bubbles in the free field were considerably more sensitive to overpressure than those that formed along a crack in the glass slide used as a target. Thus, our observations support the concept[21,40,41] that bubbles appear to be stabilized by contact with cracks and crevices. Such stabilization of bubbles could explain why modest overpressure reduces damage to cells in suspension, but minimally suppresses stone fragmentation.[17]

Our observations imply that control of PRF and overpressure could be used to improve SWL in the clinical setting. In conventional lithotripsy treatment, low PRF may be used to allow bubble dissolution between pulses and thus to suppress cavitation *in vivo*. However, the duration of a lithotripsy procedure is limited by the rate of lithotripter pulse delivery. A treatment of 2000 pulses at 1 Hz requires over 30 min. Modest overpressure (1–3 bar) accelerates the dissolution of unstabilized bubbles such as those that can develop in the vasculature. Thus, it should be possible to use faster treatment rates without an increase in tissue damage. Since irregularities and cracks in the surface of kidney stones have the potential to stabilize bubbles against overpressure, it seems feasible that application of modest overpressure could reduce cavitation-mediated damage to tissue without interfering with the role that bubble activity plays in stone comminution.

ACKNOWLEDGMENTS

We thank Dr. James E. Lingeman (Methodist-Clarian Hospital) and Dr. Andrew P. Evan [Indiana University (IU) Medical School] for use of the Dornier HM-3 lithotripter at Methodist Hospital in Indianapolis. We thank Dr. Peter Kaczkowski [Applied Physics Lab (APL)] and Francis Olson (APL) for design and manufacture of the ''pressure-bottle'' pressure chamber, David Lounsbery (IU) for manufacture of the ''pressure cube,'' and Semih Sinik (Itronics Imaging Technologies, Westlake Village, CA) and Frank Kosel (DRS Hadland, Inc., Cupertino, CA) for technical advice with the high-speed camera. This work was supported by NIH Grant Nos. P01 DK43881, R01 DK55674, NIH Fogarty FIRCA, NATO, and CRDF.

[1] C. Renner and J. Rassweiler, ''Treatment of renal stones by extracorporeal shock wave lithotripsy,'' Nephron **81**, 71–81 (1999).
[2] A. P. Evan, L. R. Willis, J. E. Lingeman, and J. A. McAteer, ''Renal trauma and the risk of long-term complications in shock wave lithotripsy,'' Nephron **78**, 1–8 (1998).
[3] W. Sass, M. Braunlich, H. Dreyer, E. Matura, W. Folberth, H. Priesmeyer, and J. Seifert, ''The mechanisms of stone disintegration by shock waves,'' Ultrasound Med. Biol. **17**, 239–243 (1991).
[4] N. G. Holmer, L. O. Almquist, T. G. Hertz, A. Holm, E. Lindstedt, H. W. Persson, and C. H. Hertz, ''On the mechanism of kidney stone disintegration by acoustic shock waves,'' Ultrasound Med. Biol. **17**, 479–489 (1991).
[5] N. Vakil, S. M. Gracewski, and E. C. Everbach, ''Relationship of model stone properties to fragmentation mechanisms during lithotripsy,'' J. Litho. Stone Dis. **3**, 304–310 (1991).
[6] M. Kuwahara, N. Ioritani, K. Kamke, S. Shirai, K. Taguchi, T. Saitoh, S. Orikasa, K. Takayama, S. Aida, and N. Iwama, ''Hyperechoic region induced by focused shock waves *in vitro* and *in vivo*: Possibility of acoustic cavitation bubbles,'' J. Litho. Stone Dis. **1**, 218–228 (1989).
[7] A. J. Coleman, M. J. Choi, and J. E. Saunders, ''Detection of acoustic emission from cavitation in tissue during extracorporeal lithotripsy,'' Ultrasound Med. Biol. **22**, 1079–1087 (1996).
[8] D. Dalecki, C. H. Raeman, S. Z. Child, D. P. Penney, R. Mayer, and E. L. Carstensen, ''The influence of contrast agents on hemorrhage produced by lithotripter fields,'' Ultrasound Med. Biol. **23**, 1435–1439 (1997).
[9] P. Zhong, Y. Zhou, and S. Zhu, ''Dynamics of bubble oscillation in constrained media and mechanisms of vessel rupture in SWL,'' Ultrasound Med. Biol. **27**, 119–134 (2001).
[10] A. P. Evan, L. R. Willis, B. A. Connors, Y. Shao, J. E. Lingeman, J. C. J. Williams, J. A. McAteer, N. S. Fineberg, M. R. Bailey, and L. A. Crum, ''Kidney damage and renal functional changes are minimized by waveform control that suppresses cavitation in SWL,'' J. Urol. (in press).
[11] M. R. Bailey, D. T. Blackstock, R. O. Cleveland, and L. A. Crum, ''Comparison of electrohydraulic lithotripters with rigid and pressure-release ellipsoidal reflectors. II. Cavitation fields,'' J. Acoust. Soc. Am. **106**, 1149–1160 (1999).
[12] M. R. Bailey, ''Control of acoustic cavitation with application to lithotripsy,'' (Technical Report No. ARL-TR-97-1, Applied Research Laboratories, The University of Texas at Austin, Austin, TX, 1997, pp. 1–210.
[13] D. L. Sokolov, M. R. Bailey, and L. A. Crum, ''Use of a dual-pulse lithotripter to generate a localized and intensified cavitation field,'' J. Acoust. Soc. Am. **110**, 1685–1695 (2001).
[14] D. L. Sokolov, M. R. Bailey, F. Pulvermakher, and L. A. Crum, ''Increased damage to stones without increased damage to cells with a dual-reflector lithotripter,'' in Proceedings of the IEEE International Ultasonics Symposium, San Juan, Puerto Rico, 2000, Vol. 2, pp. 1437–1440.
[15] X. F. Xi and P. Zhong, ''Improvement of stone fragmentation during shock wave lithotripsy using a combined EH/PEAA shock wave generator—*in vitro* experiments,'' Ultrasound Med. Biol. **26**, 457–467 (2000).
[16] P. Zhong and Y. Zhou, ''Suppression of large intraluminal bubble expansion in shock wave lithotripsy without compromising stone comminution: Methodology and *in vitro* experiments,'' J. Acoust. Soc. Am. **110**, 3283–3291 (2001).
[17] M. Delius, ''Minimal static excess pressure minimises the effect of extracorporeal shock waves on cells and reduces it on gallstones,'' Ultrasound Med. Biol. **23**, 611–617 (1997).
[18] P. Huber, K. Jochle, and J. Debus, ''Influence of shock wave pressure amplitude and pulse repetition frequency on the lifespan, size and number of transient cavities in the field of an electromagnetic lithotripter,'' Phys. Med. Biol. **43**(10), 3113–3128 (1998).
[19] R. F. Paterson, D. A. Lifshitz, J. E. Lingeman, J. C. Williams, D. L. Rietjens, A. P. Evan, B. A. Connors, M. R. Bailey, L. A. Crum, R. O. Cleveland, Y. A. Pishchalnikova, I. V. Pishchalnikova, and J. A. McAteer, ''Slowing the pulse repetition frequency in shock wave lithotripsy (SWL) improves stone fragmentation *in vivo*,'' in Proceedings of the 17th International Congress on Acoustics, Rome, Italy, 2001, pp. 200–201.
[20] M. Delius, W. Mueller, A. Goetz, H.-G. Liebich, and W. Brendel, ''Biological effects of shock waves: Kidney hemorrhage in dogs at a fast shock wave administration rate of fifteen hertz,'' J. Litho. Stone Dis. **2**, 103–110 (1990).
[21] R. E. Apfel, ''The role of impurities in cavitation-threshold determination,'' J. Acoust. Soc. Am. **48**, 1179–1186 (1970).
[22] R. O. Cleveland, M. R. Bailey, L. A. Crum, M. A. Stonehill, J. C. Williams, Jr., and J. A. McAteer, ''Effect of overpressure on dissolution and cavitation of bubbles stabilized on a metal surface,'' in Proceedings of the 16th International Congress on Acoustics, Seattle, Washington, 1998, Vol. 3, pp. 2499–2500.

[23] M. R. Bailey, R. O. Cleveland, O. A. Sapozhnikov, J. A. McAteer, J. C. Williams, Jr., and L. A. Crum, ''Effect of increased ambient pressure on lithotripsy-induced cavitation in bulk fluid and at solid surfaces,'' in *Collected Papers from the Joint Meeting 'Berlin 99:' 137th Meeting of the Acoustical Society of America; 2nd convention of the European Acoustics Association; 25th German Acoustics DAGA conference*, edited by M. Moser and J. Tichy (Deutsche Gesellschaft fur Akustik (DEGA), Berlin, 1999), pp. 1–4 (CDROM only—no true page numbers).
[24] C. C. Church, ''A theoretical study of cavitation generated by an extracorporeal shock wave lithotripter,'' J. Acoust. Soc. Am. **86**, 215–227 (1989).
[25] F. R. Gilmore, ''The growth or collapse of a spherical bubble in a viscous compressible liquid,'' California Institute of Technology Report No. 26-4, 1952, pp. 1–40.
[26] P. S. Epstein and M. S. Plesset, ''On the stability of gas bubbles in liquid-gas solutions,'' J. Chem. Phys. **18**, 1505–1509 (1950).
[27] A. J. Coleman and J. E. Saunders, ''A review of the physical properties and biological effects of the high amplitude acoustic fields used in extracorporeal lithotripsy,'' Ultrasonics **31**, 75–89 (1993).
[28] S. Zhu and P. Zhong, ''Shock wave—inertial microbubble interaction: A theoretical study based on the Gilmore formulation for bubble dynamics,'' J. Acoust. Soc. Am. **106**, 3024–3032 (1999).
[29] 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).
[30] A. Eller and H. G. Flynn, ''Rectified diffusion during nonlinear pulsations of cavitation bubbles,'' J. Acoust. Soc. Am. **37**, 493–503 (1965).
[31] W. Press, S. A. Teukolsky, W. T. Vetterling, and B. R. Flannery, *Numerical Recipes in FORTRAN*, 2nd ed. (Cambridge University Press, Cambridge, 1992), pp. 704–716.
[32] R. O. Cleveland, M. R. Bailey, N. Fineberg, B. Hartenbaum, M. Lokhandwalla, J. A. McAteer, and B. Sturtevant, ''Design and characterization of a research electrohydraulic lithotripter patterned after the Dornier HM3,'' Rev. Sci. Instrum. **71**, 2514–2525 (2000).
[33] M. R. Bailey, L. N. Couret, O. A. Sapozhnikov, V. A. Khokhlova, G. ter Haar, S. Vaezy, X. Shi, R. Martin, and L. A. Crum, ''Use of overpressure to assess the role of bubbles in focused ultrasound lesion shape *in vitro*,'' Ultrasound Med. Biol. **27**, 696–708 (2000).
[34] M. A. Stonehill, J. C. Williams, Jr., M. R. Bailey, D. Lounsbery, R. O. Cleveland, L. A. Crum, A. P. Evan, and J. A. McAteer, ''An acoustically matched high pressure chamber for control of cavitation in shock wave lithotripsy: Mechanisms of shock wave damage *in vitro*,'' Methods Cell Sci. **19**, 303–310 (1998).
[35] R. O. Cleveland, O. A. Sapozhnikov, M. R. Bailey, and L. A. Crum, ''A dual passive cavitation detector for localized detection of lithotripsy-induced cavitation *in vitro*,'' J. Acoust. Soc. Am. **107**, 1745–1758 (2000).
[36] A. J. Coleman, M. J. Choi, J. E. Saunders, and T. G. Leighton, ''Acoustic emission and sonoluminescence due to cavitation at the beam focus of an electrohydraulic shock wave lithotripter,'' Ultrasound Med. Biol. **18**, 267–281 (1992).
[37] K. Jochle, J. Debus, W. J. Lorenz, and P. Huber, ''A new method of quantitative cavitation assessment in the field of a lithotripter,'' Ultrasound Med. Biol. **22**, 329–338 (1996).
[38] M. Delius, and S. Gambihler, ''Sonographic imaging of extracorporeal shock wave effects in the liver and gallbladder of dogs,'' Digestion **52**, 55–60 (1992).
[39] R. K. Zeman, W. J. Davros, B. S. Garra, J. A. Goldberg, S. C. Horii, P. M. Silverman, E. L. Cattau, W. S. Hayes, and C. J. Cooper, ''Cavitation effects during lithotripsy. II. Clinical observations,'' Radiology **177**, 163–166 (1990).
[40] A. A. Atchley and A. Prosperetti, ''The crevice model of bubble nucleation,'' J. Acoust. Soc. Am. **86**, 1065–1084 (1989).
[41] L. A. Crum, ''Nucleation and stabilization of microbubbles in liquids,'' Appl. Sci. Res. **38**, 101–115 (1982).
[42] J. A. McAteer, M. A. Stonehill, K. Colmenares, J. C. Williams, A. P. Evan, R. O. Cleveland, M. R. Bailey, and L. A. Crum, ''SWL cavitation damage *in vitro*: Pressurization unmasks a differential response of foil targets and isolated cells,'' Proceedings of the 16th International Congress on Acoustics and 135th Meeting of the Acoustical Society of America 1998, Vol. 3, pp. 2497–2498.
[43] M. R. Bailey, L. A. Crum, N. Miller, O. A. Sapozhnikov, Y. A. Pishchalnikov, R. O. Cleveland, J. A. McAteer, P. Blomgren, B. A. Connors, and A. P. Evan, ''Localized cavitation detection in lithotripsy *in vivo*,'' in Proceedings of the 17th International Congress on Acoustics, Rome, Italy, 2001, pp. 178–179.

A novel cavitation probe design and some preliminary measurements of its application to megasonic cleaning

Gary W. Ferrell
SEZ America, Inc., 2632 Bayshore Parkway, Mountain View, California 94043

Lawrence A. Crum[a)]
Center for Industrial and Medical Ultrasound, Applied Physics Laboratory, 1013 NE 40th Street, Seattle, Washington 98105

(Received 5 September 2001; revised 28 May 2002; accepted 12 June 2002)

Cavitation is an effective physical mechanism for concentrating mechanical energy. Accordingly, it has wide applications in such diverse fields as sonochemistry, in which chemical reactions are initiated or accelerated, or in the electronic component cleaning industry in which particles (and other materials) are removed from surfaces. However, devices designed to act as cavitation monitors have had little success, partly because their intrusiveness often affects the cavitation field itself. Presented here is a brief description of a unique cavitation monitor that utilizes the phenomenon of sonoluminescence as an indirect quantifier of cavitation. It appears to work efficiently over a broad range of acoustic field intensities and its application to megasonic cleaning has provided interesting and valuable insights into this technology. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1498856]

PACS numbers: 43.80.Nd, 43.25.Yw [AJS]

I. INTRODUCTION AND SOME BACKGROUND

As feature sizes of semiconductor devices continue to shrink, the requirements to control particulate contamination in the submicron size range on product wafer surfaces becomes extremely critical. Minimizing ''killer particles,'' that is, particles on the order of one-tenth of the line width of silicon devices, means that submicron-size particles should be removed during wet processing. For example, the killer particle size is 0.04 μm for a 64-megabit silicon device. Although high-velocity spray cleaning techniques have been shown to remove larger size particles from the surface, they do not remove particles in the submicron size range because of the formation of boundary layers on the wafer surface. Boundary layer effects have been shown to disappear, or be overcome, when a megasonic wafer cleaning technique is used with both SC-1 (ammonium hydroxide/hydrogen peroxide/deionized water) and SC-2 (hydrogen chloride/hydrogen peroxide/deionized water) solutions (Syverson *et al.*, 1992). Thus, megasonic cleaning is becoming a widely used technique for removing particles from wafer surfaces; however, this technique is still poorly understood and is difficult to monitor.

There appears to be two principal classes of mechanisms through which megasonic cleaners could accomplish cleaning: (a) through a direct action of the sound field with the particle, and (b) through an indirect action, viz., acoustic cavitation. Consider first a direct interaction of the sound field with the attached particle, i.e., the oscillating acoustic field exerts periodic forces directly on particulate matter attached to a boundary or surface. These oscillating forces might eventually overcome the attractive attachment forces and free the particle. In an acoustic field, a small particle of the fluid will oscillate back and forth, exerting a drag force on obstacles in the flow. However, acoustic particle velocities (at megasonic cleaning intensities) are of insufficient displacement to result in significant particle motion. For example, at a frequency of 1 MHz, and an acoustic intensity of 1 W/cm^2, the acoustic particle displacement in water is on the order of a few Angstroms (Kinsler and Frey, 1962). Obviously, this displacement is insufficient to remove particles with sizes on the order of a few hundredths of a micron. Some authors contend that radiation pressure forces, sometimes called primary Bjerknes forces (Crum, 1975), could also result in particle removal. According to Menon (1990), these forces are principally responsible for megasonic cleaning. However, recent papers by Olson (1988) and by Geers and Hasheminejad (1991) define the conditions necessary for optimal particle removal and their theories '' . . . suggest that very high frequencies ($\sim$1 GHz) will be required to remove submicron contaminants from wafers'' (Olson, 1988).

A more plausible mechanism for a submicron particle removal is acoustic cavitation. This modality has three different specific ways in which particles can be removed. The first mechanism could be described as acoustic microstreaming. In this case, a microscopic air bubble preexisting in the liquid (or nucleated on a solid surface) undergoes stable, large-amplitude pulsations which in turn causes rapid movement of the liquid as it follows the oscillating bubble boundary. If any asymmetry exists in the flow pattern (brought about by, say, a nearby boundary), intense microstreaming patterns develop which can lead to significant shear stresses imposed along the boundary. This microstreaming phenomenon is normally associated with the presence of *stable* cavitation activity, which generally occurs when the cavitating liquid is saturated with gas. However, microstreaming must also occur for more vigorous cavitation as it is related to the volume oscillations of cavitation bubbles. Acoustic micros-

[a)]Electronic mail: lac@apl.washington.edu

treaming is a well-known phenomenon and quantitative approaches exist for evaluating the forces available for particle removal (Elder, 1959; Nyborg, 1965; Kashkoush and Busnaina, 1993). Other mechanisms such as Schlicting and Eckart streaming may be involved, but their role is considered to be minor.

A second cavitation-related, potential source of particle removal is through a phenomenon that shall be called microjet impact. If an oscillating bubble is undergoing relatively large displacement excursions, then the bubble-wall collapse velocities can also be quite large and the motion is inherently unstable. In this case, the bubble is said to be inertially controlled and any asymmetry in the flow field around the bubble normally results in an asymmetric (nonspherical) bubble collapse. Given this scenario, one wall tends to collapse faster than the other, resulting in the development of a microscopic liquid jet that propels itself through the bubble, penetrates the opposite wall, and violently impacts the very boundary that caused the asymmetry to develop in the first place. This liquid jet, which can attain supersonic velocities, is often thought to be the principal mechanism for cavitation erosion damage. Jet formulation is normally associated with the presence of *inertial* cavitation, which generally occurs when the cavitating liquid is partially degassed and the acoustic pressure amplitudes are relatively large. For more information on this phenomenon, see, for example, papers by Naude and Ellis (1961), Lauterborn and Bolle (1975) and Crum (1979,1988). Incidentally, ultrasonic cleaners operating at kilohertz frequencies are not suitable for wafer surface cleaning because these cavitation jets can damage the substrate; however, when megahertz frequencies are used (in gassy liquids such as SC-1 and SC-2), the jet impacts are apparently not sufficiently violent to induce this undesirable damage.

A third cavitation-related mechanism is a result of the violent implosion of the collapsing gas bubble. When the bubble implodes, it can create a shock wave in the liquid surrounding the bubble. The local pressures in this shock wave are known to be as high as a few kilobars—pressures sufficient to damage the surfaces of metals, and thus, a potential particle removal mechanism. If the impinging shock wave can damage the surface, it can clearly loosen lightly attached contamination. Although most cavitation researchers believe jet impacts are the principal mechanism for cavitation damage, it is not possible to eliminate shock wave effects as a possible damage source, and thus also capable of particle removal. However, shock waves are typically produced by inertial rather than stable cavitation, and are therefore more likely to be present when the cavitating liquid is moderately degassed or when the acoustic pressure amplitude is very large, or when the acoustic frequency is in the kilohertz range—conditions that are not normally met in commercial megasonic cleaners. Accordingly, we conclude that cavitation-generated shock waves do not play a major role in megasonic cleaner particle removal. Studies of shock-wave-associated acoustic cavitation damage have been performed by Ellis (1966) and Tomita *et al.* (1986).

A phenomenon that is associated with cavitation is that of sonoluminescence (SL), in which the collapsing cavitation bubble heats the gas contained within the bubble to incandescent temperatures. Although SL is generally associated with inertial, rather than stable cavitation, it is commonly known that ultrasonic cleaners, even megasonic cleaners, generate high levels of SL. Indeed, using luminol, a wave-shifting chemical, one can easily see with the naked eye the SL generated in an ultrasonic cleaner and detect areas of nonuniformity of the acoustic field.

To briefly summarize, particle removal from silicon wafer surfaces is most probably associated with cavitation. At the acoustic intensities used in ultrasonic cleaners, particularly megasonic cleaners, this cavitation produces SL, which can be detected with optical detectors. Thus, an instrument that could detect the SL produced by these cleaners can serve as a real-time, albeit indirect, monitor of the effectiveness of these cleaners.

II. APPARATUS AND SOME RESULTS

In the semiconductor industry, silicon wafers undergo many cleaning stages, a number of which utilize megasonic cleaning devices. Cavitation is a difficult phenomenon to monitor; in a real sense, it is a stochastic phenomenon, resulting from the serendipitous spatial encounter of a preexisting nucleation site with exactly the right temporal acoustic conditions. Nevertheless, its presence is essential to cleaning.

When a number of 200-mm-diam wafers are cleaned in a megasonic cleaning tank, they are typically placed in the tank for approximately 20 min, along with a test wafer. The wafers are either mechanically moved back and forth to ensure complete acoustic/cavitation field coverage, or, more recently, the acoustic field is itself spatially ''moved'' by an appropriate temporal variation of the voltage to the various transducers lining the cleaning tank. Within the tank also is a test wafer, which is removed and scanned with a microparticle detector to ensure that the necessary level of cleaning is obtained. However, there is little if any effective real-time monitoring of the cleaning efficiency or effectiveness of the megasonic system. There are a number of events that can occur within a megasonic cleaner that could result in unacceptable cleaning:

(i) One of the several transducers that line the tank can become less efficient, resulting in a ''cold'' spot or a region of ineffective cleaning; although some devices have transducer monitoring electronics, a minor change in transducer efficiency can result in such a cold spot—because cavitation is a threshold phenomenon and is nonlinearly related to the transducer's electrical characteristics.

(ii) The cleaning solution can become ineffective for supporting cavitation. It is well known that the velocity of sound in a liquid is strongly dependent on the compressibility of the liquid, which is in turn extremely sensitive to the amount of gas bubbles present. Indeed, a ''volume fraction'' of gas on the order of 10^{-4} can result in significant changes in the sound velocity. Thus, heating of the cleaning liquid or inadequate de-

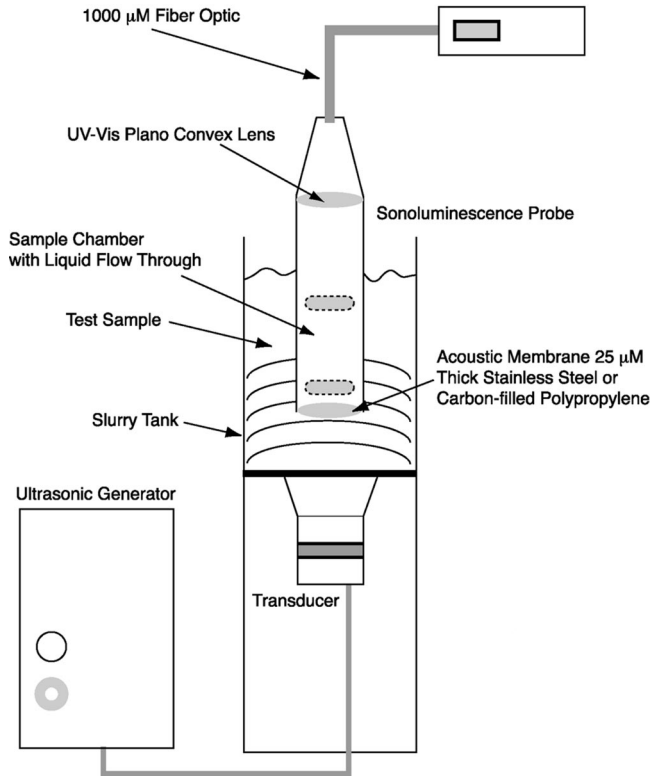


FIG. 1. Drawing of our design for a cavitation probe, and illustration of its use in a megasonic cleaner.

gassing can result in a significant loss of cleaning efficiency. It is very difficult to monitor this effect in existing megasonic cleaners.

(iii) There is no real-time monitor of the extent of cleaning. The 20-min cleaning time was obtained through empirical studies. If the conditions are favorable, complete cleaning could occur within half that time; alternatively, if they are unfavorable, and the test wafer is not adequately clean, a full 20-min recleaning is often performed. Thus, this system makes poor use of cleaning efficiency.

These and a number of other failure modes of megasonic cleaners could be eliminated if a real-time cavitation monitoring device were available to measure the effectiveness of the megasonic cleaning process. We have designed a unique device that is capable of detecting the presence of sonoluminescence with considerable spatial resolution, and indirectly through this detection of sonoluminescence, to detect the spatial and temporal existence of cavitation itself. Our initial design for a cavitation detector is shown in Fig. 1, and a photograph of one of our prototype models is shown in Fig. 2. This design if very flexible and permits the cavitation field to be measured and recorded in real time by a data acquisition system. The design is also quite simple and straightforward.

The central feature of the probe is a small closed cell which is made of metal or some other opaque material to prevent contamination by photons that originate from any other region than the test volume. An ''acoustic window'' that permits the acoustic field to enter the cell is accomplished by utilizing a 25-μm-thick stainless steel film. This film is completely opaque to background photons but is nearly transparent to the acoustic field. It is also thick enough to survive any cavitation damage that might be produced by the megasonic cleaner.

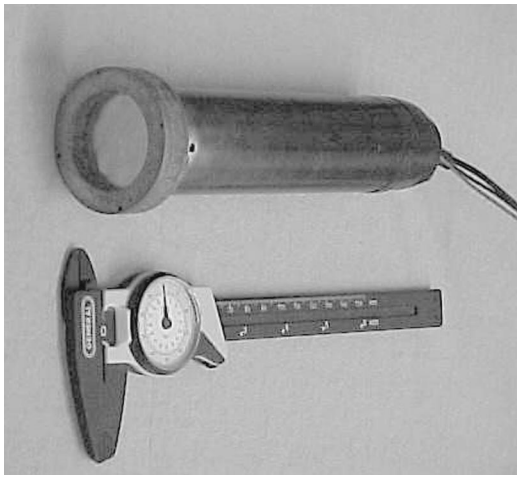


FIG. 2. Photograph of the cavitation probe. The lower dimensions on the ruler are inches; the upper ones are metric.

A second important feature of our design is that it measures the cavitation produced in the fluid of interest, not some unrelated artificial fluid. This effect is accomplished by constructing a tortuous path for the fluid to enter the test region, but one such that external photons can not. We have determined that this path can not be a simple one, but one of several turns; otherwise, even if the reflection coefficient is quite low, the ambient background has such a high density of visible photons that ''leak through'' can be a major problem. We have constructed this path by trial and error, with the result that our background count is on the order of 50 counts/second, while typical count levels in the region of intense cavitation can be on the order of a few hundred thousand.

The test region is viewed by a photomultiplier tube (PMT) that is sensitive to a range between 270 and 650 nm, although we are aware of the water cutoff that occurs below about 300 nm. Another feature of our design is to use a plano-convex lens with a focal length of 30 mm to direct as many photons as possible onto the collecting area of the detector, although with our sensitive PMT, we still have a large dynamic range in terms of the number of SL counts above background. Incidentally, we find that a photo-diode, although more robust and must less expensive, does not provide sufficient sensitivity for this operation. We have learned also that we can use an optical fiber to physically separate the PMT from the detector itself. Using an optical fiber permits us to construct the cell small enough to fit in the space between two wafers in a megasonic cleaner, for example, thus permitting us to examine quite directly the actual cavitation field experienced by the silicon wafers themselves. Without the optical fiber, the diameter of the sensitive area of the probe is on the order of 1 cm, which is determined by the size of the PMT. However, as will be shown later in the data presented (see Fig. 5, for example), spatial discrimination of the SL flux can be determined with much higher precision.

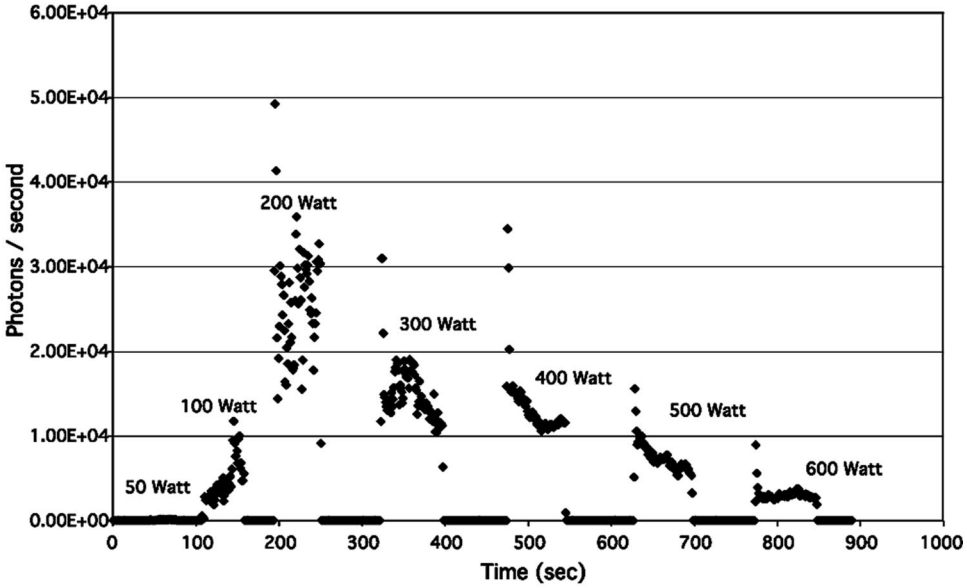


FIG. 3. In this experiment, the cavitation probe was positioned near the center of the region where a wafer would be placed to be cleaned. The electrical power supplied to the transducers that generate the acoustic field was increased from a value of 50 W to a value of 600 W in a sequential manner. The various power levels are indicated on the figure. The number of photons counted by the probe during this entire sequence is plotted on the vertical scale. Note that increasing the electrical power to the transducers beyond a saturation level has a deleterious effect on the amount of cavitation generated and, thus, presumably, on the cleaning efficiency.

In Figs. 3 and 4 we provide some initial data on the successful operation of our prototype probe, which demonstrates clearly how it can be of considerably value in determining the temporal and spatial characteristics of the cavitation field of a megasonic cleaner. In Fig. 3, we show the results of a study of the effect of increased transducer voltage on the cavitation field of a commercial megasonic cleaner. The data in Fig. 3 have important implications in the megasonic cleaning community. Until these data were obtained—and as far as we know, they are unique—it was our natural presumption that megasonic cleaners would work more effectively at electrical powers in excess of 300 W; indeed, some manufacturers advertise that their systems have the capability of operating at kilowatt power levels. We believe that this effect of reduced effectiveness at high powers is NOT due to limitations in the electrical or transduction systems, but is a cavitating fluid effect. Specifically, if there is TOO much cavitation, the increased volume fraction of gas in the liquid prevents the acoustic field from propagating. Such an effect has been observed by the sonochemistry community. It is not unreasonable to expect, and anecdotal evidence supports our view, that there are times when megasonic cleaner operators find that their system is not cleaning effectively, and subsequently increase the power beyond the saturation level, which continues to reduce their cleaning effectiveness. The use of a cavitation probe would prevent this effect from occurring.

Figure 4 demonstrates another interesting discovery that we made while testing our probe. It is well known that preexisting gas-filled nuclei are necessary for cavitation incep-

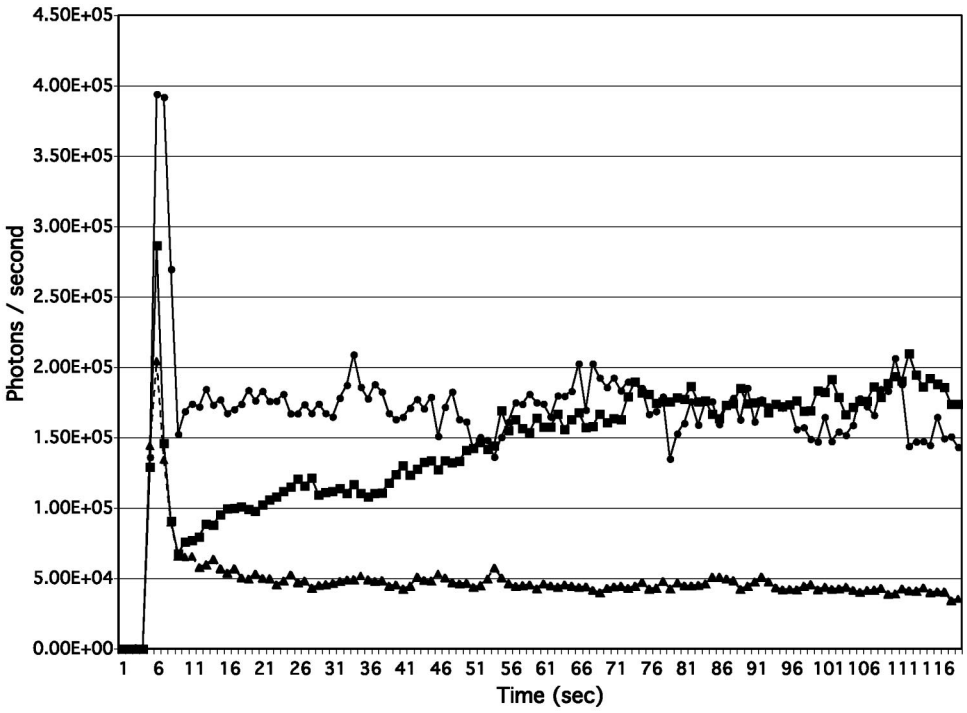


FIG. 4. In this experiment, the role of preexisting nucleation sites on the level of cavitation activity was measured. With the solid triangles, we show a case in which the liquid was permitted to remain undisturbed for several minutes, and then the cleaner was engaged. Note that the level of cavitation activity is relatively low and that there is a gradual reduction in activity over the time course of the experiment. With the solid circles, we show a case in which air was bubbled through the cleaner liquid during operation. Note that the level of activity is more than three times higher. With the solid squares, we show a case in which the liquid was permitted to remained undisturbed for several minutes, but when the cleaner was engaged, so also was the bubbler. Note the gradual increase in the level of cavitation activity. We are unsure if the transient spike in activity is electrical or cavitation in origin.

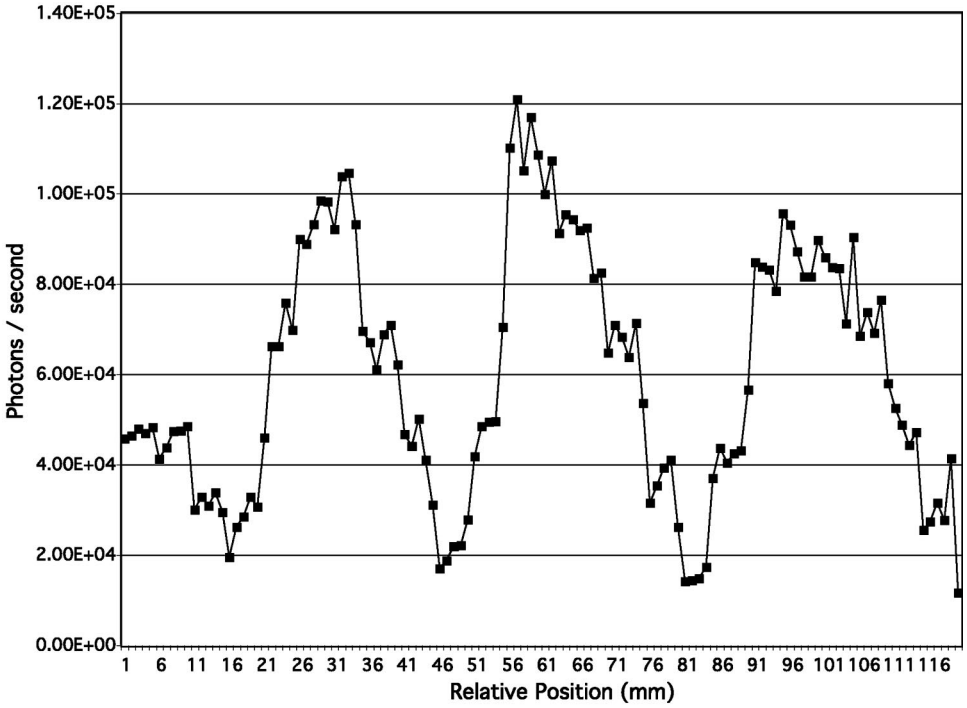


FIG. 5. Megasonic cleaners are expected to produce a uniform cavitation field so that wafer will be cleaned thoroughly and completely. We show the results of scanning the field of a commercial megasonic cleaner that utilizes four transducers. In this case, geometrical constraints prevented us from measuring the entire field produced by all four transducers; however, by scanning over a region of three of the transducers, it is seen that ''cold spots,'' or areas of reduced cavitation, are produced within the cleaning bath.

tion; i.e., a completely homogeneous liquid has an enormously high cavitation threshold. In these experiments, we examined the SL flux depending on (a) the initial conditions and (b) the presence of a source of gas nuclei. When we allowed the liquid to remain undisturbed within the megasonic cleaner for 10 min, and then measured the SL activity as the cleaner was engaged, we observed that after an initial spike, the level of cavitation activity was quite low, shown as the solid triangles in the figure, and slowly, but perceptably, showed decreasing activity over a period of about 2 min. We presume that we were gradually expending the preexisting nucleation sites. If, on the other hand, we continued to supply nucleation sites, by bubbling air through the liquid, both before and during cleaner operation, we observed a similar behavior as before, shown by the solid circles in the figure, except that in this case the level of cavitation activity was more than three times higher. Finally, we allowed the system to remain undisturbed for several minutes, then we engaged both the cleaner and the bubbler. As shown in Fig. 4, as the solid squares, the level of cavitation activity gradually increased from the lower value, observed without the bubbler, to the higher value, observed when the bubbler was active. One could make the simple observation that it would be better to bubble gas through the megasonic cleaner while it is in a cleaning mode.

In Fig. 4, we also noticed that within a few seconds after the megasonic cleaner was engaged, a large spike of cavitation activity occurred. Because we were unable to measure the electrical drive to the cleaner transducers, we are unsure if this transient is an electrical one, or a cavitation-related one.

Finally, we show in Fig. 5 the use of the cavitation probe to determine the homogeneity of the cavitation field produced by a typical, commercial megasonic cleaner. In this case, we examined a cleaner that used four transducers to generate four bands of cavitation. In various modes of operation, either all of the transducers are engaged, any one of the transducers are engaged, or the transducers are engaged sequentially—a normal mode of operation. It is presumed that, in sequential operation, there is no need to move the silicon wafers in the tank because the cavitation activity is homogeneously distributed over the entire cleaning region. Figure 5 demonstrates that, on the contrary, when the cleaning region was examined during sequential operation, the various transducers generate a band of cavitation activity with regions between the transducers where very little cavitation is occurring. During normal cleaning operations, it is not unusual to observe uncleaned bands on wafers; these data provide a plausible cavitation-dependent explanation for the existence of these bands.

III. SUMMARY AND CONCLUSIONS

We report here on a novel design for a cavitation probe and present some preliminary data on its application to megasonic cleaning. It appears to be quite robust and remarkably sensitive, permitting us to obtain real-time cavitation occurrence data. Experiments performed in a megasonic cleaner provided unexpected information on cavitation behavior, suggesting this probe has wide applications in industrial cavitation research.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation and the Defense Advanced Research Projects Agency.

Crum, L. A. (**1975**). ''Bjerknes forces in a stationary sound field,'' J. Acoust. Soc. Am. **57**, 1363–1370.

Crum, L. A. (**1979**). ''Surface oscillations and jet development in pulsating air bubbles,'' J. Acoust. Soc. Am. **40**, 131–135.
Crum, L. A. (**1988**). ''Cavitation microjets as a contributory mechanism for renal calculi disintegration in ESWL,'' J. Urology **140**, 1587–1596.
Crum, L. A. (**1994**). ''Sonoluminescence,'' Phys. Today **47**, 22–29.
Elder, S. A. (**1959**). ''Cavitation microstreaming,'' J. Acoust. Soc. Am. **31**, 54–65.
Ellis, A. T. (**1966**). ''On jets and shock waves from cavitation,'' Proceedings Sixth Naval Hydrodynamics Symposium, October 1966.
Geers, T. L. and Hasheminejad, M. (**1991**). ''Linear vibration analysis of an ultrasonic cleaning problem,'' J. Acoust. Soc. Am. **90**, 3238–3249.
Kashkoush, I., and Busnaina, A. (**1993**). ''Numerical simulation of the megasonic cleaning technique,'' Proc. Inst. Envir. Sci., pp. 386–395.
Kinsler, L. E., and Frey, A. R. (**1962**). *Fundamentals of Acoustics*, 2nd ed. (Wiley, New York), p. 121.
Lauterborn, W., and Bolle, H. (**1975**). ''Experimental investigations of cavitation bubble collapse in the neighborhood of a solid boundary,'' J. Fluid Mech. **72**, 391–402.
Menon, V. P. (**1990**). ''Particle adhesion to surfaces: Theory of cleaning,'' in *Particle Control for Semiconductor Manufacturing*, edited by R. P. Donovan (Marcel Dekker, New York), pp. 359–382.
Naude, C. F., and Ellis, A. T. (**1961**). ''On the mechanism of cavitation damage by nonhemispherical cavities collapsing in contact with a solid boundary,'' J. Basic Eng. **83**, 648–658.
Nyborg, W. M. (**1965**). ''Acoustic streaming,'' in *Physical Acoustics: Principles and Methods*, edited by W. P. Mason (Academic, New York), pp. 265–331.
Olson, L. G. (**1988**). ''Finite element model for ultrasonic cleaning,'' J. Sound Vib. **126**, 387–398.
Syverson, W. A., Fleming, M. M., and Schubring, P. J. (**1992**). ''The benefits of SC-1/SC-2 megasonic wafer cleaning,'' in *Proceedings of the Second International Symposium on Cleaning Technology in Semiconductor Device Manufacturing* (Electrochemical Society, Pennington, NJ), Vol. 92-12, pp. 10–17.
Tomita, Y., Shima, A., and Sugui, T. (**1986**). ''Mechanisms of impulsive pressure generation and damage pit formation by bubble-shock wave interaction,'' Proc. Intern. Symp. on Cavitation, 77-82.

Characterization of tissue microstructure using ultrasonic backscatter: Theory and technique for optimization using a Gaussian form factor

Michael L. Oelze[a)]
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

James F. Zachary
Bioengineering Program, University of Illinois, 1406 West Green Street, Urbana, Illinois 61801

William D. O'Brien, Jr.
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois and Bioengineering Program, University of Illinois, 1406 West Green Street, Urbana, Illinois 61801

(Received 13 December 2001; revised 14 June 2002; accepted 20 June 2002)

Characterization of tissue microstructure through ultrasonic backscatter is hypothesized to aid in detection and classification of diseased tissues. Radio frequency signals backscattered from tissues can be modeled according to the assumed shape, size, and distribution of scatterers in tissues. Power spectra of rf backscattered signals describe the frequency dependence of scatterers. Experimental measurements of ultrasonic backscatter from spontaneous mammary tumors in rats are obtained over the frequency range of 4 to 12 MHz. The power spectra measured from rat tumors are compared to theoretical power spectra derived from a 3D spatial autocorrelation function assuming a Gaussian distribution. Independent values of average scatterer diameter and acoustic concentration are obtained by approximating the measured power spectrum with a best-fit line. Enhanced B-mode images are made of the rat tumors and surrounding tissues with superimposed regions of interest quantified by estimated average scatterer sizes and acoustic concentrations. Scattering properties estimated inside the tumors and in surrounding tissues are shown to be distinct. Overall, estimates showed a 44.8% increase of average scatterer diameter inside the tumor as compared to tissues outside the tumor. With the exception of one rat, all estimates of the scatterers' average acoustic concentration inside the tumor were less than outside the tumors. [DOI: 10.1121/1.1501278]

PACS numbers: 43.80.Qf, 43.80.Vj [FD]

I. INTRODUCTION

Conventional B-mode images of living tissues created with a clinical ultrasound scanning system are made from radio frequency (rf) envelope-detected echo signals. The rf echo signals are created by reflections from interfaces between acoustically different regions and by incoherent scattering from tissue microstructures. The rf echoes are hypothesized to contain frequency-dependent details about the smaller-scale structures (less than the wavelength) in the tissues. Generally, processing of conventional B-mode images removes the frequency-dependent information. Conventional B-mode images relate the envelope of the backscattered rf echoes to a gray scale. The conventional B-mode scanners (bandwidths of 1 to 10 MHz) can resolve structures on the order of hundreds of micrometers to centimeters in scale.[1] In order to display and quantify the smaller-scale structures of tissues, the frequency-dependent information must be utilized.

The frequency dependence of scattering from small structures (compared to wavelength) has been used to extract details about the structures of different materials and media. As noted by Insana,[2] light scattering from lasers has been used to characterize the internal structure of glasses and polymers,[3] and x rays have been used to characterize the structure of atoms.[4] In acoustics, low-frequency sound waves (below 1 kHz) have been used to measure the size and distribution of turbulence in the atmosphere.[5]

In a statistical sense, the frequency dependence of ultrasound backscatter can be used to extend the resolution of conventional B-mode imaging to tens of micrometers. Ultrasonic backscattering techniques have been used to characterize different aspects of microstructure in biological tissues. Specifically, backscatter has been used to extract the average scatterer sizes and acoustic concentrations (product of the number concentration of scatterers times the relative impedance difference between the scatterers and surrounding tissues) from biological tissues. Feleppa *et al.*[6] used frequency-dependent backscatter to quantify the scatterer sizes and acoustic concentration of ocular tumors. Lizzi *et al.*[7] used the same techniques to examine the structure of the liver, while Insana *et al.*[8] used these techniques to parametrize renal tissues. Numerous other experiments and detailed scattering

a)Electronic mail: oelze@brl.uiuc.edu

models have been developed and used to describe backscattered ultrasound from biological tissues.[9]

Acoustic scattering theories for biological tissues assume that the tissues can be modeled as inhomogeneous fluids.[2] Scattering occurs when an acoustic wave propagates across a region that has a different density and/or compressibility relative to the surrounding tissue or fluid. In most cases, a sphere, a cylinder, or some simple geometric configuration has been used to approximate the scattering particle shape. If the size (diameter) of the scatterer is on the same order or smaller than the wavelength of sound, then the pattern of the scattered signal is well known for simple geometric shapes.[10] In addition, the size and shape of the scattering region determine the magnitude at which a specific frequency of sound will be scattered.

For example, in tumor pathology these scattering regions as viewed by a pathologist using an optical microscope include typical macrostructures such as tumor nodules, and microstructures such as epithelial cells. Tumor nodules could range in size from millimeters to centimeters in diameter. Epithelia cells measuring 10–20 μm in diameter could be arranged as individual cells in supporting stroma. In addition, epithelial cells could also be arranged in tubules or glands that measure 50–150 μm in diameter. These epithelial cells are supported by a stroma composed of connective tissue and blood vessels that have a unique tissue density and can also encapsulate the mass and form a tumor nodule (macrostructure).

Enhanced B-mode images can be constructed by relating scattering information (estimated average scatterer size and acoustic concentration in this study) about the tissue microstructure obtained from the backscattered rf signals to selected pixels or regions of interest (ROIs). Enhanced images are formed by superimposing the quantified estimates on conventional B-mode images.[11] An enhanced imaging technique that could be adapted for real-time *in situ* clinical diagnosis (classification) of cancerous tissues would have enormous medical significance.

This study examines the estimation of scattering properties like the average scatterer size and acoustic concentration in order to construct enhanced B-mode images. Section II outlines the theory used to model the tissue microstructures. Section III outlines a novel estimation scheme used to invert the model from the measured data and obtain the average scatterer parameters. The new estimation scheme does not assume that the scattered power spectrum has linear frequency dependence. Instead, the new estimation scheme uses a least-squares procedure to fit a line to a measured form factor and relates the slope and intercept of that line to average scatterer properties. Section IV discusses the expected variance inherent in the estimation scheme used. Section V shows enhanced B-mode images of four rats with spontaneous mammary tumors and compares scattering properties from ROIs inside the tumors and in surrounding tissues. The final section gives some concluding remarks and future directions for this work.

II. THEORETICAL DESCRIPTION

Scattering in tissues arises from spatial changes in the acoustic impedance about the local mean value. If the wavelength of ultrasound is larger than or on the same order as the size of a scattering region, some of the incident acoustic energy will be scattered in all directions. The amount of energy scattered in any one direction will depend on the size, shape, and orientation of the scatterer relative to the wavelength and propagation direction of sound.[12] Backscatter refers to that portion of the scattered field directed back to the source aperture. The frequency content of the backscattered pulse will also depend on the size and shape of the scatterer.

The tissue model used in this study assumes a sparse collection of randomly distributed, weakly scattering particles (Born approximation). Under the assumption of weak scattering, multiple scattering is assumed to be negligible. Because the scatterers are randomly distributed, the rf backscattered signal will be due to incoherent scattering. In the specific case of the rat tumors, it is expected that there may exist distinct regions of homogeneous scattering. The homogeneous regions could lead to a coherent scattering component. If the homogeneous regions were several wavelengths in size, then the effects of coherent scatter from these regions would be outside the analysis bandwidth. The effects of the coherent scatter were assumed negligible and thus were ignored.[8,13] The exact location and size of each scatterer cannot be determined, but rather a statistical description of average properties within a certain volume of interest may be deduced.[13–15]

The size, shape, distribution, and elastic properties of the scatterers in the medium will determine the frequency dependence of the backscattered signals. In the tissue model, 3D spatial autocorrelation functions can be used to describe the size, shape, elastic properties, and distribution of scatterers.[13–16] Instead of calculating the correlation functions directly for the scattering medium, some authors have found it convenient to calculate the acoustic intensity form factor F.[14,15] The form factor is proportional to the Fourier transform of the autocorrelation function and describes the frequency dependence of the scattering in terms of the size, shape, and elastic properties of the scatterers. The Fourier transform of the autocorrelation function for statistically stationary scattering is defined as the power spectrum and incorporates the form factor. Insana *et al.* showed that in the special case of randomly positioned, spherical scatterers the form factor is a function of the average effective scatterer size (diameter) and the frequency.[14]

For the present study the Gaussian form factor is used to model the soft tissue scattering from spontaneous mammary tumors in rats. The Gaussian form factor has been used to model the scattering properties of many soft tissues.[7,17,18] The choice of the Gaussian form factor over other form factors (exponential and spherical shell) was made because the average chi-square value between the measured form factor and the theoretical form factors was smallest with the Gaussian. The Gaussian form factor, as derived by Insana *et al.*, is given by[14]

$$F_{\mathrm{Gauss}}(2k) = e^{-0.827k^2a_{\mathrm{eff}}^2}, \tag{1}$$

where k is the acoustic wave number and a_{eff} is the average effective scatterer radius. The Gaussian form factor represents a spherical scatterer that varies continuously with the surrounding tissues and has a uniform particle radius (see Ref. 9, Chap. 4, Appendix B). Instead of a sharp discontinuity with the impedance like a solid or fluid-filled sphere or spherical shell, the Gaussian scattering particle is represented by a gradual change in the impedance from the surrounding tissues. Instead of having a definite radius like the spherical shell or a solid or fluid-filled sphere, the Gaussian function has an effective radius, a_{eff}, that is related to particle shape or impedance distribution. In the tumor model (rat fibroadenoma) used in this study, tubules and glands formed by neoplastic epithelial cells ranged between 50–150 μm in diameter. These structures were distributed at random in a supporting matrix composed principally of well-collagenized fibrous connective tissue.

Typical backscatter measurements from tissue volumes are made by choosing a region of interest (ROI) and gating the backscattered rf time signal corresponding to the ROI. The measured power spectrum is the magnitude squared of the Fourier transform of the gated, rf time signal. The effects of the gating function (windowing function) and beam function of the transducer (weakly focused piston transducer near the focal plane) are needed to properly model the backscattering measurement. Lizzi *et al.* incorporated the gating function (Hanning window) and beam pattern effects in the development of the model to normalize the theoretical backscattered power spectrum.[16] In the frequency domain, the normalized, theoretical power spectrum is given by

$$W(f)=\frac{185Lq^2a_{\text{eff}}^6\rho z_{\text{var}}^2 f^4}{[1+2.66(fqa_{\text{eff}})^2]}e^{-12.159f^2a_{\text{eff}}^2}, \tag{2}$$

with L the gate length (mm), q the ratio of aperture radius to distance from the region of interest, and f is the frequency (MHz). The quantity, ρz_{var}^2, is termed the acoustic concentration and is the product of the number of scattering particles per unit volume (mm^{-3}), ρ, and the fractional change in the impedance between the scattering particles and the surrounding medium, $z_{\text{var}}=(Z-Z_0)/Z_0$, where Z is the acoustic impedance of the scatterers and Z_0 is the acoustic impedance of the surrounding medium. The exponential term is the Gaussian form factor term in the frequency domain.

The measured power spectrum is the magnitude squared of the Fourier transform of the gated time signal. The effects of the equipment on the power spectrum measurement are factored out by dividing by a calibration spectrum.[13,15] To obtain the calibration (reference) spectrum, a pulse is reflected from a planar surface with known reflectivity and received by the emitting transducer. The transducer is a single element, piston style, and weakly focused. The same transducer, equipment, and settings are used to send and receive the reflected pulses that are used to obtain the backscattered signals from a selected ROI in the tissue. The planar surface is located at the same axial distance from the transducer as the center of the ROI selected in the tissue (in the focal zone of the transducer). The measured power spectrum of the signal from the gated ROI in the tissue is divided by the power spectrum of the reference signal.

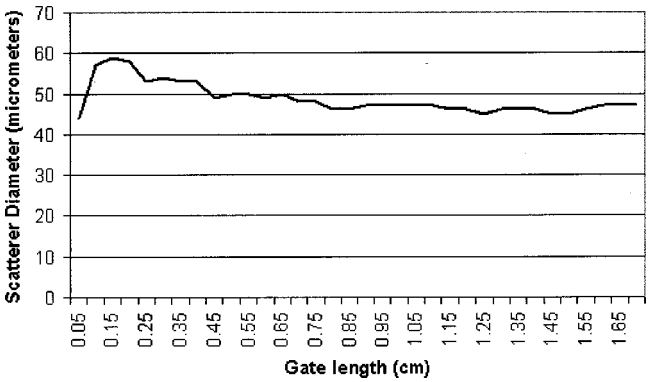


FIG. 1. Estimated scatterer diameters from Gaussian-type particles in a simulated random scattering medium versus gated length of the backscattered rf signal.

The data must be acquired from a large enough volume in order to obtain good estimates because the measured power spectrum describes statistical properties of the medium. Several authors have examined the relationship between the size of the ROI (gate length) and the estimation of scattering power and scatterer statistics.[16,19–22] As the size of ROI is increased, the statistical estimations should converge for a random scattering medium. Two important factors are key to choosing an appropriate gate length. First, the time–bandwidth product should be greater than 1. Second, a larger gate length means that more scatterers will be included in the spectrum. The sum of the random phase contributions from a large amount of scatterers is more likely to cancel than from a few randomly spaced scatterers. Topp *et al.*[19] examined the question of convergence and concluded that adequate estimations could be obtained by taking ROIs with gate lengths of 10 or more wavelengths at the center frequency of interrogation. Figure 1 shows estimations of scatterer diameter made from a computer software phantom. The scatterer size estimates were made using the estimation scheme outlined in Sec. III. The particles in the software phantom scattered according to the Gaussian form factor with an effective diameter of 49 μm. The impedance of the scattering particle, Z, was given the value of glass (density 2.38 g/cm^3, longitudinal sound speed 5570 m/s) and the impedance of the surrounding medium, Z_0, was given the value for agar (density 1.0 g/cm^3, longitudinal sound speed 1540 m/s). The center frequency of the incident pulse and analysis bandwidth was around 7 MHz ($\lambda=0.22$ mm) with a 70% bandwidth defined at -6 dB (half-power bandwidth of the pulse–echo signal). The scattering was assumed to occur in the focal zone of a transducer with an f-number of 5, where the value of q would be very small. From Fig. 1 it is seen that the estimations of scatterer diameter begin to converge to 49 μm at a gate length of about 2.5 mm, or 11 wavelengths of the center frequency.

Typically, a ROI is made up of a number of parallel gated A lines of the same length. The power spectrum measured from an individual A line may have sharp fluctuations (noise) due to the random nature of the scatterer spacings in the medium.[23] The fluctuations represent deviations from the shape of the power spectrum modeled with the form factor. Consecutive parallel A lines should have similar statistical

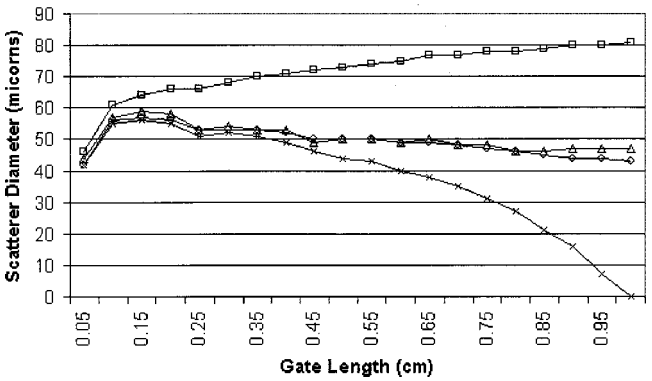


FIG. 2. Estimated scatterer diameters from Gaussian-type particles in a simulated attenuating medium (8 dB/cm @7 MHz) using attenuation compensation and no attenuation compensation; Δ, unattenuated signal; □, uncompensated attenuated signal; ×, point compensated attenuated signal; ◇, compensated signal using Eq. (6).

properties if located near each other. The difference between rf echoes from the consecutive A lines will be the difference in random spacings of the scatterers. If the statistical properties of consecutive A lines can be assumed the same, then averaging the consecutive A lines should reduce the random fluctuations seen in a single A-line spectrum. Averaging will yield a better overall description of the scatterer statistical properties. The average measured power spectrum is given by

$$W_{\text{meas}}(f)=\frac{1}{N}\frac{\Re^2}{4}\sum_{n=1}^{N}\frac{|FT\{p_n(t)\}|^2}{W_{\text{ref}}(f)}, \tag{3}$$

where $\Re$ is the reflection coefficient of the planar reflector, $p_n(f)$ is the gated rf time signal of the nth A line, N is the number of gated A lines, and $W_{\text{ref}}(f)$ is the reference power spectrum.

The measured power spectrum describes the frequency dependence of the scatterers in the medium. Frequency-dependent attenuation will affect the measured power spectrum and, if not compensated for, give rise to inaccurate estimates of scattering properties. Typically, attenuation increases with frequency. The slope of the power spectrum will be more negative if the frequency-dependent losses are not taken into account. In order to give optimal estimations of scattering properties, the frequency-dependent attenuation must be compensated. If the gate size were infinitesimally small, the frequency-dependent attenuation losses could be compensated, assuming constant attenuation, by simple-point attenuation compensation given by[14]

$$W_{\text{comp}}(f)=W_{\text{meas}}(f)e^{4\alpha(f)x}, \tag{4}$$

where $\alpha(f)$ is the frequency-dependent attenuation and x is the propagation distance through the attenuating medium to the center of the gated region. When the gate length is not small and the attenuation is not negligible, point compensation does not correctly account for frequency-dependent losses. Figure 2 shows estimations of scatterer sizes from software phantoms with large attenuation (8 dB/cm@7 MHz) as the gate length is increased using the point-attenuation-compensation function and no attenuation compensation. The large attenuation was used to coincide with the attenuation measured in rat intercostal tissues (1.1 dB/cm/MHz).[24] Point-attenuation compensation causes an underestimation of the average scatterer diameter, while not accounting for frequency-dependent attenuation leads to an overestimation of scatterer diameter. Attenuation-compensation functions exist that account for frequency-dependent losses to the measured power spectrum obtained from gates of finite length.[25–27] The attenuation-compensation functions are represented by

$$W_{\text{comp}}(f)=W_{\text{meas}}(f)A(f,L), \tag{5}$$

where L is gate length. An attenuation-compensation function derived for gated signals was used in this study, and is given by[27]

$$A(f,L)=e^{4\alpha_0(f)x_0}\left[\frac{2\alpha(f)L}{1-e^{-2\alpha(f)L}}\right]^2, \tag{6}$$

where $\alpha_0(f)$ and x_0 are, respectively, the attenuation and propagation distance of the intervening tissues and $\alpha(f)$ is the attenuation coefficient in the gated region (assumed constant).

III. MODEL INVERSION

Several techniques have been used to parametrize the shape of the measured power spectrum or to estimate properties of the scatterers in a tissue from the shape of the measured power spectrum. These techniques relate the theoretical models to the measured backscatter data. Enhanced images are then formed with the spectral parameters or the estimated scatterer properties in order to aid in identifying and characterizing diseased tissues. Several authors have successfully used enhanced (parametric) images of the backscatter coefficient and attenuation to aid in tissue characterization.[15,16,19,28–31] Zagzebski *et al.* used the backscatter coefficient parametric images to classify diseased liver.[29] Other researchers have used the slope and intercept of the measured backscattered power spectrum to create parametric images.[19] Feleppa *et al.*[30] and Lizzi *et al.*[16] have used the intercept, slope, and midband fit to the measured backscattered power spectrum and create parametric images that aided in detection and classification of cancer in the eye and prostate. Insana and Hall[15] created parametric images from scatterer size estimates, scattering strength measurements, chi-square images, and integrated backscatter coefficient images. The chi-square images were important to show how well the model represented the interrogated medium for any particular ROI.

If the deduced parameters are related to actual physical properties of the tissues, i.e., the average scatterer size and acoustic concentration, it may be possible to partially verify the particular form factor models through other means. Possible verification of the model might also be obtained through light microscopy. Ultrasound waves detect changes in the mechanical (compressibility, density) properties of the tissues, while light microscopy detects electromagnetic (dielectric) changes in the tissues. Scattering particles are expected to have both mechanical and electromagnetic differences with surrounding tissues. There will not be a one-to-

one correlation between mechanical and electromagnetic property changes in the scattering particles, but the optical characteristics should give an approximation of shape and size that can be related to acoustic estimations and aid in verification of the form factor models. Estimation of the size and acoustic concentration properties of scatterers is desirable because they are hypothesized to relate directly to morphological features in the tissues examined.

Insana and Hall[15] estimated the average scatterer size by minimizing the average squared difference (MASD) between the theory and measurements over a range of scatterer size values. The MASD between the theoretical normalized power spectrum and the measured power spectrum is calculated by[14]

$$\mathrm{MASD}=\min\left(\frac{1}{m}\sum_{i=1}^{m}(X_i-\overline{X})^2\right), \tag{7}$$

with

$$X_i=10\log[W_{\mathrm{meas}}(f_i)A(f_i,L)F_{\mathrm{Gauss}}(a_{\mathrm{eff}},f_i)/f_i^4], \tag{8}$$

and

$$\overline{X}=\frac{1}{m}\sum_{i=1}^{m}X_i, \tag{9}$$

where A is the attenuation-compensation term and m is the number of points in the frequency data bandwidth. The estimate of the average scatterer diameter is the argument that minimizes Eq. (7).

Typically, the measured power spectrum is in terms of the scatter strength, a decibel scale

$$S_{\mathrm{meas}}(f)=10\log[W_{\mathrm{comp}}(f)]. \tag{10}$$

Inverting the measurement of the backscattered power spectrum with the theoretical spectrum enables scattering parameters to be estimated. Relating Eq. (10) to Eq. (2) gives

$$S_{\mathrm{meas}}(f)=10\log\left\{\frac{185Lq^2a_{\mathrm{eff}}^6\rho z_{\mathrm{var}}^2f^4}{[1+2.66(fqa_{\mathrm{eff}})^2]}e^{-12.159f^2a_{\mathrm{eff}}^2}\right\}. \tag{11}$$

Expanding the logarithm in Eq. (11) yields

$$\begin{aligned}S_{\mathrm{meas}}(f)=&10\log[185Lq^2a_{\mathrm{eff}}^6\rho z_{\mathrm{var}}^2]\\&-10\log[1+2.66(fqa_{\mathrm{eff}})^2]\\&-10(12.159f^2a_{\mathrm{eff}}^2)\log e+10\log f^4.\end{aligned} \tag{12}$$

Typically, the ratio of transducer radius to distance from transducer to ROI is small ($q\ll1$), which allows the second term on the right-hand side to be approximated by

$$\log(1+x)\approx x\log e\quad(\text{for small } x), \tag{13}$$

because the value $2.66(fqa_{\mathrm{eff}})^2\ll1$. Subtracting the f^4 term from the right-hand side gives

$$S_{\mathrm{meas}}(f)-10\log f^4\approx M(a_{\mathrm{eff}}^2)f^2+I(a_{\mathrm{eff}}^6,\rho z_{\mathrm{var}}^2), \tag{14}$$

where

$$M(a_{\mathrm{eff}}^2)=-4.34[12.159+2.66q^2]a_{\mathrm{eff}}^2, \tag{15}$$

and

$$I(a_{\mathrm{eff}}^6,\rho z_{\mathrm{var}}^2)=10\log[185Lq^2a_{\mathrm{eff}}^6\rho z_{\mathrm{var}}^2]. \tag{16}$$

Comparing Eq. (14) to that of a line ($y=mx+b$), where $x=f^2$, it is seen that the intercept, I, is a function of the effective average radius (diameter) of the scatterers and the acoustic concentration of scatterers. The slope, M, is a function of the average effective diameter of scatterers only. The slope and intercept of the line can be found by using least squares to fit a line to the measured data (left side) of Eq. (14). From the slope value, the average scatterer diameter is estimated. Once the average effective scatterer diameter is estimated, the average acoustic concentration of scatterers can be estimated from the intercept value and the estimated value of the scatterer diameter. The best-fit line parameters derived in Eqs. (15) and (16) differ from the linear parameters used in the estimation scheme by Lizzi *et al.* in that Eqs. (15) and (16) represent the best-fit line to the measured form factor (with some constants related to the gating and beam functions).[32] The linear parameters estimated from the approach by Lizzi *et al.* are deduced from the scattered power spectrum, which is not linear over a large range of frequencies.[32]

The estimation scheme described here has a distinct advantage over the MASD estimation scheme in the number of calculations needed to derive the estimates. In the MASD estimation scheme, different values for the average scatterer diameter are chosen until a minimum average squared difference is found. The summation of the squared difference values is done every time a different average scatterer diameter is tried. Several choices of scatterer diameter may be tried before the average squared difference is minimized. For each scatterer diameter tried, the squared difference must be calculated at each data point in the analysis bandwidth. In the best-fit line estimation scheme the summation of the bandwidth values is done only once and the estimated average scatterer diameter is obtained.

IV. EXPECTED VARIANCE OF ESTIMATION SCHEME

The expected variance of the estimation scheme developed in this work was calculated in order to optimize the measurement for inverting the scatterer parameters. From Chaturvedi and Insana,[33] the expected variance of the average scatterer diameter from measurement of the Gaussian form factor is given by

$$\mathrm{var}\,\hat{D}\approx\sum_{j=1}^{N}\left\{\mathrm{var}\,\hat{F}_{\mathrm{Gauss}}(f_j)\left(\frac{\partial D}{\partial\hat{F}_{\mathrm{Gauss}}(f_j)}\right)^2\right\}, \tag{17}$$

where $\hat{D}$ is the estimated average scatterer diameter and N is the number of data points used in the analysis bandwidth. To evaluate the variance of $\hat{D}$, the relationship between the average scatterer diameter D and the measured form factor must be determined. Using the best-fit line technique, for small q, Eq. (14) can be rewritten as

$$S_{\mathrm{meas}}(f)-10\log f^4\approx10\log[\hat{F}_{\mathrm{Gauss}}(f,D)\cdot C(D,\rho z_{\mathrm{var}}^2)], \tag{18}$$

so that

$$10\log[\hat{F}_{\text{Gauss}}(f,D)\cdot C(D,\rho z_{\text{var}}^2)]=M(D)f^2+I(D,\rho z_{\text{var}}^2). \tag{19}$$

The slope M can be determined using regression to find the best-fit slope

$$M(D)=\frac{N\Sigma_{j=1}^{N}f_j^2\,10\log[\hat{F}(f_j)\cdot C]-\Sigma_{j=1}^{N}f_j^2\Sigma_{j=1}^{N}10\log[\hat{F}(f_j)\cdot C]}{N\Sigma_{j=1}^{N}f_j^4-(\Sigma_{j=1}^{N}f_j^2)^2}, \tag{20}$$

where $\hat{F}$ represents the estimated form factor. From Eq. (15), when $q\ll 1$

$$M(D)\approx -13.2D^2, \tag{21}$$

which gives, after simplification

$$-13.2D^2=\frac{N\Sigma_{j=1}^{N}f_j^2\,10\log\hat{F}(f_j)-\Sigma_{j=1}^{N}f_j^2\Sigma_{j=1}^{N}10\log\hat{F}(f_j)}{N\Sigma_{j=1}^{N}f_j^4-(\Sigma_{j=1}^{N}f_j^2)^2}, \tag{22}$$

where the factors involving $10\log C$ subtract out. Defining

$$\sigma_{f^2}=N^2\sum_{j=1}^{N}f_j^4-N\left(\sum_{j=1}^{N}f_j^2\right)^2, \tag{23}$$

and rearranging yields

$$\sum_{j=1}^{N}\left[\frac{13.2D^2}{N^2}+\frac{(f_j^2-\overline{f^2})10\log\hat{F}(f_j)}{\sigma_{f^2}}\right]=0, \tag{24}$$

which is of the form $\Sigma_{j=1}^{N}Y_j=0$. The chain rule of differentiation is used

$$\frac{\partial Y_j}{\partial\hat{F}(f_j)}=\frac{\partial Y_j}{\partial D}\frac{\partial D}{\partial\hat{F}(f_j)}, \tag{25}$$

with

$$\frac{\partial Y_j}{\partial D}=\sum_{j=1}^{N}\frac{26.4D}{N^2}=\frac{26.4D}{N}, \tag{26}$$

and

$$\frac{\partial Y_j}{\partial\hat{F}(f_j)}=\frac{4.34(f_j^2-\overline{f^2})}{\sigma_{f^2}\hat{F}(f_j)}. \tag{27}$$

Combining Eqs. (26) and (27) with Eq. (25) yields

$$\frac{\partial D}{\partial\hat{F}(f_j)}=\frac{4.34N}{26.4D}\frac{(f_j^2-\overline{f^2})}{\sigma_{f^2}\hat{F}(f_j)}. \tag{28}$$

Substituting Eq. (28) into Eq. (17) gives

$$\text{var}\,\hat{D}=\sum_{j=1}^{N}\text{var}\,\hat{F}(f_j)\left(\frac{4.34N}{26.4D}\frac{(f_j^2-\overline{f^2})}{\sigma_{f^2}\hat{F}(f_j)}\right)^2. \tag{29}$$

From Chaturvedi and Insana[33]

$$\text{var}\,\hat{F}(f_j)\approx[F_{\text{Gauss}}(f_j)]^2g(f_j), \tag{30}$$

with

$$g(f)=\left[1+\left(\frac{\sin(2\pi fN_p\Delta t)}{N_p\sin(2\pi f\Delta t)}\right)^2\right], \tag{31}$$

where Δt is the sampling interval of the time signal and N_p is the number of points in the time signal. Approximating $\hat{F}$ by F_{Gauss} yields the result

$$\text{var}\,\hat{D}=\frac{2.09\times10^{-2}N^2}{D^2\sigma_{f^2}^2}\sum_{j=1}^{N}g(f_j)(f_j^2-\overline{f^2})^2. \tag{32}$$

For a long time signal, $g(f_j)\approx 1$, which gives

$$\text{var}\,\hat{D}=\frac{2.09\times10^{-2}}{D^2}\left[\sum_{j=1}^{N}f_j^4-N\overline{f^2}^2\right]^{-1}. \tag{33}$$

Calculation of the variance of the average diameter estimation for the MASD technique gives the same value as the best-fit line estimation technique. Simulations and experiments by Chaturvedi and Insana showed that the expected variance derived for the MASD estimation scheme was valid for a Gaussian form factor.[33] The reason that the expected variance of the estimated scatterer diameter is the same for the best-fit line estimation scheme and the MASD scheme is that both techniques minimize the squared difference. The MASD finds the value of the scatterer diameter that minimizes the squared difference between the measured data and the theoretical. The best-fit line is a least-squares fit of the theoretical spectrum to the experimental spectrum. Instead of trying different values for the average scatterer diameter until the average squared difference is minimized, the best-fit line automatically minimizes the average squared difference and consequently yields the best estimate of the average scatterer diameter.

For the estimation scheme, the variance is inversely proportional to the average scatterer diameter squared. The larger the average scatterer diameter the more precise the scatterer estimation should be. However, as the relative size of the wavelength becomes smaller, the theory may not correctly describe the scattering. Likewise, estimating the scatterer parameters at higher center frequencies over a larger bandwidth also decreases the expected variance in the measurement of D. The analysis bandwidth needs to be chosen with consideration of the expected scatterer properties. Insana and Hall[15] discussed the importance of keeping the ka_{eff} value within the range of 0.5 and 1.2. Using frequencies too far below the range (larger wavelength compared to scatterer size) means the signal scattered from the microstructures may be too small to be discerned from the noise. Using frequencies too far above the range (smaller wavelength compared to the scatterer size) means that the size and shape of the scatterer no longer dictates the frequency dependence. If

a certain average size of the scatterer is expected, the analysis frequency bandwidth needs to be chosen accordingly. The calculation of the expected variance in the estimation technique exhibits the trade-off between larger bandwidth and center operating frequency and the restrictions on the ka_{eff} value. The calculation of the expected variance is important to optimizing the measurement technique. The measurement technique is optimized for the variance by choosing the largest bandwidth and largest center frequency that fit within the acceptable ka_{eff} range. The dependence of the expected variance directly on $g(f)$ also shows that a better estimation may be made from a longer gated signal.

V. ENHANCED IMAGES OF RAT MAMMARY TUMORS

The least-squares best-fit line estimation scheme utilizing the Gaussian form factor model was applied to B-mode image rf data obtained from mammary tumors that had developed spontaneously in rats. The average scatterer size and acoustic concentration (physical concentration of scatterers times the average relative impedance difference between scatterers and surrounding tissues) were estimated from the rf signals used to construct the B-mode images. The experimental protocol was approved by the Laboratory Animal Care Advisory Committee at the University of Illinois at Urbana-Champaign and satisfied all campus and NIH rules for the humane use of laboratory animals. Four Sprague-Dawley rats (Harlan, Indianapolis) that had developed spontaneous mammary tumors were evaluated. Each rat was euthanized with CO_2, and the tumor and surrounding area were immediately shaved and depilated. The rat was then placed on a holder in a tank of degassed water at 37 °C for scanning with an ultrasonic transducer. A single-element broadband transducer was used to scan laterally across the tumors and surrounding tissues. The transducer was moved laterally by a micropositioning system with step size of 100 μm between each scan. The transducer had a center frequency of about 8 MHz with an 80%–90% frequency bandwidth (Gaussian spectrum defined at -6 dB). The transducer had a diameter of 12 mm and a focal length of 50 mm. Measurements were taken within or near the focal zone. The frequency and bandwidth were chosen based on the expected values of the average scatterer sizes in the rat tissues. The frequency and bandwidth were chosen so that the variance in the measurement would be minimized while keeping within the optimal ka_{eff} range. The transducer was operated in pulse/echo mode through a Panametrics 5800 pulser/receiver (Waltham, MA). The signals were recorded and digitized on an oscilloscope (Lecroy 9354 TM; Chestnut Ridge, NY) and downloaded to a PC for postprocessing. The sampling rate of the received signals was 50 MHz. Figure 3 shows a picture of the rat in a holder ready to be scanned. The mammary tumor is located on the upper left portion of the chest. The black line across the tumor and chest represents the direction and the length of the lateral scan.

Two-dimensional B-mode images were constructed from the ultrasonic scan lines from each rat. Each scan line was taken from a gated portion of the rf backscattered signal and recorded for postprocessing. The axial distance of the B-mode images was associated with the time base of the signals, and the lateral position of the images was associated with the lateral step length where the individual scan lines were recorded. From the 2D B-mode images, ROIs were divided into regions inside and outside the tumor where the B-mode images appeared to be homogeneous (no interfaces or large echoes). The regions scanned and analyzed outside the tumors were intercostal tissues. The ROIs were separated to determine if normal surrounding tissues could be differentiated from the tissues inside the tumors based on the scatterer properties.

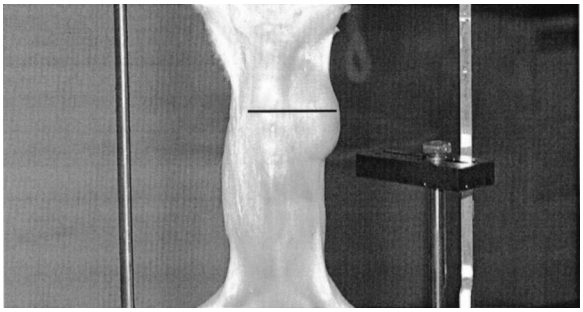


FIG. 3. Rat to be scanned in holder in tank of degassed water. Ultrasonic scans ran laterally across the chest and tumor following along the black line. The tumor is located on the upper left side of the chest.

After scanning, the tumors were excised, fixed in 10% neutral-buffered formalin, processed, and stained with hematoxylin and eosin stain for routine histologic evaluation by light microscopy. The tumors were diagnosed as fibroadenomas following histopathologic evaluation. Figure 4 shows a photomicrograph of a tissue slice through a tumor using light microscopy. The tumor consisted of well-differentiated mammary epithelial cells arranged in tubules and glands that were distributed at random in a supporting matrix composed principally of well-collagenized fibrous connective tissue. Tubules and glands formed by neoplastic epithelial cells ranged in size between 50–150 μm in diameter.

Scatterer estimates were made for each ROI using the best-fit line estimation scheme. The backscattered rf signal

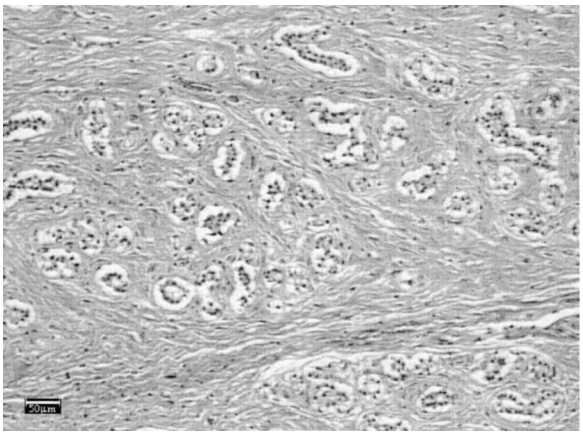


FIG. 4. Photomicrograph of a rat mammary tumor. The specimen was fixed in 10% neutral buffered formalin, dehydrated, embedded in paraffin, mounted on a glass slide, and stained with hematoxylin and eosin stain for microscopic evaluation. The tumor consists of well-differentiated epithelial cells arranged in acini surrounded by bands of fibrous connective tissue. The scale bar represents 50 μm.

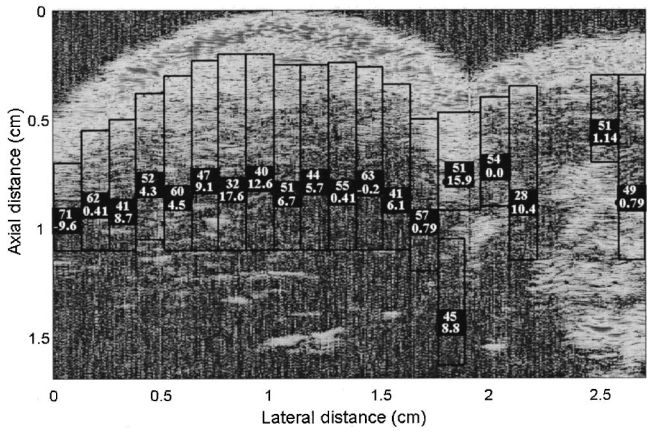


FIG. 5. Enhanced B-mode image of rat 1 with the tumor located between the lateral distance of 0 to 1.8 cm. The top number in the boxes represents the average scatterer diameter (μm) and the bottom number represents the average acoustic concentration (10* log[mm^{-3}]).

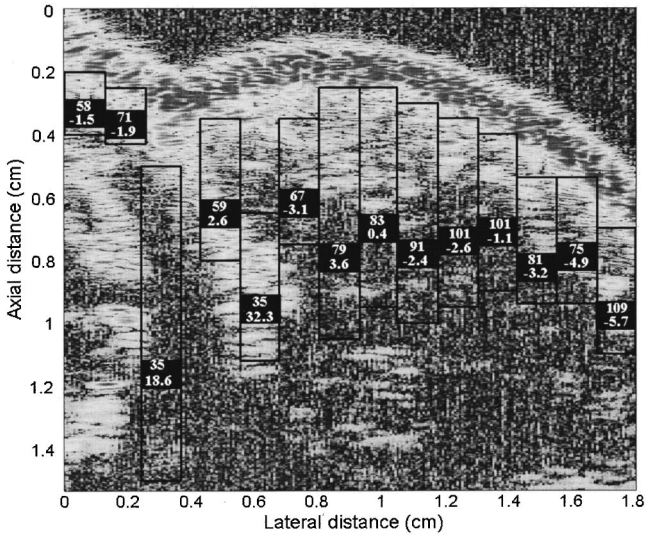


FIG. 6. Enhanced B-mode image of rat 2 with the tumor located between the lateral distance of 0.4 to 1.8 cm. The top number in the boxes represents the average scatterer diameter (μm) and the bottom number represents the average acoustic concentration (10* log[mm^{-3}]).

was gated from each ROI using a Hanning window and the measured power spectrum was calculated according to Eq. (3). The average scatterer diameter and average acoustic concentration were estimated from the measured power spectrum. The attenuation for the tissues was deduced from the backscatter according to the method listed in Shung and Thieme (p. 247).[9] Regions where the scattering appeared homogeneous with depth were selected in order to estimate the best attenuation value by examining the overall loss of the backscattered signal with depth. If the region selected is homogeneous with scattering, then the only difference between scattering from particles near and farther away is the attenuation loss over the extra propagation distance. The attenuation deduced from the backscatter in a particular rat was then used to estimate the scatterer properties throughout the particular rat. The average attenuation coefficient estimated from the backscatter was 0.9 dB/cm/MHz, assuming a linear dependence.

Conventional B-mode images were constructed of the tumors and surrounding tissues. Following the enhanced imaging model of Topp *et al.*[19] the average scatterer properties were superimposed in boxes corresponding to the gated ROIs. In the enhanced B-mode images by Topp *et al.*, the number inside the superimposed box represented the slope of the measured average backscatter coefficient. In this work, two numbers representing the average scatterer diameter and average acoustic concentration are listed in the ROI boxes. Figures 5–8 show enhanced B-mode images of the rat tumors and surrounding tissues. The top number represents the average scatterer diameter (micrometers) in the ROI, and the bottom number represents the average acoustic concentration (mm^{-3}). Noise is evident in the B-mode images of the rat mammary tumors. The noise is due to the low-power system used for imaging and not because of the large bandwidth.

The average scatterer properties estimated inside the tumor and outside the tumor from healthy tissues were then compared. Figure 9 shows the mean of the average scatterer diameters measured inside the tumors and outside the tumors of the four rats. An examination of the average estimated scatterer diameters inside and outside the tumors showed that the scatterers inside the tumors were larger than scatterers in the surrounding normal tissues. Overall, the estimates showed a 44.8% increase of average scatterer diameter inside the tumor as compared to the tissues outside the tumor. With the exception of the first rat, all differences of scatterer size in the tumors were outside the standard deviation of scatterer size estimates outside the tumors.

Figure 10 shows the estimated average acoustic concentration inside and outside the tumors of the four different rats. With the exception of the first rat, all estimates of average acoustic concentration inside the tumor were less than the average acoustic concentration outside the tumors. Error bars representing the standard deviation are not included be-

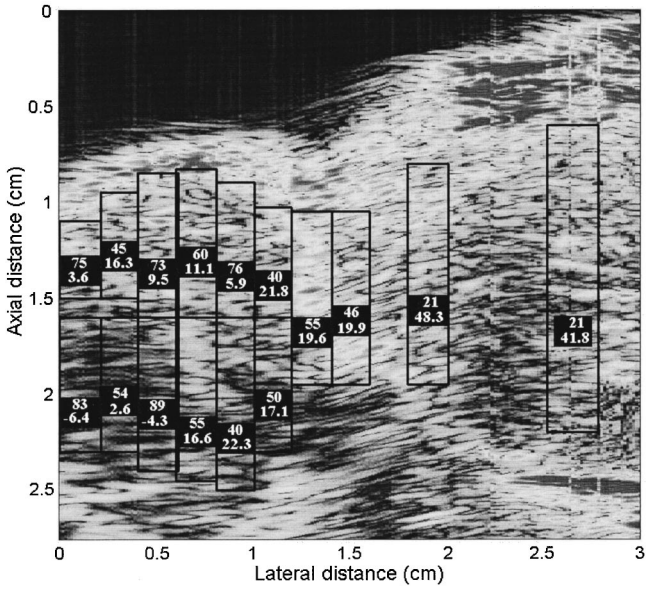


FIG. 7. Enhanced B-mode image of rat 3 with the tumor located between the lateral distance of 0 to 1.1 cm. The top number in the boxes represents the average scatterer diameter (μm) and the bottom number represents the average acoustic concentration (10* log[mm^{-3}]).

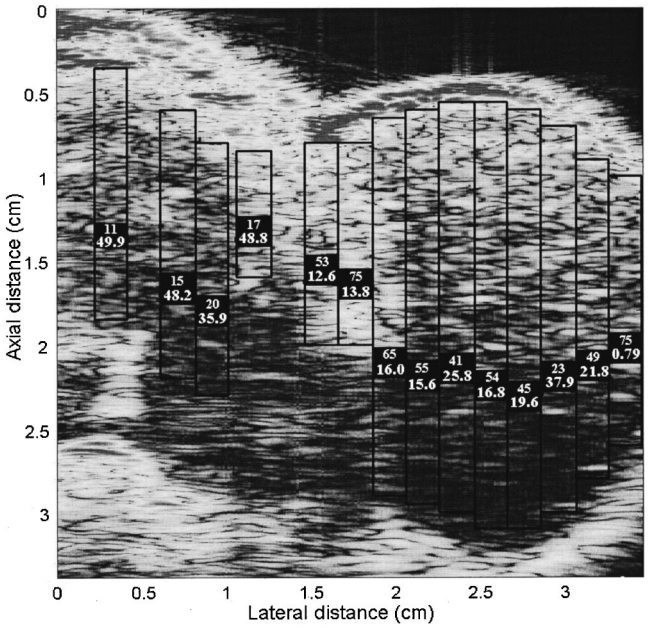


FIG. 8. Enhanced B-mode image of rat 4 with the tumor located between the lateral distance of 1.6 to 3.5 cm. The top number in the boxes represents the average scatterer diameter (μm) and the bottom number represents the average acoustic concentration (10* log[mm^{-3}]).

cause, in most of the examined cases, the error bars were actually larger than the magnitude of the acoustic concentration. The large standard deviation means that the average acoustic concentration might not be a precise estimation. An estimate of the average acoustic concentration was made from the estimated intercept of the best-fit line and the estimated average scatterer radius to the sixth power. Errors in the estimate of the average scatterer radius lead to much larger errors in the average acoustic concentration. In the case where there is definite separation (no overlap between the error bars) between the average scatterer sizes (rats 2–4), a definite difference was seen between the estimated average acoustic concentration inside and outside the tumor. Where the estimated average scatterer diameters inside the tumors and outside the tumors were close (rat 1), the difference between the average acoustic concentration is very small and not a good indicator of tissue structure difference.

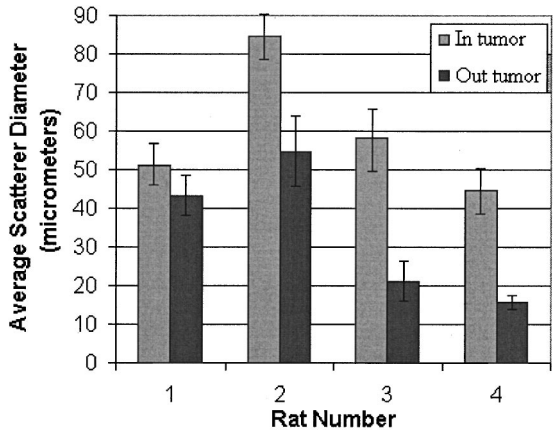


FIG. 9. Mean value of average scatterer diameters inside and outside the tumors. Error bars represent one full standard deviation about the mean.

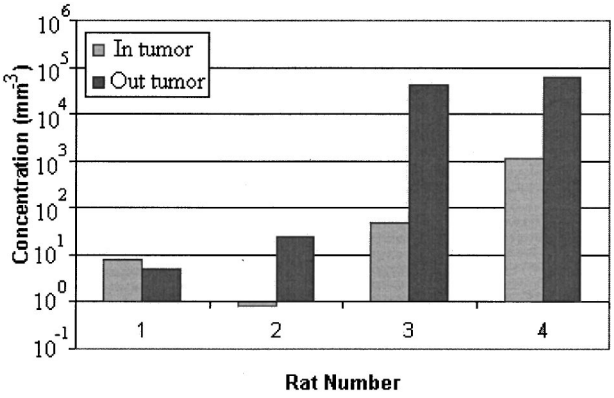


FIG. 10. Mean value of the average acoustic concentration inside and outside the tumors.

In order to verify the theory and acoustic estimates from the tumors, a comparison with the tissue microstructure through other means is necessary. Comparing the light microscopic image of Fig. 4 with the acoustic estimates of average scatterer size shows that there are structures on the same order in size. However, relating the light microscopy images with acoustic estimates may not be a valid comparison. The acoustical technique measures mechanical property changes in the tissue structure, while the light microscopy measures electromagnetic properties of the tissues. There may not be a one-to-one correlation between changes in the two properties. Much work needs to be done to verify the acoustical theory and estimation scheme by relating the estimates to actual tissue structures.

VI. CONCLUSION

An estimation scheme was devised to rapidly assess the average scatterer diameter and average acoustic concentration from soft tissues using the Gaussian form factor model. The estimation scheme works by using least squares to fit a line to the measured form factor. The precision of the estimation scheme was shown to be equivalent to the MASD scheme used by Insana *et al.*[11,12] to obtain scatterer size estimates. The best-fit line estimation scheme takes advantage of the unique functional form of the Gaussian form factor to reduce calculation time and give rapid estimates. Previous estimation schemes assumed that the power spectrum is linear with frequency.[16] The best-fit line estimation scheme assumes linear dependence for the measured form factor and not the power spectrum.

Four rats that had developed spontaneous mammary tumors were obtained and scanned with an ultrasonic transducer. The tumors and surrounding healthy tissues were interrogated in order to compare the average scatterer properties. The rf backscattered signals were gated from ROIs that were selected in each of the rats. Enhanced B-mode images were constructed from conventional B-mode images superimposed with estimated scattering parameters. The enhanced images were used to visually detect differences in tissue microstructure properties.

Comparisons of average scatterer diameters estimated inside the tumors and in the surrounding normal tissues showed that distinctions existed between the average scat-

terer properties. The distinctions could be used to classify tissues according to their microstructure. The distinctions could further be used to aid in the noninvasive diagnosis of disease.

Future work needs to be done to better account for the frequency-dependent losses produced by attenuation. The assumption that attenuation is homogeneous throughout the rat is not necessarily true. In most cases, it is expected that the attenuation and scattering properties will deviate from one region to the next. If the deviations are substantial, then a more detailed mapping of the attenuation throughout different regions will be needed to correctly account for the frequency-dependent losses to the measured power spectrum.

This study has shown that scattering properties like the average scatterer diameter and average acoustic concentration can readily be estimated from rf backscattered signals. Meaningful analysis of the statistical properties between scatterer parameters inside and outside the tumors is limited by the spontaneous nature of the tumors. The tumors grew at different places on each rat in an uncontrolled fashion. In each comparison, different surrounding tissues were compared with the tumor tissues. Future work will be done in implanting solid tumor cell lines in rats and closely managing their growth. In creating a managed line of tumors in rat subjects, a more meaningful comparison of the statistical properties of scatterers in healthy and diseased tissues can be made.

ACKNOWLEDGMENTS

Thanks to James P. Blue and Rita J. Miller, DVM, for their technical assistance. This work was supported by NIH Grant Nos. CA09067 and CA79179.

[1] J. F. Greenleaf and C. M. Sehgal, *Biologic System Evaluation with Ultrasound* (Springer, New York, 1992), pp. 101–103.
[2] M. F. Insana, ''Modeling acoustic backscatter from kidney microstructure using an anisotropic correlation function,'' J. Acoust. Soc. Am. **97**, 649–655 (1995).
[3] Y. Miyazaki, ''Light scattering of laser beams by random microinhomogeneities in glasses and polymers,'' Jpn. J. Appl. Phys. **13**, 1238–1248 (1974).
[4] M. Tobiyama, I. Endo, T. Monaka, Y. Sumi, H. Uchida, N. Yamamoto, K. Yoshida, K. Watanabe, T. Ohba, K. Baba, and T. Emura, ''Determination of atomic form factors by means of coherent bremsstrahlung,'' Phys. Rev. B **44**, 9248–9258 (1991).
[5] D. K. Wilson, J. G. Brasseur, and K. E. Gilbert, ''Acoustic scattering and the spectrum of atmospheric turbulence,'' J. Acoust. Soc. Am. **105**, 30–34 (1999).
[6] E. J. Feleppa, F. L. Lizzi, D. J. Coleman, and M. M. Yaremko, ''Diagnostic spectrum analysis in ophthalmology: A physical perspective,'' Ultrasound Med. Biol. **12**, 623–631 (1986).
[7] F. L. Lizzi, M. Ostromogilsky, E. J. Feleppa, M. C. Rorke, and M. M. Yaremko, ''Relationship of ultrasonic spectral parameters to features of tissue microstructure,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 319–329 (1986).
[8] M. F. Insana, T. J. Hall, and J. L. Fishback, ''Identifying acoustic scattering sources in normal renal parenchyma from the anisotropy in acoustic properties,'' Ultrasound Med. Biol. **17**, 613–626 (1991).
[9] K. K. Shung and G. A. Thieme, *Ultrasonic Scattering in Biological Tissues* (CRC Press, Boca Raton, 1993).
[10] J. J. Faran, Jr., ''Sound scattering by solid cylinders and spheres,'' J. Acoust. Soc. Am. **23**, 405–418 (1951).
[11] R. H. Silverman, M. J. Rondeau, F. L. Lizzi, and D. J. Coleman, ''Three-dimensional high-frequency ultrasonic parameter imaging of anterior segment pathology,'' Ophthalmology **102**, 837–843 (1995).
[12] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[13] F. L. Lizzi, M. Greenbaum, E. J. Feleppa, and M. Elbaum, ''Theoretical framework for spectrum analysis in ultrasonic characterization,'' J. Acoust. Soc. Am. **73**, 1366–1373 (1983).
[14] M. F. Insana, R. F. Wagner, D. G. Brown, and T. J. Hall, ''Describing small-scale structure in random media using pulse-echo ultrasound,'' J. Acoust. Soc. Am. **87**, 179–192 (1990).
[15] M. F. Insana and T. J. Hall, ''Parametric ultrasound imaging from backscatter coefficient measurements: Image formation and interpretation,'' Ultrason. Imaging **12**, 245–267 (1990).
[16] F. L. Lizzi, M. Astor, T. Liu, C. Deng, D. J. Coleman, and R. H. Silverman, ''Ultrasonic spectrum analysis for tissue assays and therapy evaluation,'' Int. J. Imaging Syst. Technol. **8**, 3–10 (1997).
[17] D. Nicholas, ''Evaluation of backscattering coefficients for excised human tissues: Results, interpretation, and associated measurements,'' Ultrasound Med. Biol. **8**, 17–28 (1982).
[18] D. K. Nassiri and C. R. Hill, ''The use of angular scattering measurements to estimate structural parameters of human and animal tissues,'' J. Acoust. Soc. Am. **79**, 2048–2054 (1986).
[19] K. A. Topp, J. F. Zachary, and W. D. O'Brien, Jr., ''Quantifying B-mode images of *in vivo* rat mammary tumor with frequency dependence of backscatter,'' J. Ultrasound Med. **20**, 605–612 (2001).
[20] M. Akita and M. Ueda, ''The effect of windowing on spectral estimation of echoes scattered by a random medium,'' J. Acoust. Soc. Am. **83**, 1243–1248 (1988).
[21] M. F. Insana, E. L. Madsen, T. J. Hall, and J. A. Zagzebski, ''Tests of the accuracy of a data reduction method for determination of acoustic backscatter coefficients,'' J. Acoust. Soc. Am. **79**, 1230–1236 (1986).
[22] J. F. Chen, J. A. Zagzebski, and E. L. Madsen, ''Tests of backscatter coefficient measurement using broadband pulses,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 603–607 (1993).
[23] K. A. Wear, R. F. Wagner, M. F. Insana, and T. J. Hall, ''Application of autoregressive spectral analysis to cepstral estimation of mean scatterer spacing,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 50–58 (1993).
[24] G. A. Teotica, R. J. Miller, L. A. Frizzell, J. F. Zachary, and W. D. O'Brien, Jr., ''Attenuation coefficient estimates of mouse and rat chest wall,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 593–600 (2001).
[25] R. A. Sigelmann and J. M. Reid, ''Analysis and measurement of ultrasound backscattering from an ensemble of scatterers excited by sine-wave bursts,'' J. Acoust. Soc. Am. **53**, 1351–1355 (1973).
[26] M. O'Donnell and J. G. Miller, ''Quantitative broadband ultrasonic backscatter: An approach to nondestructive evaluation in acoustically inhomogeneous materials,'' J. Appl. Phys. **52**, 1056–1065 (1981).
[27] M. L. Oelze and W. D. O'Brien, Jr., ''Comparisons of frequency-dependent attenuation-compensation functions for ultrasonic signals backscattered from random media,'' J. Acoust. Soc. Am. **111**, 2308–2319 (2002).
[28] S. L. Bridal, P. Fornes, P. Bruneval, and G. Berger, ''Parametric (integrated backscatter and attenuation) images constructed using backscattered radio frequency signals (25–56 MHz) from human aortae *in vitro*,'' Ultrasound Med. Biol. **23**, 215–229 (1997).
[29] J. A. Zagzebski, Z. F. Lu, and L. X. Yao, ''Quantitative ultrasound imaging: *in vitro* results in normal liver,'' Ultrason. Imaging **15**, 335–351 (1983).
[30] E. J. Feleppa, T. Liu, A. Kalisz, M. C. Shao, N. Fleshner, and V. Reuter, ''Ultrasonic spectral-parameter imaging of the prostate,'' Int. J. Imaging Syst. Technol. **8**, 11–25 (1997).
[31] M. F. Insana, T. J. Hall, J. G. Wood and Z-Y Yan, ''Renal ultrasound using parametric imaging techniques to detect changes in microstructure and function,'' Invest. Radiol. **28**, 720–725 (1993).
[32] F. L. Lizzi, M. Astor, A. Kalisz, T. Liu, D. J. Coleman, R. Silverman, R. Ursea, and M. Rondeau, ''Ultrasonic spectrum analysis for assays of different scatterer morphologies: Theory and very-high frequency clinical results,'' Proceedings of the 1996 IEEE Ultrasonic Symposium, 1155–1159 (1996).
[33] P. Chaturvedi and M. F. Insana, ''Error bounds on ultrasonic scatterer size estimates,'' J. Acoust. Soc. Am. **100**, 392–399 (1996).